ENJEUX PRATIQUES ET MÉTHODOLOGIQUES FACE AUX DIMENSIONS SOCIALE ET POLITIQUE DE L'ÉVALUATION DE PROGRAMME : LE CAS D'ÉVALUATEURS NOVICES ET EXPERTS EXERÇANT DANS DES CABINETS DE CONSEIL DANS LE SECTEUR PUBLIC CANADIEN

Maud Mediell

Thèse soumise à l'Université d'Ottawa dans le cadre des exigences du programme

Doctorat en philosophie en éducation (Ph.D.)

Faculté d'éducation
Université d'Ottawa

© Maud Mediell, Ottawa, Canada, 2023

Cette thèse a été réalisée sous la direction du Professeur Eric Dionne

Elle a été évaluée par un comité composé des professeurs suivants :

Robert Flynn

Michel D. Laurier

Nicholas Ng-A-Fook

Claire Tourmen (examinatrice externe)

REMERCIEMENTS

Mon parcours doctoral fut rythmé par des moments de joie, de collégialité, de succès et de fierté, mais aussi par de grandes pertes, de profondes peines, des périodes de doutes et d'isolement, et de grosses frustrations... Alors, comme promis, je tiens tout d'abord à me remercier. L'achèvement de cette thèse fut l'un des plus grands défis de ma vie, et aujourd'hui je peux dire que je me suis très reconnaissante d'avoir persévéré et produit un travail dont je peux être fière.

Bien entendu, un tel accomplissement demande le soutien de tout un village. Et quel village !

Tout d'abord, je tiens à faire part de ma gratitude envers mon directeur de thèse, Pr Eric Dionne. Eric, merci pour ton soutien, ta patience, ton mentorat et ta confiance. Merci de m'avoir transmis ton savoir, ta passion, et ta soif de rigueur. Merci d'avoir cru en moi et mes capacités malgré les épreuves, et de m'avoir poussé à persévérer et à me dépasser. Merci de m'avoir épaulé et offert une oreille attentive dans les moments de vie difficiles. On l'a fait !

Merci infiniment à la Pre Éliane Dulude de m'avoir accompagné dans la dernière ligne droite, et de m'avoir aidé à surmonter les difficultés administratives (et autres problèmes) de « dernières minutes », fortes sources d'anxiété. Et aussi un gros merci à Manon Éthier !

Je suis très reconnaissante envers les membres de mon comité de thèse, Prs Robert Flynn, Michel Laurier et Dany Lavault (*remplacé à la phase d'évaluation de la thèse*), pour leur investissement dans cette thèse, leurs attentions aux détails, leurs conseils et commentaires précieux qui ont grandement contribué à sa qualité. Merci, Pr Nicholas Ng-A-Fook, d'avoir accepté de rejoindre mon comité et d'avoir franchi la ligne d'arrivée avec nous. Pre Claire Tourmen, je suis tellement heureuse que vous ayez accepté d'être mon évaluatrice externe. Merci d'avoir partagé avec moi votre regard d'experte sur le développement et les caractéristiques de la pratique des professionnels de l'évaluation de programme.

Un clin d'œil tout particulier au Pr Flynn, qui a pris sous son aile, il y a près de treize ans maintenant, une étudiante française en échange international, et lui a fait découvrir l'évaluation de programme. Bob, merci d'avoir partagé et communiqué votre passion pour l'évaluation de programme. Merci de m'avoir accompagné, de près ou de loin, au fil des années dans mon parcours académique et professionnel.

Je dois également ce travail aux évaluateurs de programme qui ont accepté de participer à cette étude doctorale. Thank you for sharing your experiences as program evaluators. For allowing

me to explore and describe your real-world evaluation practices, the constraints that inhabit the different situations you've encountered (and are encountering), the methodological strategies you mobilize to ensure the quality and utility of your program evaluations and the challenges that remain. It is a great contribution.

À mon trio de supporteurs, mes collègues et mentors, Dre Marie-Hélène Chomienne, Dr Jean Roy et Pr Jean Grenier, merci pour votre amitié, votre soutien indéfectible, et votre confiance inébranlable. Je serais perdu sans vous. Joël et Karine, je n'ai pas les mots pour exprimer ce que votre amitié toutes ces années représente pour moi. Merci d'avoir traversé, avec moi, les hauts et les bas de ce parcours doctoral, mais aussi tout simplement de la vie. Votre soutien et votre confiance en moi ont été une source de force et de motivation. Mel, tu es une constante dans ma vie depuis plus de dix ans maintenant. Merci d'avoir été une oreille attentive toutes ces années, une source d'aventure, de joie et de réconfort en dehors de ma bulle de vie de doctorante. Ludivine, ma *coachette*, je ne sais pas ce que j'aurais fait sans toi cette dernière année. Merci ! I am also beyond grateful for my cohort sisters, Gene and Thursica, whom I spent most of my weekends and summers with at the library, keeping each other motivated and encouraging each other to stay positive while working on our thesis, sharing our joys and frustrations over snacks from Pivik. For all these moments, and many more meaningful ones, thank you! Also, thank you to Jonathan for all the times we've met at Café Nostalgica after a long day. These moments were precious. Thank you for your friendship and support throughout this journey. And to my *twinsie*, Heather, thank you for your friendship, our Shonda Rhimes Friday nights, our clay nights, and other artsy endeavours.

À ma famille, mon clan Mediell-Mourier-Orsini, merci pour votre amour et soutien inconditionnel. Merci d'avoir toujours cru en moi, même dans mes plus grands moments de doute. Je vous aime fort. À mes parents Marie-Thérèse et Didier, merci de m'avoir permis de faire *Tanguette* pendant un an (au total *lol*) afin que je puisse terminer la rédaction de cette thèse dans le meilleur environnement possible. Maminette, tu as été mon coach, ma plus fervente supportrice, ma confidente/thérapeute (et parfois le réceptacle de mes crises de colère et de frustration, désolée…), ma conseillère rédactionnelle, et j'en passe. Merci de tout mon cœur, tu es mon héros. Didou, merci d'être mon port dans la tempête, ma source de zénitude et de courage. Margaux, ma petite sœur, ma jumelle de cœur et d'esprit, merci d'être toujours là pour moi. Merci d'avoir

partagé ta force, et de m'avoir aidé à franchir cette ligne d'arrivée qui me paraissait toujours si loin. Tu es la meilleure ! Merci, Adrien, d'être un petit frère qui croit toujours en sa grande sœur. Merci, Pierre, d'être un parrain attentif et bienveillant, mais un peu brusque quand c'est pour mon bien. Merci à mon papa, Marcel, qui m'a appris que ça vaut la peine de prendre des risques, de se lancer vers l'inconnu, et qu'aucun rêve n'est trop grand. Et bien sûr une énorme pensée, remplie d'amour, pour mes grands-parents.

Enfin, la rédaction de cette thèse n'aurait pas pu être possible sans mon compte Spotify © qui m'a permis des milliers d'heures d'écoute de mes *playlists*, artistes et albums préférés où que je sois. Mentions spéciales — The Lumineers, Muse, Les Quatre Saisons de Vivaldi, Carmina Burana de Carl Orff, les bandes originales de La La Land, Le fabuleux destin d'Amélie Poulain et The Theory of Everything, les *playlists* Epic Classical, 70's Road Trip et 70'& 80's Acoustic.

À mon frère, Guillaume,
qui veille sur nous depuis l'Empire des anges...

« Entre
Ce que je pense,
Ce que je veux dire,
Ce que je crois dire,
Ce que je dis,
Ce que vous avez envie d'entendre,
Ce que vous entendez,
Ce que vous avez envie de comprendre,
Ce que vous comprenez,
Il y a dix possibilités qu'on ait des difficultés à communiquer.
Mais essayons quand même... »
– Bernard Werber

RÉSUMÉ

Au fil des générations, l'évaluation de programme a traversé de nombreux changements (nécessaires), qui l'ont conduite à devenir une activité à la fois technique et sociopolitique. Aujourd'hui, elle est d'une indéniable complexité du fait notamment de la nature des situations évaluatives rencontrées, dont les caractéristiques influencent la faisabilité, la fiabilité et l'utilité de l'évaluation. Des situations animées de contraintes sociopolitiques et organisationnelles importantes, de problèmes sur le plan des intrants disponibles, de l'accessibilité des données, etc. En outre, la qualité de l'évaluation est tributaire de la capacité de l'évaluateur à mobiliser des savoirs pratiques en situation, son expertise, et son jugement afin d'entreprendre les choix méthodologiques les plus appropriés.

Inscrite dans une perspective interprétative, cette étude explore et tente de comparer les choix méthodologiques opérés par des évaluateurs novices et experts œuvrant dans des cabinets de conseil (n=8) lors de la réalisation de leurs évaluations de programme ; et la façon dont ils appréhendent et gèrent les contraintes situationnelles. Elle mobilise deux types d'activités reconnues pour l'analyse des pratiques professionnelles des évaluateurs de programmes : 1. la métaévaluation d'un projet tierce rapportée durant un entretien non dirigé ; et 2. l'entretien d'explicitation centré sur un projet d'évaluation passé avec, en amont, l'analyse des devis évaluatifs faisant l'objet de l'exercice d'explicitation. À partir de ces deux types d'activités, il devient possible d'examiner et croiser les bonnes pratiques et les pratiques réelles des participants.

Les résultats s'inscrivent dans le contexte de l'évaluation externe de programmes publics fédéraux sous la direction d'un chef d'évaluation interne responsable de la planification initiale et la suppression de l'évaluation, et la consultation d'un comité interne composé d'utilisateurs ciblés. Au sein des équipes évaluatives externes, on observe une certaine hiérarchie dans l'organisation du travail. L'évaluateur expert est le gestionnaire de l'évaluation et il est chargé d'activités cruciales comme la conception et la communication de l'évaluation. L'évaluateur novice peut exercer dans une capacité d'exécutant/exécutant avancé de tâches de collecte et d'analyse de données (travail en silo), ou de collaborateur et participer à l'évaluation dès sa conception. Ultimement, les choix méthodologiques sont du ressort de l'évaluateur expert.

Les pratiques réelles des évaluateurs novices sont ancrées dans le développement de leurs compétences méthodologiques et de leurs capacités analytiques, ce qui transparaît sur le plan de leurs bonnes pratiques. L'articulation de leurs pratiques se centre sur une logique technique/méthodologique de production de connaissances quant aux objets évalués. Pour eux, il est essentiel de démontrer le caractère approprié de la démarche d'évaluation, l'effort de rigueur, la fiabilité et la représentativité de la méthodologie évaluative (c.-à-d. l'alignement du contenu du cadre d'évaluation, le caractère systématique des procédures, la représentativité des sources d'information, la transparence du raisonnement évaluatif, etc.). Si dans le cadre de l'analyse critique d'un devis tierce, les évaluateurs novices reconnaissent l'influence de différents facteurs contextuels sur l'évaluation, leur analyse demeure superficielle et décontextualisée comparativement aux évaluateurs experts. En effet, les pratiques réelles tout comme les pratiques recommandées des évaluateurs experts sont influencées par la situation évaluative réelle et centrée sur les résultats. La priorité, pour eux, est de veiller au caractère approprié et faisable du devis d'évaluation à la lumière des caractéristiques de l'objet évalué ou *evaluand* (complexité de la théorie de programme, envergure des objectifs, caractère facilement observable de ses résultats ou non, etc.), des exigences politico-organisationnelles (la Politique évaluative fédérale, ses directives méthodologiques et ses normes, la collaboration avec le chef de l'évaluation interne, etc.), et des ressources disponibles (temps, argent, données probantes existantes et accessibles, etc.). Et ce, tout

en assurant la fiabilité, la représentativité et l'utilité de l'évaluation. Il s'agit d'un raisonnement que les évaluateurs novices semblent commencer à s'approprier dès lors qu'ils ont l'occasion de participer à l'évaluation dès la phase de conception.

Les résultats montrent que la pertinence du devis d'évaluation repose sur la capacité des évaluateurs à réaliser un diagnostic efficace de la situation évaluative rencontrée afin de revoir et opérationnaliser les paramètres de l'évaluation initialement prescrits par les clients, et de s'assurer de la faisabilité et de la légitimité de l'évaluation. Cette entreprise nécessite la mobilisation de savoirs méthodologiques procéduraux et expérientiels, de connaissances spécifiques à l'environnement professionnel (programmes publics, système politico-organisationnel du gouvernement fédéral, politique évaluative, gouvernance de l'évaluation externe de programmes publics, etc.), de savoir-faire relationnels, et de savoir-faire cognitifs acquis dans et par l'expérience (indispensables à l'analyse de la situation et à la résolution de problèmes, à la conception et à la réalisation de l'évaluation). Les résultats soulignent ici l'incidence du niveau d'expertise de l'évaluateur de programme externe sur sa capacité à anticiper, identifier, comprendre et gérer efficacement les contraintes situationnelles afin de concevoir une évaluation fiable et utile. Ils montrent l'importance d'offrir le plus tôt possible aux évaluateurs novices l'occasion de participer à des projets d'évaluation du début à la fin, d'accroître les occasions d'observation et de contribution sur le plan de la conception et de la gestion de l'évaluation dans le cadre de petits projets où le devis est plutôt simple. Des tâches cruciales qui, dans le présent contexte de pratique, sont réalisées par les évaluateurs experts.

ABSTRACT

Throughout its history, Program Evaluation has gone through many (necessary) changes, which have led it to become both a technical and socio-political activity. Program evaluation activities can be complex depending on the nature and characteristics of the evaluative situations encountered. Situations with significant socio-political and organizational constraints, problems in terms of available resources, data accessibility, etc. Moreover, the evaluation quality depends on the evaluator's capacity to mobilize practical knowledge in situations, his expertise, and his judgment in order to make the most appropriate methodological choices.

Using an interpretative lens, this study explores and attempts to compare the methodological choices made by novice and expert evaluators working in consulting firms (n=8) when carrying out their program evaluations; and the way they apprehend and manage situational constraints. It mobilizes two types of activity recognized for analyzing professional practices of program evaluators: 1. the meta-evaluation of a third-party project reported during a nondirective interview; and 2. the "explicitation" interview focused on a past evaluation project with, beforehand, an analysis of the evaluation designs that are the subject of the interviews. Based on these two activities, it becomes possible to examine and cross-reference the participants' good practices with their actual practices. The results are set in the context of external evaluations of federal public programs under the direction of an internal evaluation manager responsible for the initial planning, and the consultation of an internal committee composed of intended users. Within external evaluation teams, there is a certain hierarchical organization of work. The expert evaluator is the evaluation manager, responsible for crucial activities such as designing and communicating the evaluation. The novice evaluator may act as an executant/advanced executant of data collection and data analysis tasks, or as a collaborator, participating in the evaluation from the beginning to the end. Ultimately, methodological choices are the responsibility of the expert evaluator.

The novice evaluators real practices are rooted in the development of their methodological and analytical skills, and reflected in their good practices. Novices' practices are focused on a technical/methodological logic of knowledge production regarding the objects evaluated. For them, it is essential to demonstrate the appropriateness of the evaluation process, the rigour, reliability and representativeness of the evaluation methodology (alignment of the evaluation framework's content, systematic nature of methodological procedures, representativeness of information sources, transparency of the evaluative reasoning, etc.). While novice evaluators recognize the influence of various contextual factors when critically analyzing a third-party evaluation, their analysis remains superficial and decontextualized compared to expert evaluators. Indeed, in the case of expert evaluators, both their real and recommended practices are influenced by the characteristics of the evaluative situation and focused on results. Their priority is to ensure the evaluation design is appropriate and feasible considering the characteristics of the object evaluated or *evaluand* (complexity of the program theory, scope of its objectives, whether or not its results are easily observable, etc.), the political and organizational context/constraints (the Federal Evaluation Policy, its methodological guidelines and standards, the collaboration with the internal evaluation chief, etc.), and the resources available (time, money, existing and accessible data, etc.). All the while ensuring the evaluation's reliability, representativeness and usefulness. This is a line of reasoning that novice evaluators seem to begin to appropriate when they start participating in evaluation projects from the design phase onwards.

The results show that the relevance of the evaluation design relies on the evaluators' ability to carry out an effective diagnosis of the evaluation situation encountered, to review and operationalize the evaluation parameters initially prescribed by the clients and ensure the feasibility and legitimacy of the evaluation. This undertaking requires the mobilization of procedural and experiential methodological knowledge, knowledge specific to the professional environment (public programs, the federal government's political-organizational system, evaluation policy, governance of the external evaluation of public programs, etc.), relational skills, and cognitive know-how acquired during and through experiences (essential for analyzing the situation and solving problems, designing and carrying out the evaluation). The results here highlight the impact of the external program evaluator's level of expertise on his ability to anticipate, identify, understand and effectively manage situational constraints to design a reliable and useful evaluation. They show the importance of offering novice evaluators the opportunity to participate in evaluation projects from start to finish, as soon and as much as possible. To invite them to observe and then contribute to the design and management of small evaluation projects with fairly straightforward methodology. Crucial tasks that are carried out by expert evaluators, in this context of practice.

Tables des matières

REMERCIEMENTS ..... **iii**
RÉSUMÉ ..... **viii**
ABSTRACT ..... **ix**
Liste des tableaux ..... **xiv**
Liste des figures ..... **xvi**
Liste des encadrés ..... **xvii**
Liste des annexes ..... **xviii**
CHAPITRE 1. PROBLÉMATIQUE **1**
1.1 Contexte général **1**
1.1.1 La traversée des courants en évaluation de programme : de la génération de la mesure à celle de la négociation. 2
1.1.1.1 Avant les années 1960 et la *Great Society Era* 2
1.1.1.2 Le début des années 1960 *et après* : l'explosion des modèles en évaluation de programme 4
1.1.2 Les enjeux contemporains en évaluation de programme 11
1.1.2.1 L'influence du mouvement pragmatiste sur les pratiques actuelles 13
1.1.2.2 La dimension « méthodologique » au cœur des préoccupations actuelles 19
1.2 Contexte spécifique **22**
1.2.1 L'influence des facteurs contextuels sur la méthodologie évaluative 24
1.2.2 Le niveau d'expertise des évaluateurs de programme 30
1.3 Objet de recherche **32**
1.4 Questions de recherche **35**
CHAPITRE 2. RECENSION DES ÉCRITS ET CADRE CONCEPTUEL **36**
2.1 La variable « méthodologie » au cœur des enjeux contemporains en évaluation **37**
2.1.1 Définitions actuelles de l'évaluation de programme 37
2.1.2 Le caractère itératif de la démarche évaluative 42
2.1.3 Définitions du devis d'évaluation et des choix dits méthodologiques 45
2.1.4 Synthèse conceptuelle 48
2.2 La dimension sociopolitique de l'évaluation et la gestion des contraintes ESR, des facteurs contextuels au cœur des enjeux méthodologiques **49**
2.2.1 Le rôle des parties prenantes clés et les enjeux méthodologiques face à la dimension sociopolitique de l'évaluation 49
2.2.1.1 Définition du concept de partie prenante 49
2.2.1.2 Les parties prenantes clés au cœur du mouvement de l'utilisation de l'évaluation 52
2.2.1.3 Les revers du mouvement de l'utilisation et de l'implication des parties prenantes clés 58
2.2.1.3.1 L'attitude réfractaire des parties prenantes clés face à l'évaluation 58
2.2.1.3.2 Les risques de mésévaluation, de non-utilisation et de mésusage 62
2.2.2 Les contraintes inhérentes au terrain (ou contraintes ESR) 65
2.2.3 Les enjeux de formation face à la dimension sociopolitique de l'évaluation 71
2.2.4 Synthèse conceptuelle 75
2.3 Les caractéristiques de la pratique des évaluateurs de programme en fonction de leur niveau d'expertise **76**

2.3.1 L'expertise et le développement professionnel 76
2.3.1.1 Qu'est-ce que l'expertise et comment se développe-t-elle? 76
2.3.1.2 Le développement professionnel (savoir agir avec compétence) 78
2.3.2 La pratique évaluative ou les choix des évaluateurs dans le cadre de leur pratique selon qu'ils soient novices ou experts 83
2.3.2.1 Les caractéristiques générales de la pratique des professionnels de l'évaluation de programme 83
2.3.2.2 Les apports de l'expérience pour apprendre à évaluer efficacement 86
2.3.3 Synthèse conceptuelle 89
2.4 Synthèse générale du cadre conceptuel 90
CHAPITRE 3. POSTURE ÉPISTÉMOLOGIQUE ET CADRE MÉTHODOLOGIQUE 100
3.1 Positionnement épistémologique 100
3.2 Devis de recherche : l'étude de cas 102
3.3 Participants et processus de recrutement 105
3.3.1 Procédures de recrutement et de sélection des cas 106
3.3.1.1 Scénarios de recrutement en milieu universitaire : procédures et critères de sélection des cas 107
3.3.1.2 Scénario de recrutement au sein du gouvernement fédéral : procédures et critères de sélection des cas 110
3.3.1.3 Scénario de recrutement au sein de cabinets de consultants : procédures et critères de sélection des cas 111
3.3.2 Participants sélectionnés 114
3.4 Déroulement de la recherche 116
3.4.1 Première collecte de données : l'entretien métaévaluatif 116
3.4.1.1 Préparation de l'activité de métaévaluation 117
3.4.1.2 Instrumentation : référentiel d'analyse d'activité et guide d'entretien métaévaluatif 119
3.4.1.3 Procédures de passation de l'entretien métaévaluatif 121
3.4.1.4 Procédures d'analyse des données de l'entretien métaévaluatif 122
3.4.2 Deuxième collecte de données : l'entretien d'explicitation 124
3.4.2.1 Le référentiel d'analyse de l'activité évaluative : la grille d'analyse de rapports d'évaluation 124
3.4.2.1.1 Procédure de création de la grille d'analyse des rapports d'évaluation 125
3.4.2.1.2 Procédure de validation du contenu de la grille d'analyse de rapports d'évaluation 127
3.4.2.1.3 Expérimentation de la grille d'analyse des rapports d'évaluation 128
3.4.2.2 Le guide d'entretien d'explicitation 130
3.4.2.2.1 Procédure de création des guides d'entretien d'explicitation à partir de l'expérimentation des grilles d'analyse des rapports d'évaluation 130
3.4.2.2.2 Procédures de passation et d'analyse de l'entretien d'explicitation 134
3.5 Conclusion du chapitre 136
CHAPITRE 4. RÉSULTATS ET INTERPRÉTATION 139
4.1 Les éléments relevés par les évaluateurs quant aux choix méthodologiques et à l'influence des facteurs contextuels sur ceux-ci (*activité métaévaluative*) 141
4.1.1 Le cas des évaluateurs juniors (novices) 142
4.1.1.1 Le manque de précision au niveau de la description des méthodes de collectes et d'analyses de données au cœur du jugement critique des évaluateurs juniors 144
4.1.1.2 Le manque de précision et de pertinence en ce qui a trait aux sources d'information et caractéristiques des échantillons 149

4.1.1.3 Les limites identifiées par les évaluateurs juniors 152
4.1.1.4 Les facteurs contextuels exerçant une influence potentielle sur les éléments clés du devis d'évaluation 162
4.1.1.5 *Interprétation de la question de recherche tertiaire 1* – les pratiques dites recommandées des évaluateurs novices de programmes publics 170
4.1.2 Le cas des évaluateurs séniors (experts) 180
4.1.2.1 Le devis d'évaluation : une analyse de contenu ancrée dans la perception de la pertinence et de la précision-justesse de la démarche évaluative, et de la gestion des facteurs contextuels 181
4.1.2.1.1 L'analyse du contenu du cadre d'évaluation 181
4.1.2.1.2 L'analyse de contenu du cadre méthodologique 194
4.1.2.2 Les limites méthodologiques identifiées par les évaluateurs séniors (experts) 204
4.1.2.3 *Interprétation de la question de recherche tertiaire 1* – les pratiques dites recommandées des évaluateurs experts de programmes publics 219
4.2 Les informations fournies par les évaluateurs pour expliquer et décrire les choix méthodologiques entrepris durant l'évaluation (*activité d'explicitation*) 223
4.2.1 Le cas des évaluateurs juniors (novices) 223
4.2.1.1 Contenu du cadre d'évaluation : perceptions et explicitations ancrées dans l'influence de la gouvernance de l'évaluation et du cadre de référence gouvernemental 225
4.2.1.2 Communication du devis d'évaluation : perceptions et explicitations des carences observées sur le plan du contenu rapporté 233
4.2.1.3 Conception du cadre méthodologique : perceptions et explicitations ancrées dans les décisions de l'évaluateur sénior chargé du projet à l'externe et le comité d'évaluation interne 237
4.2.1.4 *Interprétation de la question de recherche tertiaire 2* - les pratiques réelles des évaluateurs novices de programmes de programmes publics 256
4.2.2 Le cas des évaluateurs séniors (experts) 271
4.2.2.1 Conception du cadre d'évaluation : explicitations du processus d'opérationnalisation des paramètres de l'évaluation prescrits dans la demande de propositions 273
4.2.2.2 Communication du devis d'évaluation : explicitations des carences observées sur le plan du contenu rapporté 276
4.2.2.3 Conception du cadre méthodologique 280
4.2.2.4 *Interprétation de la question tertiaire 2* – les pratiques réelles des évaluateurs experts de programmes publics 311
4.3 *Interprétation des questions de recherche secondaires* – les différences et les ressemblances entre les pratiques réelles et les pratiques recommandées des évaluateurs de programmes publics et l'influence des facteurs contextuels sur lesdites pratiques. 324
4.3.1 *Interprétation des questions de recherche secondaires* – le cas des évaluateurs de programmes publics « experts » 325
CHAPITRE 5. CONCLUSION 342
5.1 Interprétation générale 342
5.2 Discussion sur les limites et pistes de recherche futures 353
5.3 Quelques pistes de recommandations sur le plan du soutien au développement professionnel des évaluateurs de programme novices sur le terrain 361
Références 365
ANNEXES 380

Liste des tableaux

Tableau 1. *Les rôles successivement attribués à l'évaluateur au fil de l'évolution de l'évaluation de programme* 9
Tableau 2. *Le continuum de définitions du concept de partie prenante* 50
Tableau 3. *Adaptation des choix méthodologiques en réponse aux contraintes ESR (Bamberger et Mabry, 2020)* ... 68
Tableau 4. *Les cinq stades du développement professionnel (Dreyfus et Dreyfus, 1980)* 77
Tableau 5. *L'équipement incorporé dans lequel le praticien peut puiser pour agir avec compétences (Le Boterf, 2002)* 79
Tableau 6. *Référentiel des compétences requises à l'exercice de l'évaluation de programme au Canada (SCÉ, 2018)* 82
Tableau 7. *Caractéristiques générales des professionnels de l'évaluation (débutants et expérimentés)* 84
Tableau 8. *Présentation des caractéristiques des participants de la recherche* 114
Tableau 9. *Normes et critères retenus pour la construction de la grille d'analyse de rapports d'évaluation* 126
Tableau 10. *Structure de la présentation des résultats et de leur interprétation* 140
Tableau 11. *Contexte de réalisation de l'activité d'analyse critique selon les évaluateurs juniors* 142
Tableau 12. *Pratiques méthodologiques recommandées des évaluateurs novices dans le contexte de l'évaluation de programmes du gouvernement fédéral* 178
Tableau 13. *Caractéristiques de l'analyse critique selon l'expérience des évaluateurs séniors* 180
Tableau 14. *Description des cooccurrences entre les préoccupations liées à la fiabilité, validité et représentativité sur le plan du devis d'évaluation et les facteurs contextuels (évaluateurs séniors)* 206
Tableau 15. *Description des cooccurrences entre les préoccupations liées à la fiabilité, validité et représentativité sur le plan de la synthèse et de l'interprétation des résultats, et les facteurs contextuels (évaluateurs séniors 'S2' et 'S3')* 213
Tableau 16. *Thématiques et éléments d'analyse relatifs aux différentes formes de mérepréstentation identifiées par les évaluateurs séniors 'S2' et 'S3'* 215
Tableau 17. *Description des cooccurrences entre la remise en question de l'utilité de l'évaluation et les facteurs contextuels* 217
Tableau 18. *Pratiques recommandées par les évaluateurs experts dans le contexte de l'évaluation de programmes du gouvernement fédéral* 221
Tableau 19. *Activités réalisées par les évaluateurs novices dans le cadre de leur projet d'évaluation respectif* 224
Tableau 20. *Explications ou perceptions du processus de définition du cadre d'évaluation (évaluateurs juniors)* . 232
Tableau 21. *Suppositions et explicitations des évaluateurs juniors quant au manque de précision sur le plan du contenu du cadre d'évaluation rapporté* 234
Tableau 22. *Suppositions et explicitations des évaluateurs juniors quant au manque de précision sur le plan du contenu du cadre méthodologique rapporté* 236
Tableau 23. *Explications ou perceptions du choix des méthodes de collectes de données (évaluateurs juniors)* 245
Tableau 24. *Perceptions des procédures d'identification et de sélection des sources d'information (évaluateurs juniors 'J2', 'J3' et 'J4')* 247
Tableau 25. *Explicitations des procédures d'identification et de sélection des sources d'information (évaluateur junior 'J1')* 248
Tableau 26. Description ou explicitation des procédures de conception et de passation de l'instrumentation *(évaluateurs juniors)* 250
Tableau 27. *Explicitations ou perceptions sur les procédures d'analyse de données et de synthèse des résultats (évaluateurs juniors)* 251
Tableau 28. *Perceptions et interprétations sur l'obtention de faibles taux de réponse aux enquêtes (évaluateurs juniors 'J2' et 'J3')* 252
Tableau 29. *Explicitations ou perceptions/suppositions sur les procédures d'assurance de la fiabilité et de la scientificité (évaluateurs juniors)* 255

Tableau 30. *Synthèse des observations dans le cadre de l'analyse du contenu des différents devis d'évaluation rapportés par les évaluateurs séniors* 272
Tableau 31. *Explicitations des évaluateurs séniors quant aux carences observées sur le plan du contenu du devis d'évaluation présenté dans leur rapport d'évaluation respectif* 278
Tableau 32. *Objectifs des méthodes de collecte de données utilisées par les évaluateurs séniors* 286
Tableau 33. *La démarche de cadrage de l'évaluation entreprise par les évaluateurs séniors.* 315

Liste des figures

Figure 1. *Représentation temporelle de l'évolution de l'évaluation de programme* 1
Figure 2. *Représentation graphique de la démarche d'évaluation de programme* 13
Figure 3. *Evaluation Theory Tree (Alkin et Christie, 2013, p.14)* 14
Figure 4. *Synthèse du cadre conceptuel* 37
Figure 5. *Le cycle itératif de la démarche d'évaluation de programme* 45
Figure 6. *Étayage des choix méthodologiques propres à la conception et à la mise en œuvre du devis d'évaluation* 47
Figure 7. *Synthèse conceptuelle [1/3]* 48
Figure 8. *Conceptualisation de la définition de l'utilisation (Alin et King 2017)* 53
Figure 9. *Les répercussions de l'attitude réfractaire des parties prenantes clés sur l'évaluation* 61
Figure 10. *Représentation graphique du mésusage de l'évaluation* 64
Figure 11. *Synthèse conceptuelle [2/3]* 76
Figure 12. *Processus cognitif aboutissant à la formation de théories implicites (McPherson, 2015).* 88
Figure 13. *Synthèse conceptuelle (3/3)* 89
Figure 14. *Processus de codage éclectique des données qualitatives de l'activité métaévaluative* 123
Figure 15. *Extrait du tableau de validation de contenu de la grille d'analyse de rapports d'évaluation* 128
Figure 16. *Processus de codage éclectique des données qualitatives de l'activité d'explicitation* 136
Figure 17. *Protocole de l'étude de cas multiples sur les processus décisionnels relatifs aux choix dits méthodologiques des évaluateurs dans le cadre de leurs opérations d'évaluation de programme (inspiré de Kinney-Walker, 2015 ; Yin, 1994)* 138
Figure 18. *Les principaux axes et thématiques d'analyse mobilisés par les évaluateurs juniors dans le cadre de l'activité métaévaluative.* 143
Figure 19. *Thématiques et éléments d'analyse relatifs aux méthodes de collectes et d'analyse de données (évaluateurs juniors)* 144
Figure 20. *Thématiques et éléments d'analyse relatifs aux sources d'informations et échantillonnages (évaluateurs juniors)* 150
Figure 21. *Thématiques et limites identifiées par les évaluateurs juniors face au manque de description et d'explicitation du cadre méthodologique (évaluateurs juniors)* 153
Figure 22. *Distribution des fréquences de cooccurrences dans le cadre des interactions entre chaque groupe de facteurs contextuels et les éléments clés du devis d'évaluation* 163
Figure 23. *Fréquences de cooccurrence entre les éléments constitutifs de chacune des dyades [devis d'évaluation - groupe de facteurs contextuels]* 164
Figure 24. *Thématiques et éléments d'analyse relatifs au cadre d'évaluation (évaluateurs séniors)* 182
Figure 25. *Fréquences de cooccurrence entre les éléments constitutifs de chacune des dyades [cahier des charges - groupe de facteurs contextuels]* 184
Figure 26. *Principales thématiques et éléments d'analyse du cadre méthodologique (évaluateurs séniors).* 194
Figure 27. *Thématiques et éléments d'analyse relatifs aux sources d'information et échantillonnages (évaluateurs séniors)* 196
Figure 28. *Thématiques et éléments d'analyse sur les procédures de collectes de données (évaluateurs séniors)* .. 200
Figure 29. *Distribution des cooccurrences dans le cadre des interactions entre les différents axes de limites et les facteurs contextuels* 205
Figure 30. *Thématiques et éléments d'analyse sur la relation entre le manque d'explication du cadre méthodologique (et ses conséquences), et certains facteurs contextuels clés (évaluateurs séniors)* 207
Figure 31. *Thématiques et éléments d'analyse sur la relation entre le manque de pertinence et de précision-justesse sur le plan du contenu du devis d'évaluation et certains facteurs contextuels clés (évaluateurs séniors)* 209
Figure 32. *Dynamique entre les différentes entités gouvernantes de l'évaluation* 312
Figure 33. *Représentation graphique des caractéristiques des situations évaluatives rencontrées.* 317
Figure 34. *Les contributions de l'enseignement centré sur les cas (Bourgeois, Wildschut et Steinberg, 2021, p.41)* 364

Liste des encadrés

Encadré 1. *Définition canadienne « proxy » de l'évaluation de programme (SCÉ, 2015)* 42
Encadré 2. *Procédures du 1er scénario de recrutement en milieu universitaire* 108
Encadré 3. *Procédures du 2e scénario de recrutement en milieu universitaire* 109
Encadré 4. *Procédures du scénario de recrutement au sein du gouvernement fédéral* 111
Encadré 5. *Procédures du scénario de recrutement au sein de cabinets de consultants* 113
Encadré 6. *Extrait de l'utilisation de la grille d'analyse du rapport du participant [COH7]* 129
Encadré 7. *Extrait du guide d'entretien d'explicitation du participant [COH7]* 133
Encadré 8. *Thématiques et éléments d'analyse sur le manque de pertinence et précision-justesse du devis d'évaluation, de la faiblesse des preuves, et du rôle joué par certains facteurs contextuels (évaluateurs 'S2' et 'S3').* 214
Encadré 9. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J2'* 267
Encadré 10. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J4'* 268
Encadré 11. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J3'* 269
Encadré 12. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J1'* 270

Liste des annexes

Annexe 1. *Le pragmatisme et les approches évaluatives « participatives » contemporaines* 380
Annexe 2. *Le processus de formulation de la définition canadienne de l'évaluation de programme* 382
Annexe 3. *Les définitions actuelles de l'évaluation de programme* 383
Annexe 4. *La Taxonomie des compétences essentielles de l'évaluateur de programme (Stevahn, King, Ghere et Minnema, 2005) (Traduction libre)* 384
Annexe 5. *Certificats d'approbation éthique* 385
Annexe 6. *Matériel de recrutement des participants* 389
Annexe 7. *Instrumentation de la collecte de données numéro 1 – Référentiel d'analyse de l'activité et Guide d'entretien métaévaluatif* 394
Annexe 8. *Procédure de validation de contenu de la grille d'analyse de rapports d'évaluation centrée sur le devis d'évaluation* 396
Annexe 9. *Version finale de la grille d'analyse de rapports d'évaluation centrée sur le devis d'évaluation* 412

CHAPITRE 1. PROBLÉMATIQUE

1.1 Contexte général

La popularité croissante de l'évaluation de programme en fait aujourd'hui une discipline en vogue, dans de nombreuses sphères comme les sciences humaines et sociales et les sciences de la santé. L'évaluation telle que nous la pratiquons aujourd'hui a été façonnée dans sa traversée par quatre générations qui couvrent six périodes distinctes de son histoire (Stufflebeam et Coryn, 2014 ; Dubois, Champagne et Bilodeau, 2011 ; Stufflebeam et Shinkfield, 2007 ; Fishman, 1992 ; Madaus, Stufflebeam et Scriven, 1989 ; Guba et Lincoln, 1989) (Figure 1).

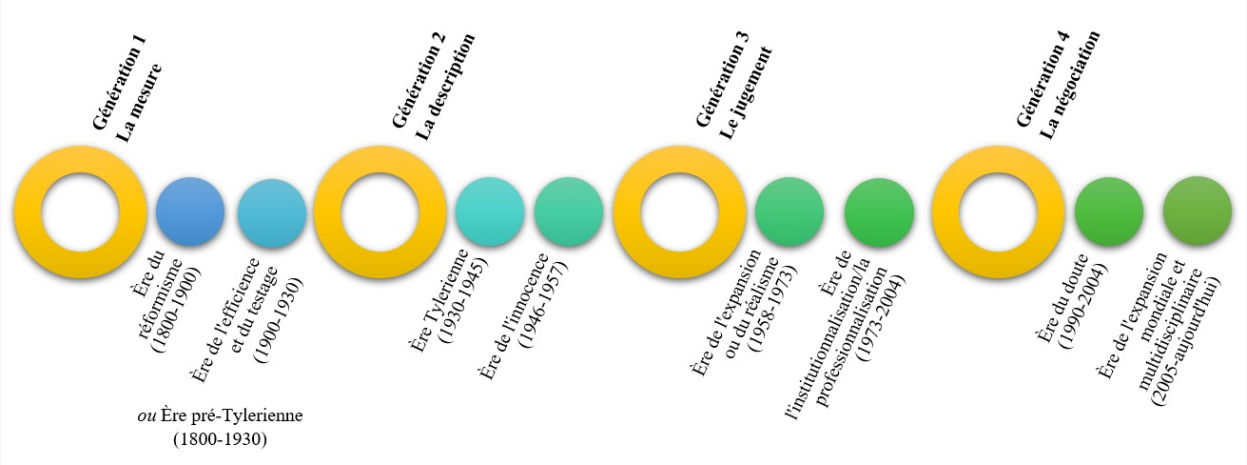


Figure 1. *Représentation temporelle de l'évolution de l'évaluation de programme*

La transformation de la discipline s'est fait au rythme de l'évolution des recherches en sciences sociales et a été ponctuée, entre autres, par le débat entre les deux traditions scientifiques, *qualitative* versus *quantitative*. L'évolution des pratiques en évaluation de programme a été rythmée par les contributions de théoriciens de l'évaluation que Stufflebeam et Shinkfield (2007) qualifient de « rebels ». Ils ont remis en question des philosophies, des traditions en matière d'évaluation qu'ils considéraient débilitantes. Par exemple, la prédominance des tests standardisés dans l'évaluation des programmes éducatifs, la philosophie de l'évaluation en tant que science exempte de valeur, la désignation des méthodes expérimentales telles qu'appliquées en laboratoire comme le « *gold standard* » des études sur le terrain, la tendance à poursuivre l'adéquation de l'évaluation avec les standards de la recherche expérimentale au détriment de son utilité, etc. L'évolution de l'évaluation de programme et les nombreuses problématiques rencontrées sur le

terrain ont ainsi entrainé différentes variations dans les approches méthodologiques, et un compromis évident (pour les chercheurs en évaluation et les évaluateurs), à savoir le recours aux méthodes quantitatives et qualitatives. Pour les praticiens et les chercheurs en évaluation, il est question de s'assurer d'une part de la validité de l'interprétation des données et de la qualité des méthodologies employées, et, d'autre part, de l'utilité des résultats et de la nécessité de travailler plus ou moins en partenariat avec les parties prenantes.

L'histoire de l'évaluation de programme exerce une influence majeure sur la manière dont nous la pratiquons aujourd'hui et elle contribue à la complexité de cette jeune discipline. Aussi, afin de dépeindre au mieux le contexte général dans lequel se déroule cette étude doctorale ainsi que les pratiques actuelles en évaluation, nous débutons par une brève section historique qui décrit l'évolution de l'évaluation de programme au fil des courants qu'elle a traversé, mais aussi de la complexification du terrain et de la relation que les évaluateurs entretiennent avec ce dernier. Ces quelques pages d'histoire nous permettent de rendre compte, entre autres, de l'émergence de certains enjeux méthodologiques auxquels la communauté évaluatrice a dû faire face, enjeux à l'origine de problèmes contemporains en évaluation de programme.

1.1.1 La traversée des courants en évaluation de programme : de la génération de la mesure à celle de la négociation.

Au cours de son histoire, l'évaluation a traversé quatre grandes générations et six périodes distinctes qui font ce qu'elle est aujourd'hui (Figure 1, pagre 1). Notons que l'Ère de l'expansion ou du réalisme, le mouvement de la *Great Society Era* et la reconnaissance de l'évaluation de programme en tant que champ disciplinaire, marquent un tournant majeur en évaluation. Dans cette courte description des courants de la discipline, nous nous concentrons particulièrement sur l'évolution méthodologique de l'évaluation, une évolution rythmée par l'expansion des pratiques évaluatives et la gestion des problématiques de terrain.

1.1.1.1 Avant les années 1960 et la *Great Society Era*

1re génération (de 1900 à 1930) : la mesure

Cette première génération est caractérisée par le *testing*, le perfectionnement des instruments de mesure ainsi que l'élaboration de tests standardisés et de méthodes de collecte

systématique de données (Dubois, Champagne et Bilodeau, 2011), telles que le test de mesure du quotient intellectuel de Binet (1905/1916). Plusieurs facteurs contextuels ont en outre contribué au développement de ce courant (Guba et Lincoln, 1989) : le premier a trait à la montée du phénomène « science sociale », qui proposait l'application de l'approche scientifique pour étudier les phénomènes humains et sociaux. Le second fut l'apparition du mouvement de la « gestion scientifique » dans le domaine des affaires et de l'industrie, avec le taylorisme et la standardisation des opérations pour améliorer le rendement des travailleurs et l'efficience des systèmes de production (Stufflebeam et Coryn, 2014; Stuffelbeam and Shinkfield, 2007). Enfin, durant cette période inscrite dans une tradition positiviste, on se concentrait alors sur l'élaboration de tests en mobilisant les techniques quantitatives, empruntées à la psychométrie et aux statistiques (Dubois, Champagne et Bilodeau, 2011).

2e génération et naissance de l'évaluation de programme (de 1930 à 1957 environ) : la description

La génération de la description s'est développée avec la nécessité d'évaluer l'impact des changements apportés aux curriculums dans le cadre d'expériences en éducation (Fishman, 1992). En effet, il s'agissait de développer une approche caractérisée par la description de la réalisation des curriculums (du processus), l'identification des points forts et des points faibles quant à leurs mises en application et la réponse aux objectifs préalablement fixés par ceux-ci (Guba et Lincoln, 1989). Il n'était donc plus question de se baser uniquement sur la mesure du fonctionnement individuel des élèves pour évaluer les effets des programmes en éducation. Par ailleurs, ce sont les travaux du chercheur positiviste Tyler (The Eight Years Study*,* 1933) qui illustreront cette deuxième génération ; celui-ci ayant développé une méthode qui permet la description du degré selon lequel certains objectifs éducatifs étaient accomplis (Dubois, Champagne et Bilodeau, 2011 ; Fishman, 1992). De 1930 à 1945, on se trouve alors dans l'Ère tylerienne à l'issue de laquelle l'approche de Tyler est grandement acceptée dans le domaine de l'éducation aux États-Unis, et influencera pendant des décennies la manière dont les éducateurs et les évaluateurs de programme perçoivent l'évaluation (Stufflebeam et Coryn, 2014). Cependant, pendant une dizaine d'années (Ère de l'innocence), les recommandations de Tyler feront l'objet de discussions plutôt que d'applications concrètes. Stufflebeam et Shinkfield (2007) parlent d'une période d'apathie sociale.

Au demeurant, selon le *Joint Committee on Standards for Educational Evaluation* (2011, 1994, 1981), pour certains évaluateurs et chercheurs en évaluation, le concept de description du

degré d'atteinte des objectifs d'un programme marquerait le réel point de départ de l'évaluation de programme et identifierait Tyler comme le père de l'évaluation. Tyler a d'ailleurs inventé (*coined*) le terme évaluation éducationnelle au début des années 1930. Cette deuxième génération qualifiée de courant de la description marque la naissance de l'évaluation de programme et l'utilisation de la mesure comme un instrument qui permet sa réalisation (Guba et Lincoln, 1989).

1.1.1.2 Le début des années 1960 *et après* : l'explosion des modèles en évaluation de programme

3e génération (1957-1989) : le jugement

L'ère de l'expansion ou du réalisme (1957-1973)

Cette période marque un renouveau, le début de la prolifération des évaluations de programmes, principalement dans le domaine de l'éducation (Stufflebeam et Coryn, 2014 ; Dubois, Champagne et Bilodeau, 2011 ; Stufflebeam et Shinkfield, 2007). Précisons qu'un évènement historique est à l'origine de l'arrêt brutal de l'ère de la complaisance (de l'apathie sociale), à savoir la mise sur orbite du satellite russe Spoutnik I en 1957. En effet, une telle prouesse met en évidence le retard technologique des États-Unis par rapport à l'Union soviétique et conduit le gouvernement américain à lancer immédiatement un projet de réforme visant à améliorer les programmes scolaires (plus particulièrement en sciences, mathématiques et langues étrangères). L'octroi d'investissements massifs par les pouvoirs publics dans le cadre de cette réforme s'accompagne alors d'une condition, soit l'évaluation des nouveaux programmes éducationnels. Le concept de reddition de compte se dessine alors, avec l'évaluation de la pertinence et le questionnement des retombées positives des investissements des pouvoirs publics tels que le ministère de l'Éducation. Face à cette recrudescence des évaluations de programmes, des chercheurs tels que Cronbach (1963) constatent une inadéquation entre les volontés des bailleurs de fonds et gestionnaires de programme et les méthodes évaluatives mobilisées à l'époque. Selon l'auteur, les méthodologies d'évaluation de la 2e génération utilisées pour évaluer le degré de réalisation des objectifs ne permettent pas de juger de la valeur et de la pertinence de ceux-ci au regard des besoins décelés. De plus, pour Cronbach (1963), en mesurant l'atteinte des objectifs *a posteriori*, on observe seulement les lacunes des programmes après les faits. On rejoint ici les frustrations de l'époque qui ont suscité un renouveau de l'évaluation de programme en éducation

et l'élargissement de son cadre conceptuel avec la proposition de nouvelles approches évaluatives. Des approches qui découlent de la nécessité, pour les évaluateurs de programme, d'évaluer les objectifs, d'examiner les intrants, d'examiner la mise en œuvre et la prestation des services, et de mesurer les résultats prévus et imprévus du programme (Stufflebeam et Coryn, 2014; Stuffelbeam and Shinkfield, 2007). Parmi les nouveaux modèles d'évaluation élaborés, on trouve des modèles qui proposent une reformulation de l'approche de Tyler tels que le modèle d'appréciation des écarts de Provus (1969). On trouve également des modèles qui divergent complètement des approches utilisées jusque-là, soit le modèle bimatriciel de Stake (1967), le modèle d'évaluation par induction de Scriven (1972), le modèle orienté vers la décision de Stufflebeam (1967). Les théoriciens concepteurs de ces nouvelles approches soulignent également la nécessité pour l'évaluateur de porter (ou de recueillir) des jugements sur le mérite ou la valeur de l'objet évalué (Stufflebeam et Coryn, 2014; Stuffelbeam and Shinkfield, 2007).

Par ailleurs, cette période est également marquée par l'institution Johnson (États-Unis) et la *Great Society Era*, connue pour le lancement d'un grand projet d'interventions sociales dont les premières évaluations n'ont pas été en mesure de démontrer l'utilité des programmes implantés, et les bénéfices réels des investissements financiers octroyés. Elles ont également conduit à la remise en question de l'efficacité de l'administration publique responsable de l'implantation des programmes. L'évaluation apparait alors comme une nécessité, une ressource incontournable de développement et de gestion de programmes. Elle doit être utilisée pour guider le processus de conception des programmes, de planification (élaboration des objectifs et intégration de facteurs observables facilitant la compréhension et l'analyse des processus de réalisation), de rationalisation de l'allocation/la distribution des ressources publiques ; et pour surveiller les processus administratifs et assurer l'imputabilité (Dubois, Champagne et Bilodeau, 2011 ; Fishman 1992).

Au regard des faits marquants de l'ère de l'expansion (ou du réalisme), les chercheurs en évaluation et les évaluateurs de programme s'accordent sur le fait qu'ils doivent également assumer le rôle de juge, car ils sont les experts désignés pour attester de la valeur et de l'efficacité des programmes, par le biais d'une investigation rigoureuse (Dubois, Champagne et Bilodeau, 2011 ; Fishman, 1992 ; Guba et Lincoln, 1989). Enfin, la 3^{e} génération de l'évaluation (*le*

jugement) marque également la professionnalisation de l'évaluation de programme, qui devient alors un champ disciplinaire à part entière (Dubois, Champagne et Bilodeau, 2011).

L'Ère de l'institutionnalisation et de la professionnalisation (1973 à 2004)

À partir des années 1970, différents éléments ont contribué à l'émergence de l'évaluation de programme comme un champ professionnel à part entière. Des efforts considérables sont entrepris pour la diffusion des connaissances, et toute une série de revues et de collections spécialisées apparait : *Educational Evaluation and Policy Analysis, Studies in Educational Evaluation, CEDR Quarterly, Evaluation Review, New Directions for Program Evaluation, Evaluation and Program Planning* (Stufflebeam et Madaus*,* 2000). La professionnalisation de l'évaluation entraine la création de plusieurs associations professionnelles qui visent à la promouvoir, mais aussi à soutenir les évaluateurs ainsi que l'avancement des connaissances (Dubois, Champagne et Bilodeau*.*, 2011). Parmi les associations qui voient le jour, on compte la Société canadienne d'évaluation. De plus en plus, le domaine s'est tourné vers la métaévaluation comme moyen d'assurer et de contrôler la qualité des évaluations (Stufflebeam, 2001, 1978). En 1975, treize associations professionnelles crées le *Joint Committee on Standards for Educational Evaluation* (dorénavant JCSEE). En 1981, le JCSEE publie des normes pour juger les évaluations des programmes, projets et matériels éducatifs. Il établit un mécanisme d'examen et de révision continu des normes ainsi que des ressources pour aider les évaluateurs à les utiliser (Stufflebeam et Coryn, 2014 ; Stufflebeam et Shinkfield, 2007). Il publiera une deuxième édition en 1994, puis une troisième en 2011. L'évolution vers la professionnalisation de l'évaluation est également associée à l'élargissement des possibilités de carrière. Plusieurs institutions publiques consacrées à l'évaluation sont créées dans divers pays, comme les bureaux nationaux d'évaluation et d'audit (Dubois, Champagne et Bilodeau, 2011).

La transition entre la 3e et la 4e génération et le débat animé, méthodes « quantitatives » versus méthodes « qualitatives »

Pendant très longtemps, l'évaluation était dépeinte comme une entreprise principalement quantitative (Cook, 1997). Dans les années 1960, les méthodes qualitatives étaient encore très peu considérées, même si on traversait progressivement la 3e génération et que des bouleversements théoriques et méthodologiques s'annonçaient. À cette époque, la principale préoccupation des

évaluateurs porte sur « *the black-box task of generating unbiased, precise estimates of the causal consequences of programs or their major constituent parts* » (Cook, 1997, p.30). En effet, les débuts de la 3e génération (avant le milieu des années 1970) restent ainsi inscrits dans la tradition du postpositivisme, où la pratique de l'évaluation de programme, comme toute recherche sociale, se concentrait principalement sur l'évaluation de la force des relations de cause à effet (entre une intervention et les résultats), par l'intermédiaire de devis expérimentaux ou quasi expérimentaux, appuyés par des mesures standardisées (Shaw, Greene et Mark, 2006).

Progressivement, avant le début des années 1980, les chercheurs ont commencé à remettre en question la pertinence et l'applicabilité des approches postpositivistes pour étudier les phénomènes humains. Eu égard à l'évaluation de programme, le débat est également né de l'utilisation, à ces débuts, des méthodes traditionnelles qui ont fait preuve de succès mitigés dans l'évaluation de la qualité et de l'efficacité des programmes et des politiques publics (Shaw, Greene et Mark, 2006). Ainsi cela marque-t-il une ouverture vers des approches plus qualitatives, voire mixtes (et leurs paradigmes associés), afin de comprendre les phénomènes humains. Cette période de changement est alors incarnée par la volonté de rendre légitimes les méthodes qualitatives et marque également l'expansion de l'utilisation des méthodes mixtes (Cook, 1997). Certains chercheurs (tels que Guba et Lincoln, par exemple) formés dans la tradition du postpositivisme et des méthodes quantitatives (particulièrement en éducation) ont alors « rejeté » leur formation qu'ils semblaient juger épistémologiquement inadéquate, et exprimés une préférence pour le qualitatif (Stake, 1998 ; Guba et Lincoln, 1981). De surcroit, grâce à l'influence de chercheurs positivistes tels que Weiss (1970, 1997), Cronbach (1982) et Cook (1997), à partir des années 1990, la majorité de la communauté évaluatrice de tradition positiviste reconnait les limites du statut de *Gold Standard* du devis expérimental en évaluation de programme et la contribution essentielle des méthodes qualitatives, notamment en ce qui a trait à la compréhension et à la description du contexte dans lequel est implanté le programme, à la description des processus (influence de la 2e génération), à l'utilisation des résultats d'évaluation, etc. (Dubois, Champagne et Bilodeau, 2011 ; Alkin et Christie, 2003).

Enfin, cette reconnaissance de l'utilité des méthodes qualitatives témoigne de leur légitimité, mais montre aussi que les deux méthodes (quantitative et qualitative) ne sont pas interchangeables, mais plutôt complémentaires (Cook, 1997 ; Shadish, 1995). À partir de cette

époque, et avec le mouvement de la 4e génération de Guba et Lincoln (1989), le quantitatif et le qualitatif sont deux axes qui coexistent dans le monde de l'évaluation ce a engendré de vifs débats entre 1980 et la fin des années 1990.

4e génération (1989 – aujourd'hui) : la négociation

L'ère du doute (1989-2004)

Jusqu'au début des années 1990, l'entreprise évaluative se déroule principalement dans la tradition postpositiviste, avec la mobilisation de méthodologies majoritairement quantitatives. Cependant, l'ensemble de la communauté de chercheurs dits quantitatifs reconnait l'importance du rôle des méthodologies qualitative dans la mise en œuvre d'une évaluation de programme. De plus, des chercheurs tels que Stufflebeam (2013, 1971) et Patton (2008, 1982) commençaient à percevoir les limites des méthodes évaluatives employées à l'époque, en mentionnant notamment qu'il serait approprié de travailler de concert avec les parties prenantes et d'orienter les approches évaluatives sur l'utilisation des résultats.

L'émergence de l'évaluation de 4e génération prend appui sur le caractère relatif des principes de l'évaluation et sur l'aspect contextuel et négocié de ses résultats (Dubois, Champagne et Bilodeau, 2011). Elle est apparue en réponse à trois grandes limites, que Guba et Lincoln (1989) ont identifiées au niveau des générations précédentes. La première limite concerne le désir d'objectivité des évaluateurs, puisque l'évaluation est trop souvent conçue par les gestionnaires de programme comme une activité qui dessert des stratégies politiques et non la population qui reçoit l'intervention. La seconde limite concerne l'hypothèse d'un jugement objectif élaboré dans le cadre d'une évaluation qui est en inadéquation avec la nature pluraliste des valeurs de notre société, des interventions sociales, et de leur système d'action. En effet, ces éléments contextuels exercent une influence directe sur l'élaboration du devis d'évaluation (questions d'évaluation, méthodes de collectes de données, etc.), sur le jugement opéré par l'évaluateur et sur les résultats de l'évaluation. Enfin, la troisième limite porte sur l'omniprésence des méthodes expérimentales rattachées à la tradition postpositiviste avec les mesures quantitatives et l'analyse des relations causales directes (entre une intervention et les résultats observés), qui engendre la mise à l'écart des éléments contextuels, sociopolitiques, culturels, etc. Ainsi, la non-utilisation des méthodologies qualitatives et la non-prise en compte des multiples éléments contextuels dans

lesquels le programme évolue diminueraient la pertinence et l'applicabilité des résultats évaluatifs, ces derniers étant trop éloignés de la réalité locale des décisionnaires et gestionnaires de programme (Fishman, 1992).

Au regard des limites qu'ils ont pu identifier, Guba et Lincoln (1989) mettent de l'avant que la pertinence d'une théorie et des procédures évaluatives mobilisées doit être négociées au niveau des parties prenantes (gestionnaires et bénéficiaires du programme). De plus, l'évaluateur doit être réceptif vis-à-vis de la perspective des parties prenantes en ce qui a trait au développement du devis d'évaluation. Cette caractéristique « collaborative » de la relation entre les évaluateurs et les parties prenantes, et l'importante place des approches postpositivistes, est d'ailleurs à l'origine des approches contemporaines, plus ou moins participatives, inscrites dans le paradigme pragmatiste, tentant d'allier rigueur scientifique et accompagnement des décisionnaires dans l'utilisation des résultats d'évaluation. Nous nous attardons davantage sur le mouvement pragmatiste dans la prochaine section qui porte sur les enjeux contemporains en évaluation de programme, ce dernier exerçant une influence importante sur les pratiques actuelles. Des enjeux qui s'inscrivent dans une nouvelle ère selon Stufflebeam et Coryn (2014), soit celle de l'expansion mondiale et multidisciplinaire.

En conclusion, la traversée historique vers la légitimation de l'évaluation de programme en tant que discipline à part entière fut marquée par de nombreux débats méthodologiques ainsi que de multiples tentatives de résolution de problèmes liés à la complexité du terrain (et de ses contraintes). Ce parcours vers l'évaluation telle que nous la pratiquons aujourd'hui a donc conduit à des changements significatifs dans la pratique des évaluateurs qui ont entrainés, notamment, une accumulation des rôles de l'évaluateur de programme (Tableau 1) :

Tableau 1. *Les rôles successivement attribués à l'évaluateur au fil de l'évolution de l'évaluation de programme*

	Rôle(s) de l'évaluateur	**Description**
1re génération : **La mesure**	**Technicien**	**En tant que technicien**, expert de l'instrumentation, l'évaluateur doit connaitre l'ensemble des instruments de mesure quantitative disponibles (en psychométrie, en statistique, etc.), tout en étant apte à développer de nouveaux dispositifs puisque toute variable concernée par une enquête scientifique peut être mesurable (Guba et Lincoln, 1989).

2e génération : **La description (naissance de l'évaluation de programme)**	Technicien **Descripteur**	**En tant que descripteur**, l'évaluateur doit non seulement développer des procédures de mesure de performance individuelle, il doit également recueillir les aspects descriptifs (qualitatifs) des composantes d'un programme, de sa mise en œuvre, de ses forces, de ses faiblesses et du degré d'atteinte de ses objectifs, afin de proposer des recommandations (Dubois, Champagne et Bilodeau, 2011).
3e génération : **Le jugement**	Technicien Descripteur **Juge**	**En tant que juge**, l'évaluateur a la responsabilité d'investiguer et d'utiliser toutes les informations qui permettent d'attester de l'efficience d'un programme. À titre d'expert, il a pour mission de mettre en œuvre des méthodes d'investigations rigoureuses afin de prononcer un jugement évaluatif crédible (Fishman, 1996 ; Guba et Lincoln, 1989).
4e génération : **L'évaluation de 4e génération**	Technicien Descripteur Juge **Modérateur**	**En tant que modérateur**, l'évaluateur doit inscrire sa pratique dans une approche plus ou moins participative afin de faciliter la mise en œuvre et l'utilisation des résultats de l'évaluation. Dans le cadre de la réalisation « collective » du processus évaluatif, il doit s'assurer de réunir toutes les conditions indispensables afin de faciliter le processus de collaboration, de négociation avec les parties prenantes. Ces dernières doivent pouvoir faire part de leurs questions et de leurs recommandations en ce qui a trait à l'élaboration du devis d'évaluation (Dubois, Champagne et Bilodeau, 2011).

Nous venons de voir que la transformation de la pratique et de la recherche en évaluation de programme s'est faite au gré de la complexification de l'activité évaluative, de la gestion des problématiques de terrain, des débats méthodologiques, et de la quête de la qualité et de l'utilité de l'évaluation. L'effort constant associé à la professionnalisation de l'évaluation de programme a conduit à la proposition et l'adoption de normes et lignes directrices qui visent à guider les pratiques et les prises de décisions lors de la conduite de projets d'évaluation (normes du JCSEE [2011, 1994, 1981] approuvées par les principales associations professionnelles nord-américaines, et adaptées/adoptées à l'échelle mondiale). Cependant, face à l'expansion multidisciplinaire et multisectorielle de l'évaluation et de son utilisation ainsi que l'accumulation des rôles de l'évaluateur, il est devenu impératif de développer un référentiel de compétences en évaluation de programme. Pour Stevahn, King, Ghere et Minnema (2005), il importe de définir les connaissances, aptitudes, et attitudes requises pour répondre efficacement aux normes, principes et lignes directrices de la profession, accroître la qualité et l'efficacité des efforts d'évaluation et

de formation en évaluation. Dans cet ordre d'idée, les auteurs publient une *taxonomie des compétences essentielles* (2001, révisée en 2005), qui organise 61 compétences selon six domaines ou volets de compétences :

1. *Pratique professionnelle* — volet de compétences qui se concentrent sur les normes et les valeurs fondamentales qui sous-tendent la pratique de l'évaluation, comme le respect des normes évaluatives et de l'éthique.
2. *Enquête systématique* — volet de compétences qui se concentrent sur les aspects plus techniques de la pratique de l'évaluation, tels que la conception, la collecte de données, l'analyse, l'interprétation et la production de rapports.
3. *Analyse situationnelle* — volet de compétences qui se concentrent sur l'analyse et la réponse aux intérêts, problèmes et circonstances contextuelles uniques d'évaluation donnée.
4. *Gestion de projet* — volet de compétences qui se concentrent sur les rouages de la conduite d'une évaluation, comme la budgétisation, la coordination des ressources et la supervision des procédures.
5. *Relations interpersonnelles* — volet qui se concentre sur les compétences interpersonnelles utilisées dans la conduite d'études d'évaluation, telles que la communication, la négociation, la gestion/résolution de conflit, la collaboration et les compétences interculturelles.
6. *Pratique réflexive* — volet de compétences qui se concentrent sur la prise de conscience de l'expertise en évaluation et des besoins en termes de croissance professionnelle (y compris la connaissance de soi en qu'évaluateur, l'identification de ses propres besoins pour améliorer sa pratique et l'engagement dans le développement professionnel vers cet objectif).

À présent, nous allons discuter des enjeux liés à la pratique et à la recherche en évaluation qui façonnent selon nous le monde de l'évaluation de programme actuel. Des enjeux ancrés dans les dimensions technique et sociopolitique de l'activité évaluative contemporaine, la poursuite de la quête de la qualité et de l'utilisation de l'évaluation, la persistance de débats méthodologiques, etc.

1.1.2 Les enjeux contemporains en évaluation de programme

Dans la continuité de l'histoire de l'évaluation de programme, nous discutons à présent de certaines problématiques actuelles en évaluation. Des problématiques en lien notamment avec l'héritage des générations passées et son influence sur les pratiques méthodologiques d'aujourd'hui.

Avant de débuter, il nous parait important de définir l'évaluation de programme et de décrire brièvement la démarche évaluative. En 2015, à la suite des travaux de recherche de Poth, Kay, Lamarche, Yapp, Sulla et Chisamore (2014) et de Lamarche et Poth (2013), la Société canadienne d'évaluation (SCÉ) a adopté la présente définition de l'évaluation dans le contexte canadien[1] : « l'évaluation est l'appréciation systématique de la conception, de la mise en œuvre ou des résultats d'une initiative pour des fins d'apprentissage ou de prise de décision. » La description des différents termes soulignés fait partie de la définition adoptée par la SCÉ[2]. Aussi, l'*appréciation* évaluative porte sur la valeur, le mérite, la portée, l'importance ou la qualité d'une *initiative* (un programme, un projet, un sous-programme, un sous-projet, une intervention ou leurs composantes). L'activité évaluative doit être *systématique* et doit être basée sur le recueil et l'analyse de données empiriques (à l'aide des méthodes de recherche propres aux sciences sociales), permettant la formulation de conclusions. Enfin, en ce qui a trait aux *finalités* d'une évaluation, celles-ci permettent à partir des résultats obtenus de former un jugement, de formuler des conclusions, de prendre des décisions relativement au programme, d'appuyer le développement organisationnel, etc. L'ultime but est de conduire à la formulation de recommandations pour appuyer la prise de décisions informées en ce qui a trait au futur du programme.

La démarche d'évaluation prescrit un certain nombre d'étapes incontournables pour la réalisation d'une évaluation de programme, et ce, quel que soit l'approche ou le modèle évaluatif utilisé (*Better Evaluation*, 2014; Hurteau et Houle, 2012; Stufflebeam, 2004; Darabi, 2002). Nous en proposons la représentation graphique suivante (Figure 2, p. 13) :

[1] Nous abordons en détail le processus de définition et d'adoption de cette définition dans le cadre conceptuel.

[2] La définition de l'évaluation et ces termes clés est disponible sur le site internet de la SCÉ : http://evaluationcanada.ca/fr/quest-ce-que-levaluation

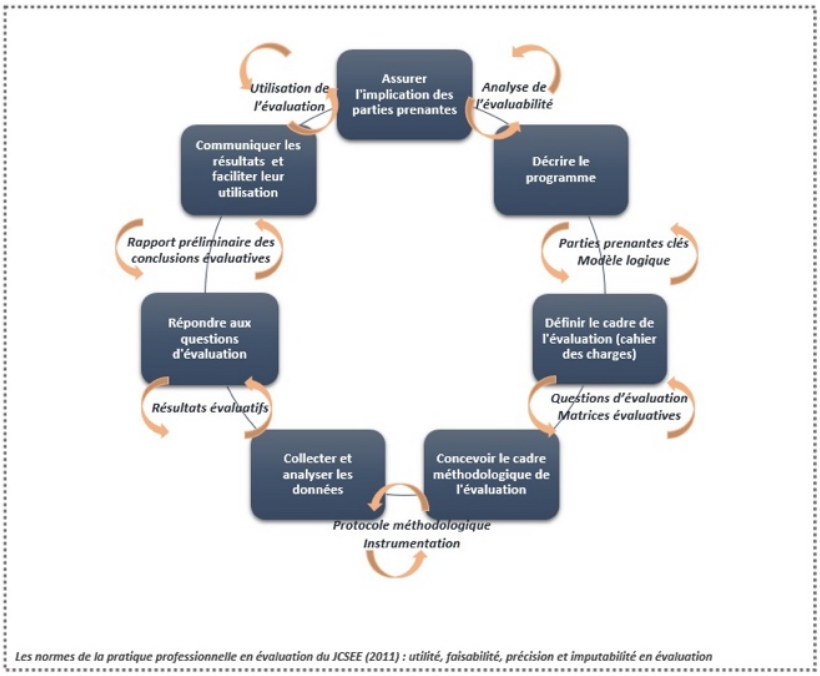


Figure 2. *Représentation graphique de la démarche d'évaluation de programme*

Veuillez noter que nous discutons plus en profondeur de la démarche d'évaluation et des processus décisionnels engagés au fil de sa mise en œuvre dans le cadre conceptuel (Chapitre 2). Ce qu'il faut retenir ici, c'est que la succession des étapes est itérative et que de nombreux éléments contextuels influencent le déroulement de l'évaluation, particulièrement au niveau de la méthodologie. C'est ce dont nous allons discuter à présent.

1.1.2.1 L'influence du mouvement pragmatiste sur les pratiques actuelles

L'un des changements notables dans l'histoire de l'évaluation de programme fut la recherche indispensable de compromis entre les deux traditions scientifiques, les méthodes quantitatives et les méthodes qualitatives. En effet, les postpositivistes ont assoupli le devis expérimental, en développant le devis quasi expérimental et en reconnaissant l'utilité des méthodes qualitatives, et par extension du recours aux méthodes mixtes. Les constructivistes ont proposé quant à eux d'impliquer les parties prenantes dans le processus évaluatif afin de s'assurer que celui-ci répond bien à leurs besoins ainsi qu'à leurs attentes. En outre, un tel rapprochement entre les différentes traditions scientifiques (épistémologiques et méthodologiques) stimule l'apparition d'approches plus ou moins participatives, approches qui s'inscrivent dans le paradigme

pragmatiste (Annexe 1[3]). Le mouvement pragmatiste représente alors l'alliance entre la rigueur méthodologique et la quête de scientificité des résultats du processus évaluatif (postpositivisme) ainsi que la nécessité d'impliquer les parties prenantes dans le processus évaluatif afin de faciliter le déroulement de l'évaluation et l'utilisation des résultats (constructivisme). Il représente selon nous une influence majeure sur la manière dont nous pratiquons l'évaluation de programme aujourd'hui. Aussi, afin de rendre compte de l'histoire de l'évaluation et de l'évolution des approches évaluatives, Alkin et Christie (2013) ont proposé une modélisation arborescente (*Evaluation Theory Tree*, Figure 3), qui permet de visualiser l'évolution et les contributions des différents courants (paradigmes) :

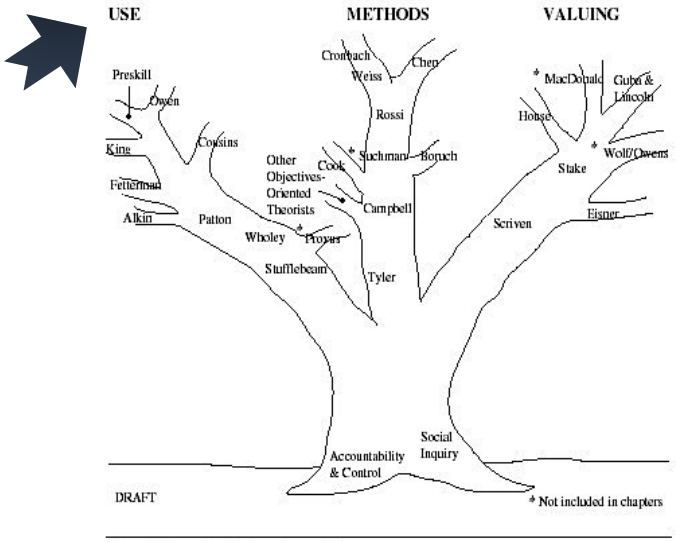


Figure 3. *Evaluation Theory Tree (Alkin et Christie, 2013, p.14)*

Sur cet arbre nous retrouvons la branche « méthodologie » (paradigme postpositiviste), la branche « jugement/valeur » (paradigme constructiviste), et enfin la branche qui nous intéresse,

[3] Vous trouverez en Annexe 1 un tableau dans le cadre duquel nous présentons quelques théoriciens important du courant pragmatiste en évaluation (majoritairement de la branche de l'utilisation) et nous offrons une brève description de certaines approches évaluatives contemporaines, inscrites dans une perspective plus ou moins participative (l'évaluation axée sur l'utilisation [Patton, 2008, 2005, 1982], l'évaluation de type *Empowerment* [Fetterman, 2005, 1994], et l'évaluation participative [Cousins et Whitmore, 1998]).

celle de l'« utilisation » (paradigme pragmatiste), une branche qui tente d'allier les premières traditions scientifiques (postpositivisme/quantitatif et constructiviste/qualitatif).

Les théoriciens pragmatistes de l'évaluation sont apparus à partir de la 3e génération (courant du jugement). Mertens (2015) nous explique qu'à l'époque, les évaluateurs (particulièrement au niveau fédéral) ne pensaient pas forcément aux destinataires des rapports d'évaluation qui, la plupart du temps, étaient mis de côté par les gestionnaires des programmes évalués. C'est ainsi que des chercheurs tels que Stufflebeam (1971), Weiss (1997, 1970) et Patton (2008, 1982) ont jugé qu'il était temps de prendre conscience de l'utilité de l'évaluation et de l'applicabilité des résultats par les destinataires de l'évaluation, et de s'assurer d'accompagner au mieux l'ensemble des parties prenantes (bailleurs de fonds, gestionnaires, intervenants, etc.) dans les prises de décisions relatives aux résultats et aux recommandations présentées par les évaluateurs. De plus, nous avons pu voir avec le mouvement de la *Great Society Era* que les logiques et les précisions propres à l'expérimentation en sciences sociales étaient alors inadaptées au caractère « désordonné » et dynamique des politiques et des programmes publics (Greene, 2005), et que, de ce fait, les résultats de ces évaluations apparaissaient peu défendables ou peu utiles auprès des parties prenantes.

Selon Patton (2018, 2008), les évaluateurs peuvent et doivent alors travailler avec les gestionnaires de projet afin de définir, mettre en place et exécuter les évaluations, combinant ainsi le meilleur de ce que chacune des parties peut apporter à la relation. En effet, l'implication des gestionnaires de projet (et autres parties prenantes clés) facilite le déroulement de l'évaluation en accompagnant l'évaluateur dans la compréhension de la théorie du programme, en favorisant l'accès au terrain (aux sources d'information). De plus, les impliquer dans le processus évaluatif c'est les engager à devenir « ambassadeurs » de l'évaluation et ainsi promouvoir l'utilisation des résultats d'évaluation ainsi que l'évaluation de programme dans les organisations. En outre, les dynamiques relationnelles, plus précisément le degré de collaboration entre les évaluateurs et les parties prenantes sont déterminées en fonction de ce que les évaluateurs considèrent comme approprié au regard du cas étudié (caractéristiques du programme évalué, objets d'évaluations, *evaluand*, etc.) et de ses besoins pour atteindre les objectifs de l'évaluation. En ce qui a trait à la méthodologie, chaque méthode ou source de données doit être spécifiquement associée aux questions et aux objectifs de l'évaluation. Le courant pragmatiste préconise le recours aux

méthodologies mixtes (conciliation entre différentes traditions : le courant postpositiviste avec les approches quantitatives et le courant constructiviste avec les approches qualitatives), et ce, afin d'assurer la scientificité et l'utilisation des résultats de l'évaluation.

L'évaluation contemporaine s'inscrit dans l'Ère de l'expansion mondiale et multidisciplinaire (Stufflebeam et Coryn, 2014). Une ère où de nombreuses organisations internationales telles que la Banque Mondiale commencent à écarter les méthodes évaluatives expérimentales et économétriques qu'elles ont historiquement privilégiées, et ce, au profit des approches évaluatives pragmatistes (évaluation basée sur la théorie, évaluation participative, etc.). C'est aussi une ère dont les caractéristiques reflètent nos observations sur l'évolution de la discipline et le caractère perméable des frontières entre les différents courants paradigmatiques (Mertens, 2015). On prend ici l'exemple de l'évolution de l'évaluation de programmes publics et du lien étroit qu'il existe aujourd'hui entre l'évaluation, la mesure de la performance et la gestion axée sur les données probantes. Nous avons vu qu'à partir des années 1970, la politisation de l'évaluation a conduit à l'abandon de la conception de l'évaluation inspirée de la recherche scientifique. L'évaluation se voit dès lors attribuer la fonction d'aide à la prise de décision au sein de l'administration publique de nombreux pays comme le Canada et les États-Unis. L'évaluation est utilisée comme une source d'information pour l'affectation des ressources, l'amélioration des programmes et l'imputabilité des gestionnaires (Secrétariat du Conseil du Trésor du Canada, 1981). Des théoriciens pragmatistes de la branche de la méthode comme Weiss (1994, 1970) se sont alors penchés sur les méthodes d'évaluation de l'efficacité des programmes publics et l'amélioration de l'utilisation des méthodes de recherche propres aux sciences sociales, en évaluation. Pour Weiss (1994, 1970), le devis expérimental n'est pas le plus approprié en évaluation, et les mesures d'impacts ne sont pas les plus pertinentes pour vérifier et renforcer l'efficacité des services et programmes sociaux. La complexité du contexte évaluatif freine la possibilité pour l'évaluateur de garantir une parfaite objectivité, aussi le recours à de multiples méthodes de collectes de données, quantitatives et qualitatives, est le compromis idéal pour générer des résultats à la fois rigoureux et utiles pour le programme. Au cours des 30 dernières années, la fonction de l'évaluation en administration publique a beaucoup évolué, et elle est aujourd'hui intrinsèquement reliée à la gestion fondée sur les données probantes (mesure de la performance), qui est devenue une caractéristique importante de l'administration des gouvernements dans les pays occidentaux (Campbell, Benita, Coates, Davies et Penn, 2007). Au Canada, une mise à jour

majeure de la politique d'évaluation du gouvernement fédéral a été annoncée en 2009. Le principal élément de cette politique est l'exigence pour les ministères et les organismes fédéraux d'évaluer tous leurs programmes sur un cycle de cinq ans (Secrétariat du Conseil du trésor du Canada, 2009). Cette évaluation des programmes est explicitement liée à l'évaluation du rendement, soit l'évaluation de l'efficacité, de l'efficience et de l'économie. Pour la première fois, la fonction de mesure du rendement dans tous les ministères et les organismes, qui était auparavant une activité de gestion distincte, est maintenant liée à la fonction d'évaluation (McDavid, Huse et Hawthorn, 2019). À l'image des politiques évaluatives gouvernementales d'autres pays comme l'Angleterre et les États-Unis, la politique canadienne accorde un rôle plus important à la fonction de l'évaluation à l'appui du système de gestion des dépenses (Secrétariat du Conseil du trésor du Canada, 2015). Il s'agit de renforcer la qualité et l'utilisation des données d'évaluation afin de déterminer si les dépenses de programmes sont efficaces et correspondent à une utilisation optimale des fonds publics. L'évaluation a pour objet de répondre au besoin de preuves neutres et dignes de foi sur l'optimisation des dépenses, et elle cible « la collecte et l'analyse systématique de données probantes sur les résultats du programme afin d'en évaluer la pertinence et le rendement et de trouver d'autres modes de prestation ou d'autres façons d'obtenir les mêmes résultats. » (Secrétariat du Conseil du trésor du Canada, 2009).

Ainsi, les attentes associées au mandat fédéral (évaluation fondée sur les données probantes) réactivent des débats de longue date au sein de la communauté évaluative en ce qui a trait aux méthodes appropriées pour inférer le lien de causalité entre le programme et les résultats observés et le maintien du devis expérimental, voire quasi-expérimental comme le *gold standard* de l'évaluation de l'efficacité (McDavid, Huse et Hawthorn, 2019 ; Christie et Fleischer, 2010 ; Cook, Scriven, Coryn et Evergreen, 2010 ; Donaldson et Christie, 2005). Les questions liées à l'efficacité invitent à recourir à des devis méthodologiques sophistiqués qui répondent aux normes pour évaluer les relations de cause à effet. Or, la réalité de terrain et les contraintes qui l'occupent (état de l'*evaluand*, ressources insuffisantes, problèmes important d'accès aux données, etc.) posent de gros problèmes de faisabilité qui freinent la possibilité de produire des résultats valides, fiables et représentatifs de l'objet d'évaluation (Bamberger et Mabry, 2020). Une approche pragmatiste est donc recommandée, soit le recours à un devis d'évaluation qui : 1) mobilise une approche de méthodes multiples; 2) mise sur la complémentarité des méthodes quantitatives et qualitatives; 3) s'appuie sur le jugement de l'évaluateur jugement pour déterminer les relations de

causalité et 4) assure un processus rigoureux et systématique de triangulation (méthodes, sources, évaluateurs) (Bamberger et Mabry, 2020; McDavid, Huse et Hawthorn, 2019; Greene, 2015, 2005c; Stufflebeam et Coryn, 2014).

Les évaluations de programme sont souvent menées dans des conditions où les données appropriées pour déterminer la question de l'attribution du lien de causalité entre le programme et les résultats observés sont difficiles à obtenir (Bamberger et Mabry, 2020). En l'absence d'une base de données sur la mesure du rendement du programme, ou lorsque celle-ci présente un problème de justesse ou d'exhaustivité, d'autres données provenant de la documentation de programme sont généralement disponibles. Cependant, un examen plus approfondi de ces données indique souvent qu'elles mesurent des caractéristiques du programme qui ne sont que partiellement alignées avec les questions clés que doit traiter l'évaluation (McDavid, Huse et Hawthorn, 2019). Les efforts nécessaires pour combler ces lacunes nécessitent généralement de revoir la distribution des intrants, et ils peuvent être très coûteux selon les circonstances (c'est-à-dire la complexité/envergure des programmes gouvernementaux, nature et ampleur des données à collecter, etc.). Nous verrons un peu plus loin que la perspective de reddition de compte associée aux exigences de la politique évaluative semble avoir l'ascendance sur la faisabilité et (ou) l'utilité de l'évaluation. En effet, les commanditaires de l'évaluation semblent se voir dans l'obligation de répondre à des impératifs (plan d'évaluation quinquennale, renouvellement de financements, etc.) qui réduisent la marge de manœuvre des évaluateurs (échéancier très serré, budget limité, manque de maturité du programme, etc.). On semble être davantage préoccupé par le respect du rituel et des exigences de l'activité évaluative que par les limites sur le plan de la rigueur et de la validité des évaluations produites (Lamari, Poulin-Larivière et Lemieux, 2016). À l'instar de l'observation de McDavid, Huse et Hawthorn (2019), Dumaine (2018) remarque que la mesure des résultats ultimes, de l'impact, qui porte sur l'attribution chiffrée de la causalité (devis expérimentaux ou quasi expérimentaux) est un exercice difficile, très peu réalisé au gouvernement fédéral, particulièrement lors d'une évaluation. Pour l'auteur, si une étude de l'impact doit/peut être entreprise, elle devrait se faire en marge du processus d'évaluation quinquennale des programmes fédéraux. Les résultats d'une telle étude serviraient alors de sources de données lors de l'évaluation du programme.

La diversification des contextes de pratiques, des approches et des tâches évaluatives ainsi que la dimension sociopolitique de l'évaluation représentent pour nous des enjeux importants pour les pratiques actuelles en évaluation de programme, enjeux qui peuvent avoir un impact déterminant sur la qualité des procédures évaluatives.

1.1.2.2 La dimension « méthodologique » au cœur des préoccupations actuelles

Eu égard à l'histoire de l'évaluation de programme, de la recherche en sciences sociales et de leurs vigoureux débats, nous remarquons que des préoccupations fondamentales perdurent et façonnent l'évolution des générations en évaluation et en recherche. En effet, pendant près de 70 ans, l'évolution de l'évaluation de programme a été rythmée par les travaux de chercheurs et d'évaluateurs en quête de qualité et d'utilité du processus évaluatif. Une quête aux prises, notamment, avec les contraintes de terrain, les débats épistémologiques et méthodologiques, la nécessité croissante de collaborer avec les parties prenantes, la multiplication des rôles et des tâches de l'évaluateur (expert méthodologique, gestionnaire de projet, médiateur, éducateur, etc.). La complexité actuelle de la pratique et de la recherche en évaluation de programme est intrinsèquement liée aux facteurs contextuels et à l'éventail d'activités qui sont propres à la pratique de terrain (Smith et Brandon, 2008). En effet, de nombreux problèmes, souvent perçus comme insolubles, se manifestent de manière récurrente dans la pratique des évaluateurs : les difficultés de communication avec les clients, certains dilemmes éthiques, la formation des nouveaux évaluateurs, la collaboration avec les services gouvernementaux, les difficultés méthodologiques et les influences potentielles sur la réalisation d'une évaluation (contraintes budgétaires, temporelles, politiques, etc.), etc. Aussi, lorsqu'on désire étudier certains de ces problèmes fondamentaux, il est important de se questionner sur la façon dont les facteurs constitutifs de ces problèmes sont enchevêtrés au cœur des quatre dimensions de l'univers de l'évaluation de programme (Smith et Brandon, 2008 ; Association Amérique en Évaluation [AEA], 2004) : (1) la théorie (2) la méthodologie (3) la pratique et (4) la profession. Ainsi, la *théorie* fait référence à la réflexion des praticiens sur le comment et le pourquoi ils s'engagent dans la mise en place d'une évaluation (ex. le but de l'évaluation, est-ce que c'est à des fins de redditions de compte, de monitorage ou de développement ? Est-ce que l'évaluation vise la production de connaissances, le service à la clientèle, la réforme sociale ou le contrôle politique ?). La *méthodologie* concerne les outils et les ressources méthodologiques que les praticiens mobilisent

afin de mener à bien leurs évaluations (ex. les outils d'enquête, de gestion des ressources, de collaboration, etc.). La *pratique* rappelle, quant à elle, le monde immédiat de la politique, des clients, des ressources, de l'ambigüité des rôles et de l'évolution des conditions sur le terrain, et l'intérêt pratique de l'accomplissement d'un travail bien fait et le souci de faire la différence. Enfin la *profession*[4] porte sur l'évaluation en tant que guilde socialement définie de praticiens et collègues, et la définition de leur travail, de leur éthique, de leurs responsabilités, ainsi que leurs moyens de subsistance.

Au fil des générations, de nombreux modèles et approches ont été développés en réponse à ces différentes préoccupations afin d'accompagner au mieux les organismes publics et à but non lucratif dans l'amélioration, l'évaluation et la mesure de la performance de leurs programmes, et ce, en favorisant l'utilisation des résultats et des recommandations produits par les évaluateurs. Malgré tout, les enjeux relatifs au contrôle de la qualité du processus évaluatif, à la sélection et à la formation des évaluateurs, aux normes et à l'éthique du travail évaluatif ainsi qu'à l'utilisation des résultats par les parties prenantes, etc. perdurent (Hatry, Newcomer, Wholey, 2015). En effet, comme nous avons pu le voir plus tôt, les évaluateurs se trouvent périodiquement confrontés à ces différents problèmes, problèmes qu'ils parviennent à régler momentanément. À chaque fois, le processus de résolution de ces problèmes reflète les valeurs, les considérations techniques, les préoccupations, les ressources disponibles, les technologies émergentes, etc. contemporaines (de l'époque). Il est donc essentiel de reconnaitre, et d'accepter qu'il demeure utopique de vouloir résoudre ces problématiques importantes de manières définitives. Toutes nouvelles procédures de gestion de ces problèmes sont inéluctablement temporaires, contextuelles, situationnelles et conditionnelles (Smith, 2008). Par conséquent, l'évaluation de programme présente de nombreux défis selon le contexte dans lequel elle est réalisée, ce qui rend d'autant plus importants le questionnement et la compréhension de l'influence des facteurs problématiques récurrents sur la

[4] En 2004, l'AEA a organisé sa conférence annuelle sur le thème des enjeux fondamentaux en évaluation de programme. Pour être en mesure de saisir pleinement ces problématiques et leurs influences sur l'activité évaluative, le président de l'association (Nick Smith) et le comité d'organisation composé d'experts de la discipline, ont découpé l'univers de l'évaluation en quatre dimensions imbriquées.

dimension *méthodologie* (et par extension, la dimension *pratique*) de l'activité évaluative et les processus décisionnels des évaluateurs (de la conception du plan d'évaluation à la présentation des résultats évaluatifs).

Enfin, rappelons que la complexification des pratiques évaluatives a contribué à l'accumulation des rôles et des tâches des évaluateurs afin qu'ils puissent s'adapter et répondre au contexte évaluatif. En effet, comme Beaudoin, Lefrançois et Ouellet (1986) l'indiquaient déjà à l'époque, dans le cadre de sa pratique, « l'évaluateur met toujours en jeu sa compétence à analyser avec rigueur, cohérence et justesse une action ou un projet, étant donné les contraintes inhérentes au milieu, c'est-à-dire les attitudes parfois réfractaires du client ou encore les conditions défavorables (délai imposé, limite budgétaire) dans l'exercice de son mandat. » (p.192). Aujourd'hui, l'évaluateur (ou l'équipe d'évaluation) doit donc mettre à profit un nombre important de compétences très diverses en gestion de projet, en méthodologie évaluative et en méthodes d'investigation scientifiques (p. ex. méthodes quantitatives, qualitatives, mixtes), en éthique et déontologie, en analyse de pratique, etc. (Société canadienne d'évaluation, 2018, 2014 ; Centre d'excellence du Secrétariat du trésor du Canada, 2010; Stevahn, King, Ghere, Minnema, 2006, 2005). Selon nous, un tel constat fait émerger un autre enjeu actuel (récurrent) en évaluation : la sélection et la formation des évaluateurs de programme. En effet, trouver des professionnels qualifiés et formés en évaluation constitue un prérequis important pour assurer la qualité d'une évaluation (Hatry, Newcomer et Wholey, 2015). Or, de nouveaux évaluateurs entrent constamment sur le terrain, soit à la suite de l'obtention d'un diplôme universitaire en évaluation de programme, soit *« accidentellement »* dans le cadre de leur poste en assumant des responsabilités d'évaluation avec peu (ou pas) de préparation (Hatry, Newcomer et Wholey, 2015; Ghere, King, Stevahn et Minnema, 2006); et très peu (ou pas) de connaissances relatives à l'exercice de cette discipline ainsi qu'aux standards de la profession (King, Stevahn, Ghere et Minnema, 2001). Le profil de l'évaluateur semble donc constituer un facteur contextuel important dans le cadre de l'évaluation et la rigueur méthodologique.

À la lumière des éléments de problématiques exposés jusqu'à présent, l'évaluation de programme se trouve confrontée à de nombreux enjeux relatifs à la praxis, enjeux qui peuvent avoir une influence déterminante sur la dimension « *méthodologie* » de la pratique évaluative. Aussi cela nous conduit-il à nous interroger sur les facteurs qui peuvent exercer une influence sur

les choix dits méthodologiques des évaluateurs, de la formulation des questions d'évaluation à la présentation des résultats évaluatifs. En effet, le processus décisionnel relatif à la méthodologie évaluative semble être un élément crucial, mais aussi un point de tension, dans la pratique des évaluateurs de programme. De plus, la complexification de la pratique évaluative, l'accumulation des rôles et des responsabilités des évaluateurs ainsi que la présence sur le terrain d'évaluateurs qui n'ont pas nécessairement une formation en évaluation soulèvent des questions quant aux capacités d'adaptation des évaluateurs face à un environnement évaluatif donné, et ce, en fonction de leur niveau d'expertise. En effet, de nombreux auteurs s'accordent sur le fait que la qualité de l'évaluation est intrinsèquement reliée à la capacité de l'évaluateur à mobiliser des savoirs pratiques en situation réelle, son expertise, ses valeurs et son jugement afin d'opérer les choix méthodologiques les plus appropriés à la lumière des caractéristiques de l'environnement évaluatif, des contraintes inhérentes au terrain (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019; Bamberger, Rugh et Mabry, 2012; Greene, Lipsey, Schwandt, Smith et Tharp, 2007; Greene, 2006; King, 2003).

L'objectif principal de cette thèse doctorale est de décrire les pratiques des évaluateurs de programme (novices et experts) et plus particulièrement leurs choix méthodologiques et la potentielle influence des contraintes situationnelles sur leurs opérations d'évaluation.

1.2 Contexte spécifique

En 2013, l'année 2015 fut déclarée l'année internationale de l'évaluation (*EvalYear2015*) par EvalPartenaires[5], un mouvement mondial visant à renforcer les capacités nationales en évaluation. Cette décision fut renforcée par l'adoption de la résolution A/RES/69/237 « Renforcement des capacités en vue de l'évaluation des activités de développement au niveau des pays » durant l'Assemblée générale des Nations Unies, le 14 décembre 2014. Des organismes tels que le Groupe d'évaluation des Nations Unies (UNEG) ont alors joint le mouvement, et le groupe

[5] Le lecteur peut consulter la liste des membres EvalPartenaires à l'adresse suivante : https://www.evalpartners.org

EvalPartenaires entreprit un processus de consultation mondial invitant les parties prenantes à réfléchir aux priorités et aux dimensions clés du programme mondial pour l'évaluation 2016-2020 (*EvalAgenda2020*). Ce projet de cinq ans visait à aider l'évaluation de programme à atteindre son plein potentiel et être reconnu pour sa contribution considérable dans l'amélioration de la société. Le processus de consultation se poursuivit durant les 92 évènements nationaux, régionaux et internationaux d'*EvalYear2015*, où les membres de la communauté évaluatrice ont été appelés à faire part de leurs suggestions pour le programme. Durant le Forum international en évaluation (novembre 2015), les parties prenantes clés ont alors finalisé le guide d'implantation du Programme mondial d'évaluation 2016-2020.

En signant la déclaration de soutien au Programme mondial *EvalAgenda2020*[6], les signataires (organismes gouvernementaux et non gouvernementaux, sociétés régionales, nationales et internationales d'évaluation, etc.) reconnaissent que l'évaluation peut contribuer à assurer que les politiques publiques, les programmes et les processus décisionnels s'appuient sur des données probantes et engendrent des résultats efficaces afin d'améliorer la qualité de vie de l'ensemble de la population. En outre, ils reconnaissent qu'aujourd'hui il importe de réduire l'écart entre la valeur potentielle de l'évaluation et le manque actuel de reconnaissance de son utilité, et ce, en redoublant d'efforts pour améliorer la qualité et l'utilité des pratiques évaluatives, et faire reconnaitre ses avantages dans la société (EvalPartenaires, 2016). Ainsi, le programme fait état d'un certain nombre de priorités en évaluation sur cinq ans, priorités visant notamment le renforcement d'un environnement favorable à l'évaluation, le développement des capacités institutionnelles et le renforcement des capacités individuelles en évaluation.

Nous prenons le temps de discuter du mouvement mondial pour l'évaluation depuis ces trois dernières années, car parmi les premiers signataires se trouvaient, entre autres, la Société canadienne d'évaluation (SCÉ), la Société québécoise d'évaluation de programme (SQEP), le Réseau francophone de l'évaluation (REF), le bureau indépendant d'évaluation des Nations Unies

[6] Le lecteur peut consulter l'EvalAgenda2020 à l'adresse suivante : https://www.evalpartners.org/global-evaluation-agenda

(UNDP) pour lequel beaucoup d'évaluateurs accrédités du Canada sont agissent comme consultants. Cette mobilisation globale autour de l'*EvalAgenda2020* fait, selon nous, partie intégrante du contexte spécifique dans lequel se déroule cette recherche doctorale en évaluation dont les participants sont membres des sociétés professionnelles nationales, régionales et internationales signataires. De surcroit, notre volonté de décrire les pratiques méthodologiques des évaluateurs de programme (novices et experts) et les facteurs contextuels pouvant influencer ces dernières s'aligne sur un certain nombre de priorités du Programme mondial pour l'évaluation 2016-2020, notamment en ce qui a trait au renforcement de la qualité des évaluations et de la rigueur des pratiques méthodologiques, au soutien de la recherche en évaluation ainsi qu'à la formation des évaluateurs de programme.

La présentation du contexte général nous a permis d'illustrer la complexité de la pratique de l'évaluation de programme et d'introduire différents facteurs contextuels qui peuvent avoir une influence sur le développement et la mise en œuvre de l'évaluation ainsi que sur les choix méthodologiques des évaluateurs (implication des parties prenantes, contraintes politico-organisationnelles, ressources disponibles pour l'évaluation incluant l'accessibilité des données probantes, profil des évaluateurs, etc.). À présent, nous abordons de manière plus spécifique le processus d'opérationnalisation des choix méthodologiques de l'évaluateur de programme et l'influence des facteurs contextuels sur lesdits choix. Nous discutons également des caractéristiques de la pratique des évaluateurs selon leur profil et leur niveau d'expertise, et de l'incidence qu'elles peuvent avoir sur leurs capacités d'adaptation à l'environnement évaluatif. Il est important de garder à l'esprit que cette partie constitue un aperçu des éléments clés qui seront développés de manière plus précise dans le cadre conceptuel.

1.2.1 L'influence des facteurs contextuels sur la méthodologie évaluative

Comme l'expliquent Hurteau et Houle (2012), si aujourd'hui il existe des approches multiples et variées (en fonction du paradigme, du contexte, du programme, de l'organisme évaluatif, etc.), certaines étapes demeurent essentielles pour planifier et mener à terme une évaluation. L'une de ces étapes est la conception et la mise en œuvre du devis d'évaluation, un élément qui doit rendre compte de la démarche évaluative, avec clarté, précision et transparence (Rog, 2005). Il est donc nécessaire qu'il décrive de manière rigoureuse les raisons d'être, la portée et la méthodologie (les sources d'information, les variables et les concepts mesurés qui devraient

être déterminés à l'aide de la théorie du programme, les procédures de collecte et d'analyse de données, ainsi que les limites inhérentes) de l'évaluation. Ainsi, le devis d'évaluation englobe le cadre de l'évaluation (le cahier des charges ou mandat) et le cadre méthodologique. Il est important de noter que la succession des étapes de la démarche évaluative n'est pas figée, elle est itérative, car l'évaluateur doit constamment réajuster ses choix méthodologiques, afin de s'adapter aux conditions de terrain.

De nombreux évaluateurs s'accordent à dire que le contexte évaluatif (et tout ce qui le constitue) est à la fois important pour les objets de l'évaluation et pour le déroulement et l'efficacité du travail de l'évaluateur (Greene, 2005a). En effet, les conditions de terrain, la nécessité de s'adapter au contexte politique et organisationnel et la gestion des contraintes (les problèmes de temps, de budget ou encore d'accès aux données) conduisent très souvent les évaluateurs à amender le devis d'évaluation en cours de route, à improviser pendant la collecte de données, etc., et ce, afin de s'assurer que les résultats sont crédibles et défendables (Rossi, Lipsey et Freeman, 2019). De plus, de récents courants tels que l'évaluation en situation réelle (Bamberger et Mabry, 2020 ; Bamberger, Rugh et Mabry, 2012) proposent d'accompagner les évaluateurs dans leur pratique, mais l'écart entre les contraintes inhérentes au terrain et la théorie semble persister. L'approche de l'évaluation en situation réelle propose des stratégies pour gérer le plus efficacement possible les contraintes inhérentes au terrain (ou contraintes ESR) et l'utilisation et la réalisation de méthodologies évaluatives adaptées et rigoureuses. Si à l'origine elle était destinée à la réalisation d'évaluation d'impact dans les pays en développement, ses auteurs soulignent sa pertinence pour tout type de mandat évaluatif, et ce, peu importe l'environnement socio-économique et politique, du programme évalué (Bamberger et Mabry, 2020; Bamberger, Rugh et Mabry, 2012 ; Bamberger, Rugh, Church et Fort, 2004). En effet, un bon nombre des situations évaluatives traitées dans le cadre de cette approche ESR s'avèrent courantes et communes pour l'ensemble des milieux évaluatifs, par exemple lorsque : (1) l'évaluation arrive tardivement dans le cycle de vie du programme et qu'aucune base de données comparative n'est disponible ; (2) le temps et le budget alloués à l'évaluation sont inappropriés ; (3) l'accès à des données de base est limité ou les données probantes existantes sont incomplètes ou manquent de qualité ; (4) les pressions politiques liées à la manière dont l'évaluation devrait être réalisée et ce que les résultats devraient dire ; ou encore (5) la résistance de parties prenantes face à l'évaluation qu'ils perçoivent comme une menace. La définition de la méthodologie de collecte de l'information repose sur un

certain compromis nécessaire entre les exigences/attentes/besoins des utilisateurs ciblés et la pertinence des méthodes (sources de données) sur le plan de la faisabilité et de la rentabilité (utilité réelle, rapport coût-bénéfice), et ce, tout en veillant à la fiabilité et la représentativité des informations collectées. Ultimement, pour les auteurs, les plus gros défis proviennent de la nécessité de tenir compte (et de satisfaire) des différentes attentes et exigences organisationnelles, dans un système administratif souvent rigide et inadapté pour accommoder les stratégies méthodologiques envisagées en réponse aux contraintes ESR (Bamberger, Rugh, Church et Fort, 2004). L'évaluateur fait également souvent face à de fortes pressions méthodologiques, notamment parce que les parties prenantes clés peuvent avoir des idées bien arrêtées sur ce qui est « juste » en ce qui a trait aux ressources nécessaires, aux méthodes d'investigation, aux objets d'étude, aux sources d'information, à la communication des résultats évaluatifs. Bamberger et Mabry (2020) font d'ailleurs remarquer que les jeunes évaluateurs découvrent très rapidement que les questions relatives aux choix méthodologiques (le recours à un devis expérimental ou non, le recours à un devis méthodologique qualitatif, quantitatif ou mixte, la sélection des sources d'information, la sélection des questions d'entrevues, etc.) peuvent provoquer de vifs débats.

En dépit des contraintes inhérentes à la complexité du contexte évaluatif, on observe une augmentation des évaluations systématiques d'impact (mesure de performance) commanditées par les agences de financements, les organismes gouvernementaux, les organismes sociaux et à but non lucratif ainsi que les bénéficiaires des programmes. On peut citer, par exemple, la Politique sur l'évaluation du Secrétariat du Conseil du trésor du Canada de 2009, aujourd'hui la Politique sur les résultats (2016), qui exige, entre autres, l'évaluation quinquennale, de la pertinence et de la performance (efficacité, efficience et économie) des programmes ministériels de subventions et de contributions. En outre, si l'institutionnalisation des pratiques évaluatives constitue une bonne chose sur le plan de la transparence (reddition de compte), du renforcement des capacités évaluatives des organisations, de l'allocation de ressources pour l'évaluation, etc., cela vient également avec son lot de limites notamment en ce qui concerne la qualité méthodologique et la crédibilité de l'activité d'évaluation. Parmi ces éventuels problèmes se trouvent : (1) *l'utilisation de conformité du processus évaluatif*, c'est-à-dire entreprendre une évaluation avec pour seule motivation l'obligation de respecter une exigence (Patton, 2015a). Les activités d'évaluation peuvent alors devenir mécaniques et consister en l'application d'une méthodologie généralement prescrite, presque routinière, et pas nécessairement pertinente. Ceci complique aussi l'accès aux

données puisque les évaluateurs doivent anticiper le fait qu'une source soit moins encline à participer faute de temps, voire d'épuisement face à la perpétuelle sollicitation évaluative ; et (2) *l'utilisation symbolique de l'évaluation*, c'est-à-dire mandater une évaluation pour répondre à une obligation sans réelles intentions d'utiliser les résultats et de s'impliquer sérieusement dans l'activité évaluative, ou encore d'utiliser l'évaluation pour justifier des décisions déjà prises (Taut et Brauns, 2003). Nous verrons un peu plus loin que ces formes d'utilisation sont de potentielles sources de mauvaise évaluation et mauvaise utilisation.

Enfin, la promotion de bonnes pratiques évaluatives dans un environnement aussi complexe, nécessite développer une collaboration étroite et compréhensive avec les utilisateurs ciblés (Patton, 2018 ; Taut, 2008). Ces dernières occupent donc aujourd'hui une place centrale au sein de l'activité évaluative, et ce, afin de promouvoir l'utilisation du processus et des résultats d'évaluation[7], de favoriser la faisabilité, la validité et la crédibilité de la démarche évaluative (compréhension du programme et de l'environnement évaluatif, anticipation des contraintes, accès aux données, formulation de recommandations utiles et utilisables, etc.) et de contribuer au renforcement des capacités évaluatives des organisations (Bamberger et Mabry, 2020 ; Patton, 2018 ; Alkin et King, 2016 ; Chen, 2015; Bryson, Patton et Bowman, 2011). En outre, la rigueur de la démarche évaluative, sa qualité méthodologique et l'utilité de ses résultats reposent sur un certain nombre de facteurs en lien avec l'implication des utilisateurs ciblés de l'évaluation, tels que l'attitude des parties prenantes envers l'évaluation (p. ex. leurs expériences antérieures, leurs attentes méthodologiques, leurs perceptions de la crédibilité de l'évaluateur et de son capital relationnel, etc.), l'établissement d'une relation de confiance entre ces dernières et l'évaluateur (p. ex. pour les parties prenantes les habilités sociales de l'évaluateur constituent un élément majeur), les origines du mandat d'évaluation, le contexte sociopolitique, les caractéristiques de l'*evaluand*,

[7] Le mouvement de l'utilisation de l'évaluation ne fait pas simplement référence à l'utilisation des résultats évaluatifs. Afin d'assurer la rigueur et l'utilité de l'évaluation, il est nécessaire que les parties prenantes clés soient engagées dans le processus évaluatif, afin qu'ils acquièrent des connaissances en ce qui a trait à la démarche d'évaluation, aux processus décisionnels en amont des choix méthodologiques, etc. Il est important qu'ils s'approprient l'évaluation, qu'ils développent une pensée évaluative afin qu'ils deviennent eux-mêmes garants de sa qualité, de sa crédibilité et de sa légitimité (Patton, 2015a, 2008).

l'anxiété face à l'anticipation de conséquences négatives, la résistance, etc. (Bamberger et Mabry, 2020 ; Marchand, 2020 ; Dubeau, Bérubé, Lafantaisie, Coutu et Turcotte, 2018 ; Patton, 2018, 2015a, 2008 ; Alkin et King 2017 ; King et Stevahn, 2013 ; Taut et Alkin, 2003 ; Donaldson, Gooler et Scriven, 2002). L'évaluateur doit donc faire en sorte d'inviter les parties prenantes clés à faire part de leurs inquiétudes, de leurs interrogations, de leurs attentes ainsi que de leurs recommandations, et ce, tout au long du processus évaluatif. Il ne doit pas simplement les écouter, il doit prendre en compte les éléments qu'elles partagent avec lui et qu'il les considère comme des collaboratrices indispensables.

Cependant, il est important d'envisager les effets « pervers » et les potentielles influences négatives de l'implication des parties prenantes clés, de l'institutionnalisation de l'évaluation. En effet, avec le renforcement de l'utilisation et l'omniprésence des principaux utilisateurs ciblés viennent aussi les risques de mésévaluation (*misevaluation*) et de mésusage (*misuse*). Des risques qui touchent principalement la qualité des opérations d'évaluation, l'utilité et l'utilisation du processus et des résultats évaluatifs (Alkin et King, 2017; Patton, 2015a, 2008 ; Alkin, 2011). La mésévaluation fait référence à une situation où l'évaluateur réalise une faible performance ou n'a pas su respecter les standards et les principes propres à la pratique de cette discipline. Par exemple, la pression liée à la mise en œuvre d'une évaluation sous d'importantes contraintes et pressions politiques engendre des problèmes en ce qui a trait à l'application de solides méthodes de recherche et à la gestion des facteurs affectant la validité des inférences faites à partir des données collectées et faisant l'objet de conclusions évaluatives sur la performance du programme (Bamberger et Mabry, 2020). Le poids des contraintes sociopolitiques et organisationnelles sur la qualité et l'utilité de l'évaluation est un élément caractéristique des situations évaluatives rencontrées dans le contexte de l'évaluation de programmes publics. Lamari, Poulin-Larivière et Lemieux (2016) mettent de l'avant les problèmes fréquents liés au calibrage inadéquat de l'évaluation au regard des ressources disponibles et mobilisables. Ils parlent ici du phénomène paradoxal lié à l'exigence de livrer une évaluation fiable et fondée sur suffisamment de données probantes, alors que très souvent l'échéancier est trop serré, le budget insuffisant pour pallier efficacement les difficultés importantes d'accès aux données. Les contraintes situationnelles limitent la collecte des données requises pour répondre aux exigences de rigueur et de validité. Les résultats de l'évaluation de la Politique d'évaluation de 2009 (Secrétariat du Trésor du Canada, 2015) mettent en avant la même problématique, à savoir : l'obligation pour les ministères d'évaluer l'ensemble de leurs

programmes de subventions et de contributions, tous les cinq ans, et de répondre à l'ensemble des questions fondamentales de la politique évaluative fédérale sur le rendement de leur programme (efficacité, efficience et économie), et ce, en dépit des problèmes de faisabilité (nature et maturité du programme, échéanciers serrés, budgets insuffisants, lacunes sur le plan des données de mesure du rendement du programme, etc.). Dumaine (2018) parle d'attentes irréalistes en ce qui a trait à l'évaluation des résultats ultimes / de l'impact, et de l'efficience et l'économie des programmes dans le contexte des évaluations externes réalisées en réponse aux mandats d'évaluation quinquennale.

Quant au mésusage, celui-ci se manifeste lorsque les utilisateurs de l'évaluation manipulent l'évaluation de sorte que cela « falsifie » les résultats ou corrompent l'enquête. Les possibilités de détournements et de corruption des évaluations font par ailleurs partie des préoccupations de la profession. Il peut s'agir de parties prenantes qui posent la mauvaise question, de pressions exercées sur les évaluateurs internes afin qu'ils ne présentent que les résultats positifs, de l'utilisation de conformité en réponse à une obligation d'évaluation, ou encore de manifestations d'intérêts personnels (Alkin et King, 2017 ; Patton, 2015a ; Alkin et Coyle, 1988).

Comme nous avons pu le voir tout au long de cette problématique, de nombreux éléments contextuels influencent la pratique de l'évaluation, et sont susceptibles d'avoir des répercussions plus ou moins considérables sur le développement et la mise en œuvre de la méthodologie d'évaluation ainsi que sur la qualité et l'utilité des résultats évaluatifs. Aujourd'hui, l'évaluateur doit être en mesure de produire une évaluation rigoureuse en s'assurant à la fois d'impliquer les parties prenantes clés tout au long de l'activité évaluative, d'adapter ses choix méthodologiques en fonction des contraintes sociopolitiques, budgétaires, temporelles et d'accès aux données ainsi qu'en veillant à prévenir les risques de mésusage et de mésévaluation. En effet, la validité, la pertinence et la fiabilité d'une évaluation reposent sur la capacité de l'évaluateur à adapter efficacement ses choix méthodologiques face aux différentes pressions exercées par l'environnement socio-économique, politique et culturel dans lequel l'évaluation se déroule. Aussi, la dimension à la fois technique et sociopolitique de l'évaluation de programme nous pousse à nous questionner sur la formation des évaluateurs, et s'ils sont bel et bien équipés pour répondre à ces exigences sur le terrain.

1.2.2 Le niveau d'expertise des évaluateurs de programme

Les évaluateurs mobilisent un ensemble considérable et varié de compétences que Stevahn, King, Ghere et Minnema (2005) ont réparties sur six volets ou domaines : (1) pratique professionnelle (2) enquête systématique (3) analyse situationnelle (4) gestion de projet (5) relations interpersonnelles et (6) pratique réflexive. Ils doivent donc faire preuve d'un total de 61 compétences, de la gestion de projet à la formation des parties prenantes en évaluation de programme. En outre, en fonction des caractéristiques du programme à évaluer, du mandat et du devis d'évaluation, etc., la liste des rôles, des connaissances et des compétences à mobiliser chez peut être assez conséquente, ce qui nécessite davantage le recours à une équipe évaluative plutôt qu'à un seul évaluateur. Cependant, dans la réalité ce n'est pas nécessaire toujours le cas et un évaluateur peut aisément se trouver à conduire une évaluation de programme, seul, alors que celle-ci exige la mobilisation d'une équipe évaluative. D'ailleurs, si une équipe évaluative est mobilisée et que l'ensemble des connaissances et des compétences exigées ne résident pas dans un seul individu, il n'en demeure pas moins que l'évaluateur responsable du projet d'évaluation (généralement un évaluateur chevronné) se doit de détenir l'ensemble des connaissances théoriques et pratiques relatives à toutes les composantes de l'activité évaluatives (de la planification de l'évaluation à la communication des résultats). Notons que cette Taxonomie des compétences essentielles (Stevahn, King, Ghere et Minnema, 2005) a été adaptée et révisée par la Société canadienne d'évaluation dans le cadre de la conception et l'adoption de son propre référentiel de compétences requises à l'exercice de l'évaluation (SCÉ, 2018).

De surcroit, selon Hatry, Newcomer et Wholey (2015), le type et le niveau de formation sont informatifs, mais ils ne sont en aucun cas déterminants dans le cadre du processus de sélection du personnel évaluatif. En effet, leur réussite dans le cadre de cours ou d'une formation en évaluation augmente la probabilité que les candidats sachent ce qu'ils font, au moins, en ce qui a trait à l'aspect technique du travail évaluatif. Cependant, les auteurs constatent que certains candidats avec ce type de profil ne s'adaptent pas (ou moins bien) dans un environnement où les contraintes en matière de ressources neutralisent la possibilité de mettre en œuvre des méthodologies prescrites, de recourir à des méthodes sophistiquées (devis expérimentaux, quasi expérimentaux), ou encore à des procédures de triangulation (méthodes de collecte de données, sources, évaluateurs). Selon nous, ce constat s'expliquerait, entre autres, par un manque

d'expérience pratique qui freine leurs capacités d'adaptation à l'environnement évaluatif en question. En effet, avoir des compétences dites techniques ou plutôt détenir des ressources (savoirs théoriques et procéduraux, habiletés techniques, etc.) est une condition nécessaire, mais insuffisante, pour mettre en œuvre une pratique professionnelle pertinente et adaptée à l'environnement (McDavid, Huse et Hawthorn, 2019 ; McPherson, 2015 ; Le Boterf, 2011). Une pratique professionnelle pertinente s'inscrit dans une succession de prises de décisions, d'actions, d'interactions et de processus d'autorégulation en réponse à l'interprétation des critères de réalisation souhaitables d'une activité. Ce processus itératif repose sur le recours à des schémas opératoires et leurs combinatoires de ressources appropriées (soit le jumelage de savoirs théoriques et procéduraux, avec des savoirs pratiques et expérientiels). Aussi, c'est l'expérience pratique qui permet le développement et le stockage de schèmes opératoires et de leurs combinatoires pertinentes en fonction d'une situation ou d'une famille de situations donnée. La pratique professionnelle évolue donc sur un continuum développemental, au fil de l'expérience des professionnels sur le terrain, permettant l'enrichissement progressif d'un répertoire de situations vécues au sein duquel chaque situation ou famille de situations détermine une action immédiate, intuitive, pertinente et adaptée à l'environnement. On parle aussi de la mobilisation de théories implicites (McPherson, 2015 ; Urban et Tronchin, 2009), de théories spontanées (Shadish, Cook et Leviton, 1991), ou encore de connaissances pragmatiques ou concepts en actes (Tourmen, 2009, 2007 ; Pastré, Mayen et Vergnaud, 2006), qui sont développés dans et par l'expérience. Il s'agit de systèmes de représentation en actes composés de « propositions tenues pour vraies sur le réel », et de raisonnements conditionnels de type « si..., alors... ». Dans le contexte de la pratique des évaluateurs de programmes publics, ces systèmes de représentation portent sur trois éléments cruciaux qui permettent de diagnostiquer le contexte, de déduire l'enchainement des actions et d'en anticiper les conséquences possibles : i. l'objet évalué ; ii. les stratégies des acteurs autour de l'évaluation (exigences politico-organisationnelles sur le plan de l'évaluation, des attentes, des approches, des méthodes, etc.) ; et iii. les moyens méthodologiques - ressources disponibles pour évaluer (Tourmen et Droyer, 2013 ; Tourmen, 2009, 2007).

Enfin, Hatry, Newcomer et Wholey (2015) expliquent qu'au cours de leurs carrières ils ont travaillé avec de nombreux excellents évaluateurs qui détiennent des formations surprenantes (p. ex. des historiens et des avocats). Pour eux, les évaluateurs doivent recourir à des approches systématiques, logiques et faire preuve d'humilité culturelle, et que l'habileté à appliquer de telles

approches peut se trouver chez des individus ayant des formations très différentes. Une telle remarque est très importante dans le contexte de cette thèse doctorale, puisque les participants exercent dans des cabinets de consultants en gestion et en évaluation de programme et ils ont des profils très diversifiés. En effet, certains d'entre eux sont avocats, d'autres sont sociologues, psychologues ou encore docteurs en économie. Cependant, des auteurs tels que Ghere, King, Stevahn et Minnema (2005) soulèvent certaines interrogations quant aux potentielles limites liées à l'arrivée constante de nouveaux évaluateurs sur le terrain qui n'ont pas nécessairement de formation initiale en évaluation. En effet, s'il existe aujourd'hui des diplômes universitaires en évaluation de programme, des programmes d'accréditation professionnelle, beaucoup de jeunes évaluateurs le deviennent de manière accessoire, dans le cadre de leur travail, lorsque leurs supérieurs leur donnent le mandat d'évaluer un de leur programme alors que ces derniers n'ont aucune expérience en évaluation. Par ailleurs, un autre profil fréquent se dessine, à savoir de nouveaux évaluateurs qui n'ont pas de formation en évaluation, mais qui possèdent des compétences que l'on retrouve dans le socle de compétences essentielles en évaluation de programme. Aussi, de telles situations impliquent que certains évaluateurs sont conscients qu'il existe des normes et des lignes directrices en évaluation de programme, mais d'autres ne le sont pas nécessairement. Cela nous conduit donc à nous interroger sur leurs pratiques, et ce, en fonction de leur niveau d'expertise (novice et expert [Le Boterf, 2002]).

1.3 Objet de recherche

En créant le Programme mondial pour l'évaluation 2016-2020, les EvalPartenaires (UNEG, SCÉ, SQEP, REF, etc.) ont invité les membres de la communauté évaluatrice (incluant les organismes gouvernementaux, les parties prenantes, les praticiens et les chercheurs en évaluation, etc.) à s'engager à faire de la qualité et de l'utilité de l'évaluation de programme une priorité. L'EvalAgenda2020 stipule que la recherche en évaluation constitue une priorité afin de favoriser le développement des capacités institutionnelles en évaluation. Selon nous, en soutenant et en encourageant la recherche en évaluation, la communauté évaluatrice reconnait le rôle indispensable qu'elle peut jouer dans l'atteinte des priorités en évaluation. En effet, les recherches empiriques ont le pouvoir de permettre la promotion et la facilitation d'évaluations de qualité, la proposition de ressources en évaluation afin de soutenir la qualité des données évaluatives ainsi que le développement de formation en évaluation afin d'assurer la qualification des évaluateurs.

Elles peuvent, mais surtout doivent guider la pratique en évaluation de programme. Et, il est de notre ressort en tant que chercheur en évaluation, d'enquêter en profondeur sur l'importance de certaines problématiques, et sur l'impact qu'elles ont sur le travail des évaluateurs et développer ainsi des pratiques de résolution de problèmes efficaces (Smith, 2008).

Au fil des générations, l'évaluation de programme a traversé de nombreux changements (nécessaires), tant sur le plan méthodologique que sur le plan social, faisant d'elle une activité à la fois technique et sociopolitique. Aujourd'hui, elle est d'une indéniable complexité du fait notamment de la nature des situations évaluatives rencontrées, dont les caractéristiques influencent la faisabilité, la fiabilité et l'utilité de l'évaluation. Des situations animées de contraintes sociopolitiques et organisationnelles importantes, de problèmes sur le plan des intrants disponibles, de l'accessibilité des données, etc. En outre, la qualité de l'évaluation est tributaire de la capacité de l'évaluateur à mobiliser des savoirs pratiques en situation, son expertise, et son jugement afin d'entreprendre les choix méthodologiques les plus appropriés. Un raisonnement méthodologique qui s'appuie, entre autres, sur des théories implicites ou spontanées acquises dans et par l'expérience de l'évaluateur.

Dans le cadre de cette recherche doctorale, nous nous intéresserons aux choix méthodologiques de l'évaluateur de programme, selon qu'il soit novice ou expert, et de l'influence que les facteurs contextuels tels que la relation avec les parties prenantes clés et la gestion des contraintes inhérentes au terrain (contraintes budgétaires, temporelles, politico-organisationnelles, d'accès aux données) peuvent avoir sur l'opérationnalisation desdits choix. Notons qu'afin d'être en mesure de comprendre et décrire pleinement les pratiques évaluatives d'ordre méthodologique des évaluateurs de programme participants à notre recherche ainsi que l'influence des facteurs contextuels sur lesdites pratiques et les stratégies mises en œuvre pour y faire face, nous avons décidé d'examiner à la fois les perceptions des évaluateurs quant aux « bonnes » pratiques évaluatives (pratiques recommandées) et les explications fournit quant aux choix méthodologiques opérés dans le cadre d'projet d'évaluation auxquels ils ont pris part (pratiques réelles).

Dans le contexte de cette recherche, l'exercice professionnel des évaluateurs de programmes s'inscrit dans une activité de services où, en tant que prestataires (ou membres de l'équipe évaluative prestataire), ils répondent à des commanditaires et produisent des connaissances permettant l'attribution d'un jugement de valeur sur un programme public fédéral.

Il s'agit là d'un environnement de pratique hautement complexe qui est habité de contraintes caractéristiques de l'évaluation en situation réelle (Bamberger et Mabry, 2020). Il importe ici de fournir quelques informations en ce qui trait à l'environnement politique et organisationnel dans lequel est ancrée l'activité évaluative, plus précisément l'organisation du travail.

L'évaluation au gouvernement fédéral canadien s'inscrit dans une Politique évaluative accompagnée d'une Directive qui inclut les Procédures obligatoires pour l'évaluation et les Normes de l'évaluation élaborées par le Secrétariat du Conseil du Trésor du Canada. Dans le cadre de cette étude doctorale, les pratiques évaluatives s'inscrivent dans le contexte de la Politique d'évaluation de 2009 et la Directive « Fonction de l'évaluation » qui s'y rattache (la Politique de 2009 a été remplacée par la Politique sur les résultats en 2016). Au niveau des différents ministères, la fonction de l'évaluation repose sur les épaules de leur unité d'évaluation respective, qui est dirigée par un chef d'évaluation responsable de l'ensemble des activités d'évaluation de l'agence (l'unité d'évaluation est indépendante des programmes – souci d'évaluation neutre). Les évaluations individuelles des programmes sont planifiées (et calibrées) au moment de la conception du plan d'évaluation quinquennal (révisé annuellement) par l'agence gouvernementale.

Dans le cadre de la préparation de la demande de propositions (sous-traitance à l'externe), le chef de l'évaluation en collaboration avec les gestionnaires de programme définit les paramètres de l'évaluation (conception initiale de l'évaluation). L'évaluation est dirigée par le chef d'évaluation de l'agence commanditaire, en consultation avec un comité d'évaluation interne composé d'utilisateurs ciblés (bailleurs de fonds, décideurs, gestionnaires de programme). En tant que prestataire de services, l'équipe évaluative externe réalise l'évaluation. Cette équipe est dirigée par un évaluateur sénior (la plupart du temps un partenaire de la firme contractante) qui à titre de gestionnaire du projet est chargé des activités cruciales de l'évaluation : collaboration avec l'unité d'évaluation interne, conception du devis, supervision des procédures de collecte et d'analyse de données, agrégation et communication des résultats, et formulation des constatations évaluatives. La version finale du rapport d'évaluation est de la responsabilité du chef de l'évaluation, ce qui inclut la formulation des recommandations. Outre la prise en compte des exigences de la Politique évaluative et de sa Directive ainsi que des ressources et recommandations pratiques et méthodologiques fournies par le Centre d'excellence en évaluation du Secrétariat du Conseil du trésor du Canada (aujourd'hui remplacé par la Division sur les résultats), chaque ministère semble

avoir ses propres attentes, préférences en matière de communication de l'évaluation et de ses résultats.

1.4 Questions de recherche

Question principale

En quoi les évaluateurs novices et les évaluateurs experts œuvrant dans des cabinets de conseil en évaluation de programmes publics se distinguent-ils dans les choix méthodologiques qu'ils opèrent relativement à la réalisation de leurs évaluations de programme ?

Questions secondaires

(1) Quelles sont les ressemblances et les différences entre les pratiques réelles et les pratiques recommandées des évaluateurs externes de programmes publics en ce qui a trait aux choix dits méthodologiques ?

(2) Quelles sont les perceptions des évaluateurs de programme externes quant aux possibles influences des facteurs contextuels sur les choix méthodologiques opérés durant leurs évaluations de programmes publics ?

Questions tertiaires

(1) Lors de l'analyse critique d'une évaluation de programme tierce, quels sont les éléments mis en exergue par les évaluateurs de programmes publics relativement aux choix méthodologiques, et à l'influence des facteurs contextuels sur ceux-ci ?

(2) Quelles sont les informations fournies par les évaluateurs, pour expliquer et décrire les choix méthodologiques opérés dans le cadre de leurs évaluations de programme ?

Enfin, ce présent écrit constitue notre thèse doctorale et il est composé d'une recension des écrits, de la présentation du cadre conceptuel, de la méthodologie qui a été mobilisée afin de répondre à nos questions de recherche, de l'analyse et de l'interprétation des résultats, et enfin d'une conclusion.

CHAPITRE 2. RECENSION DES ÉCRITS ET CADRE CONCEPTUEL

À travers le premier chapitre, nous avons vu que l'histoire de l'évaluation de programme et l'évolution des pratiques évaluatives furent marquées par certaines problématiques fondamentales et les différentes tentatives visant à les résorber. Aussi, chaque tentative a été modelée par une quête de qualité et d'utilité de l'évaluation de programme et de ces processus méthodologiques. Or, actuellement, comme ce fut le cas au cours des générations précédentes, la volonté de promouvoir l'utilisation de l'évaluation de programme et la qualité de ses procédures et de ses résultats se trouve confrontée aux contraintes inhérentes au terrain, aux débats méthodologiques, à l'accumulation des rôles et des tâches des évaluateurs ainsi qu'à la nécessité d'impliquer, plus ou moins, les parties prenantes dans le processus évaluatif. De fait, ce sont de tels constats qui nous ont conduits à nous interroger, dans le cadre de cette recherche doctorale, sur les enjeux relatifs à la dimension méthodologique de la pratique évaluative et à l'influence que différents facteurs contextuels tels que les contraintes inhérentes au terrain, l'implication des parties prenantes et le niveau d'expertise des évaluateurs de programme peuvent avoir sur l'opérationnalisation des choix dits méthodologiques dans le cadre d'opération d'évaluation.

Dans le cadre de ce deuxième chapitre, nous présentons le cadre conceptuel de la recherche, soit la description des concepts clés qui ont permis son opérationnalisation (Figure 4, p. 37). Dans un premier temps, nous proposons un tour d'horizon de l'évaluation de programme et des pratiques évaluatives actuelles, plus particulièrement le processus itératif de la planification et de la mise en œuvre de la méthodologie évaluative. Dans un deuxième temps, nous abordons les facteurs contextuels susceptibles de représenter de potentiels défis dans le cadre de la pratique évaluative. Nous montrons rapidement que la place centrale accordée aux parties prenantes clés et les efforts entrepris pour renforcer l'utilisation de l'évaluation, ainsi que les contraintes inhérentes au terrain sur le plan des ressources disponibles, des exigences et du fonctionnement politico-organisationnels, etc., sont des facteurs déterminants dans l'opérationnalisation des choix dits méthodologiques. Troisièmement, nous revenons en détail sur un autre facteur qui selon nous joue un rôle important dans la pratique évaluative et plus particulièrement dans la conception et la mise en œuvre de procédures méthodologiques de qualité, à savoir le niveau d'expertise des évaluateurs de programme. Enfin, nous proposons une synthèse conceptuelle illustrant le lien entre notre objet

d'étude, l'opérationnalisation des choix dits méthodologiques dans le cadre d'évaluation et l'influence de certains facteurs contextuels sur ces choix.

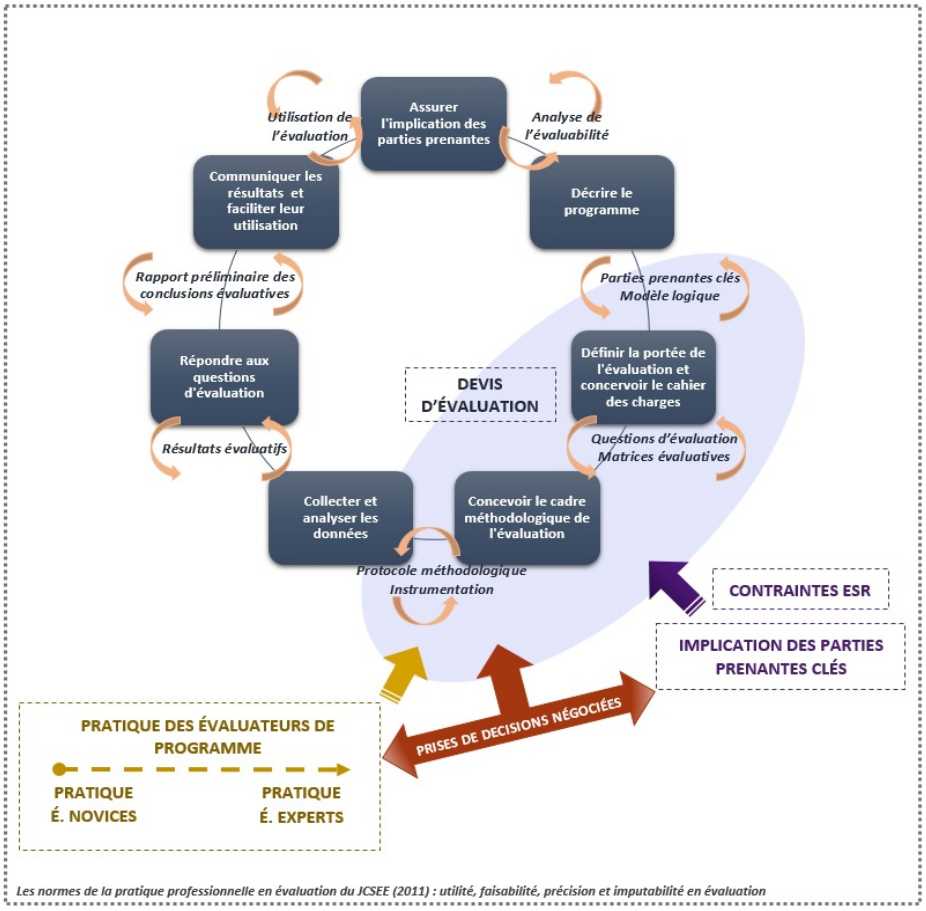


Figure 4. *Synthèse du cadre conceptuel*

2.1 La variable « méthodologie » au cœur des enjeux contemporains en évaluation

2.1.1 Définitions actuelles de l'évaluation de programme

À divers moments dans la vie d'un programme, les différentes parties prenantes (les financeurs, les décisionnaires, les gestionnaires, les clients ainsi que le public en général) doivent être en mesure de déterminer s'il est utile et rentable, et ce, afin de pouvoir le poursuivre, l'améliorer ou encore le remplacer par un nouveau programme (Rossi, Lipsey et Freeman, 2019). Pour être en mesure de prendre ce genre de décision, il importe de répondre à des questions du type : *le programme (et ses interventions) répond-il bel et bien aux besoins de la population ciblée ? Les interventions du programme sont-elles implantées telles que planifiées ? Est-ce que les services (activités, formations, interventions, etc.) prescrits par le programme ont bel et bien*

été fournis ? Est-ce que les objectifs du programme ont été atteints ? Au regard des relations coûts-bénéfices et coûts-efficacité, le coût total du programme est-il raisonnable ?

Dans les écrits nord-américains, un programme peut faire référence à un projet, une formation, un service, une politique, un curriculum, etc. (Tourmen, 2007). Pour Hurteau et Houle (2005), il s'agit d'« un système ouvert et systématique composé des dimensions qui lui sont propres – intrants, processus et résultats, environnements et rétroactions » (p. 86). Selon nous, cette définition demeure trop restrictive puisqu'elle ne rend pas compte de la logique (ou de la théorie) du programme et de ses raisons d'être. Pour le Conseil du Trésor du Québec (2002), « le programme désigne un ensemble cohérent et structuré d'objectifs, d'activités et de ressources humaines, financières, matérielles et informationnelles, regroupés pour offrir des produits et des services particuliers qui satisfassent les besoins ciblés d'une population prédéterminée » (p.6). Si, aux premiers abords cette définition est plus complète, finalement elle n'illustre pas non plus les fondements et les processus du programme. Aussi, selon nous, la définition de Plante (1994) reprise par Ridde et Dagenais (2012) constitue un bon compromis puisqu'elle met de l'avant à la fois sa structure et les processus qui l'animent, sa raison d'être, et les personnes responsables de sa mise en œuvre et de sa gestion, soit autant d'éléments déterminants pour la compréhension de la théorie du programme, et son évaluation (repérage des informations manquantes et des potentielles contraintes, identification des sources d'information, etc.):

> Un programme est « un ensemble cohérent, organisé et structuré d'objectifs, de moyens et de personnes qui l'animent. Il se justifie sur la base de besoins, définis comme une carence ou un manque qui affecte des individus, une collectivité ou une société. Il est sous le contrôle d'une ou plusieurs personnes responsables de la qualité de sa formulation et de son bon fonctionnement. Il est mis en place pour transformer des choses ou l'état des choses » (p. 16).

Au-delà de porter un jugement sur un programme, l'évaluation doit permettre de formuler des constats concernant son déroulement ainsi que la proposition de changements nécessaires à l'amélioration des services procurés à la population qu'il cible, d'accompagner les prises de décision, de renforcer les capacités évaluatives de l'organisation, d'accompagner, etc. (Bamberger et Mabry, 2020 ; Rossi, Lipsey et Freeman, 2019 ; Ridde et Dagenais, 2012 ; Bryson, Patton et Bowman, 2011 ; Yarbrough, Shulha, Hopson et Caruthers, 2011 ; Patton, 2008 ; Donaldson et Christie, 2006 ; Tard, Ouellet et Beaudoin, 1997). En outre, de telles tâches supposent de faire

appel à des praticiens de l'évaluation de programme (nous verrons plus tard que ce n'est pas toujours le cas), une discipline qui mobilisent les méthodes de la recherche en sciences sociales afin de collecter des informations permettant de faire le point sur la pertinence, la conceptualisation, la mise en œuvre, les effets et (ou) l'efficience d'un programme dans son ensemble, ou de certaines de ses composantes (Rossi, Lipsey et Freeman, 2019 ; Tourigny et Dagenais, 2005). C'est la finalité même de l'évaluation qui la distingue de la recherche. En effet, la démarche évaluative vise à porter un jugement sur la valeur, le mérite, la portée et (ou) la qualité du programme, de ses activités, de son organisation, etc. (Scriven, 2016, 1991 ; Chen, 2015 ; American Evaluation Association, 2014 ; Patton, 2008 ; Fournier, 2005). Enfin, la qualité et la crédibilité du jugement évaluatif est fondées sur une démarche méthodologie rigoureuse et systématique de collecte et d'analyse de données probantes (Société canadienne d'évaluation, 2015 ; Secrétariat du trésor du Canada, 2016 ; Patton, 2008 ; Fournier, 2005 ; Plante, 1991). Thibaut et *coll.* (2011) précise que la crédibilité de l'évaluation repose non seulement sur un jugement scientifiquement valide, mais aussi sur un jugement qui est socialement légitime, c'est-à-dire l'acceptabilité et la recevabilité du jugement auprès des parties prenantes.

Le recours exponentiel à l'évaluation de programme depuis ces 20 dernières est lié à sa popularité croissante et à la reconnaissance de son utilité dans de nombreux domaines tels que l'administration publique, la santé, l'éducation, la santé psychosociale et communautaire des populations, l'environnement, le développement international, etc. Ainsi trouve-t-on de nombreux *organismes publics* (le Centre d'excellence en évaluation du Secrétariat du trésor du Canada, l'Agence comptable générale des États-Unis, etc.) et *à but non lucratif* (les Nations Unies avec UNEG, UNDP, etc.) ainsi que des *centres et des instituts de recherche* (l'Institut de la santé publique et des populations du Canada, le Centre de recherche sur les services éducationnels et communautaires de l'Université d'Ottawa, le Centre de recherche et d'expertise de l'ENAP, etc.), qui se spécialisent en recherche et en évaluation. En Amérique du Nord, on retrouve également de nombreux cabinets de consultants en gestion et en évaluation, particulièrement dans les capitales ou autres grandes villes administratives, qui pratiquent aussi bien l'évaluation de programmes publics que l'évaluation de programmes d'organismes à but non lucratif.

L'évolution de l'évaluation de programme et l'élargissement de son champ d'exercices ont donc logiquement conduit à l'émergence d'associations professionnelles telles que la Société

canadienne d'évaluation (SCÉ). Elles ont globalement pour but de développer une culture de l'évaluation de programme commune et contextuelle, de prodiguer des normes et des critères relatifs à la qualité de l'exercice évaluatif, de proposer des formations, voire d'octroyer des accréditations. La définition des lignes directrices, des objectifs ainsi que la création d'une définition commune de l'évaluation de programme permettent ainsi aux évaluateurs et aux parties prenantes d'acquérir une compréhension collective des pratiques évaluatives propres à l'exercice de l'évaluation de programme dans un contexte donné (Patton, 2018, 2008 ; Gauthier et *coll.*, 2010). En outre, il est important de mentionner que la majorité des définitions de l'évaluation de programme utilisées aujourd'hui ont émergé du contexte américain (Poth et *coll.*, 2014). Or, au regard de l'évolution et de l'expansion de la pratique de l'évaluation de programme au Canada et de sa professionnalisation (programme d'accréditation), il est devenu nécessaire de formuler notre propre définition.

La SCÉ s'est ainsi engagée entre 2012 et 2014 dans un processus d'examen systématique et de consultation auprès de ses membres afin de faciliter la formulation d'une définition canadienne commune de l'évaluation de programme[8]. Ce projet fut laborieux et à notre sens beaucoup plus complexe que ce que les instigateurs avaient anticipé (Annexe 2). En effet, en 2012, un premier panel de consultation a été organisé auprès des membres du Conseil national de la SCÉ. Ces derniers ont été invités à identifier les caractéristiques indispensables définissant l'activité évaluative à partir des *6 définitions actuelles* de l'évaluation de programme (Annexe 3), puis à formuler une définition canadienne commune. Cette tâche s'est avérée difficile pour ces derniers qui ont eu du mal à associer leur pratique à une ou plusieurs de ces définitions et ils n'ont pas été en mesure de s'entendre sur la formulation d'une définition commune. En 2013, un second panel de consultation a été organisé, cette fois-ci auprès des membres de la SCÉ durant le congrès annuel de la SCÉ, où les participants ont été invités à commenter les résultats préliminaires de l'étude de

[8] Nous nous sommes référés aux notes de pratique publiées dans la Revue canadienne d'évaluation de programme : Poth, C., Lamarche, M. K., Yapp, A., Sulla, E. et Chisamore, C. (2014). Toward a Definition of Evaluation Within the Canadian Context: Who Knew This Would Be So Difficult? *Canadian Journal of Program Evaluation, 29*(1), 87-103.

Lamarche et Poth (2013). Une fois de plus les membres de la communauté évaluatrice ont eu du mal à associer les caractéristiques de leur pratique à une définition globale de l'évaluation de programme. En effet, ces derniers ne semblent pas se référer à une définition en particulier, mais plutôt à la description des étapes de la démarche évaluative. Pour eux, engager leur pratique professionnelle au sein d'une définition commune signifie restreindre ou simplifier (à l'excès) l'étendue de leurs activités évaluatives. En outre, si les membres de la communauté évaluatrice canadienne ne reconnaissent pas nécessairement le besoin « urgent » de proposer une définition de l'évaluation dans le contexte du Canada, il n'en demeure pas moins qu'avec les procédures d'accréditation des évaluateurs et de professionnalisation de l'évaluation de programme (avec la définition des compétences essentielles, etc.), qu'il est devenu essentiel de parvenir à la proposition d'une définition canadienne. Par conséquent, la SCÉ adopta en 2015 une définition canadienne « *proxy*[9] » de l'évaluation de programme.

Si selon nous, il est plus pertinent de proposer une modélisation canadienne des étapes clés du processus évaluatif, en prenant le soin de définir les éléments caractéristiques de chacune d'entre elles, nous avons tout de même décidé de retenir la définition canadienne adoptée par la SCÉ (Encadré 1, p. 42), dans le cadre de notre recherche doctorale. Une définition qui rappelons-le s'appuie notamment les éléments caractéristiques des *6 définitions actuelles* de l'évaluation de programme (Annexe 3). Cette définition représente un compromis nécessaire visant à offrir une première définition canadienne, bilingue (français et anglais) de l'évaluation de programme, tout en rendant compte des multiples facettes de l'évaluation et de l'étendue des activités entreprises dans le cadre de ses opérations. En effet, les auteurs ont pris le soin de décrire chaque concept clé de la définition afin de rendre compte des caractéristiques propres aux activités évaluatives. Et selon nous cela nous répond davantage aux besoins des évaluateurs. Cette définition correspond au contexte de notre recherche en évaluation et de la pratique de nos évaluateurs participants

[9] Face à l'impératif de définir l'évaluation de programme en contexte canadien, et ce, en dépit de l'absence de consensus au sein de la communauté évaluatrice canadienne, la SCÉ à proposer et adopter une définition approximative que les membres sont invités à enrichir, à modifier, etc., en attendant qu'un modèle davantage unanime soit proposé.

puisqu'ils sont pour la majorité membres de la SCÉ et que certains d'entre eux sont également des évaluateurs accrédités (ÉA) par la Société. D'ailleurs, la section de la région de la capitale nationale de la SCÉ où ils exercent comptait 381 membres en 2017, dont 92 ÉA, ce qui représente la deuxième section la plus importante après l'Ontario (408 membres, dont 56 ÉA), et devant l'Alberta (255 membres, dont 41 ÉA). Enfin, comme Gauthier et *coll.* (2010) le soulignent, il est important de promouvoir la recherche et la pratique en évaluation de programme dans le contexte canadien, un contexte animé par un certain nombre d'enjeux liés notamment à l'exercice professionnel en milieu plurilingue, en milieu politique à la fois fédéral et provincial, en milieu autochtone, etc.

Encadré 1. *Définition canadienne « proxy » de l'évaluation de programme (SCÉ, 2015)*[10]

« L'évaluation est l'appréciation systématique de la conception, de la mise en œuvre ou des résultats d'une initiative pour des fins d'apprentissage ou de prise de décision ».

L'**appréciation** évaluative porte sur la valeur, le mérite, la portée, l'importance ou la qualité d'une **initiative** (un programme, un projet, un sous-programme, un sous-projet, une intervention ou leurs composantes). L'activité évaluative doit être **systématique** et doit être basée sur le recueil et l'analyse de données empiriques (à l'aide des méthodes de recherche propres aux sciences sociales), permettant la formulation de conclusions. Enfin, en ce qui a trait aux **finalités** d'une évaluation, celles-ci permettent à partir des résultats obtenus de former un jugement, de formuler des conclusions, de prendre des décisions relativement au programme, d'appuyer le développement organisationnel, etc. L'ultime but étant de conduire à la formulation de suggestions pour amélioration, pour appuyer des prises de décision informées en ce qui a trait au futur du programme.

2.1.2 Le caractère itératif de la démarche évaluative

La troisième génération de l'évaluation fut le point marquant de la légitimation de l'évaluation de programme en tant que discipline à part entière et de la sophistication des approches évaluatives et de la pratique des évaluateurs de programme. En effet, de nombreuses approches et

[10] La définition de l'évaluation et ces termes clés est disponible sur le site internet de la SCÉ : http://evaluationcanada.ca/fr/quest-ce-que-levaluation

méthodologies évaluatives ont été rapidement développées pour répondre au besoin croissant d'évaluation des politiques administratives et sociales et de leurs programmes respectifs (Rossi, Lipsey et Freeman, 2019). L'évaluateur a alors la responsabilité d'investiguer et d'utiliser toutes les informations à sa disposition pour attester de la performance d'un programme. À titre d'expert, il doit recourir à des méthodes d'investigation précises et systématiques qui permettent de prononcer un jugement évaluatif crédible et générer des résultats évaluatifs utiles (Fishman, 1996 ; Guba et Lincoln, 1989).

Aujourd'hui, le besoin d'évaluation n'a pas diminué (bien au contraire), et la notoriété grandissante de l'évaluation de programme conduit de plus en plus de domaines à recourir à cette pratique afin d'accompagner le développement, l'implantation et le monitorage de leurs programmes. Nous l'avons mentionné plus tôt, des organismes tels que le Centre d'excellence en évaluation du Secrétariat du trésor du Canada ont été créés pour guider et réglementer les pratiques évaluatives au niveau des gouvernements fédéral et provincial, pratiques désormais institutionnalisées par la politique sur l'évaluation de 2009. Dans la même lignée, des associations professionnelles ont vu le jour, notamment, afin de faire avancer la théorie et la pratique en évaluation de programme et de prodiguer des lignes directrices, des normes et des critères de pratiques évaluatives, etc. (p.ex. le code déontologique de la SCÉ, approuvée initialement en 1996). Si l'évaluation constitue l'outil privilégié pour rendre compte de l'efficacité d'un programme et accompagner les prises de décision relatives à la vie de ce dernier, elle est cependant synonyme d'enjeu pour les organismes politiques, sociaux, éducationnels, sanitaires, etc. En effet, elle est souvent instrumentalisée dans le cadre de prises de décisions relatives à l'allocation de ressources pour les programmes et les politiques (gestion axée sur les résultats), et ce, dans un paysage économique fragilisé par la multiplication des programmes, les pressions politiques internationales, les impératifs éducationnels, sociaux et sanitaires, etc.

En dépit des enjeux sociopolitique et économique qui animent l'environnement évaluatif, l'évaluation doit permettre une description précise et minutieuse des entités et des processus d'un programme ainsi que de sa performance, et ce, en ayant recours des méthodes rigoureuses et en assurant la qualité des procédures et des résultats ainsi que l'utilité de l'évaluation (Yarbrough, Shulha, Hopson et Caruthers, 2011). Aussi trouve-t-on diverses approches et méthodes évaluatives permettant de se prononcer sur une ou plusieurs dimensions d'un programme, soit sur : (1) la

nécessité du programme ; (2) la conception du programme ; (3) l'implantation du programme et les services offerts ; (4) les effets du programme ; (5) la performance et la rentabilité du programme (efficience) ; et (6) l'impact (Rossi, Lipsey et Freeman, 2019 ; Alkin, 2013 ; Ridde et Dagenais, 2012 ; Patton, 2008). En outre, quels que soient l'approche évaluative mobilisée et le type d'évaluation sélectionnée, la démarche évaluative doit tenir compte d'un certain nombre d'étapes indispensables (McDavid, Huse et Hawthorn, 2019 ; Hurteau et Houle, 2012 ; Ridde et Dagenais, 2012 ; Stufflebeam, 2004 ; Darabi, 2002) (Figure 5, p. 45) :

- *Étapes 1 et 2 - Évaluation de l'évaluabilité* : analyser les finalités de l'évaluation, identifier les utilisateurs ciblés ainsi que leurs besoins/attentes, décrire et comprendre le programme et ses composantes, analyser les intrants disponibles pour l'évaluation, évaluer le contexte évaluatif, etc. (extrants : identification des parties prenantes clés impliquées dans l'évaluation et création du modèle logique).
- *Étape 3 - Définition du cadre de l'évaluation (cahier des charges)* : définir la portée de l'évaluation, le type d'évaluation, les questions d'évaluation, les procédures méthodologiques envisagées, identifier les sources d'informations disponibles, etc. (extrant : questions d'évaluation, matrice d'évaluation).
- *Étape 4 - Conception du cadre méthodologique de l'évaluation* : déterminer le processus de collecte de données, développer les instruments de collecte et d'analyse de données, recruter les participants, développer le plan de communication des résultats (extrant : protocole d'évaluation, instrumentation).
- *Étape 5 - Collecte et analyse des données* : collecter les données, monitorer le déroulement de la collecte de données, analyser les données, étudier les résultats préliminaires (extrant : résultats évaluatifs)
- *Étape 6 - Rédaction des conclusions préliminaires* : rédiger un premier rapport d'évaluation et discuter des résultats ainsi que des conclusions avec les parties prenantes clés (extrant : rapport préliminaire des conclusions évaluatives)
- *Étape 7 - Communication des résultats et facilitation de leur utilisation* : rédiger et présenter le rapport final aux parties prenantes clés et discuter de l'implantation des recommandations (extrant : utilisation de l'évaluation).

Les opérations d'évaluation mobilisent généralement des méthodes de recherche en sciences sociales, cependant elles ne suivent pas à la lettre les principes et l'ensemble des règles prescrites d'un ou plusieurs types de recherche (quantitatif ou qualitatif, expérimental ou

participatif, positiviste ou naturaliste, etc.) (Martens et Wilson, 2010). En effet, les conditions environnementales, la nécessité de s'adapter au contexte politique et organisationnel, les caractéristiques du programme, la gestion des contraintes de terrain (les problèmes de temps, de budget ou encore d'accès aux données), etc., conduisent très souvent les évaluateurs à amender le devis d'évaluation à plusieurs reprises, à improviser pendant la collecte de données, etc., et ce, afin de s'assurer que les résultats évaluatifs soient crédibles et défendables (McDavid, Huse et Hawthorn, 2019 ; Rossi, Lipsey et Freeman, 2019). L'évaluateur se retrouve souvent engagé dans des négociations, et ce, dès la première rencontre avec les commanditaires de l'évaluation. Ils négocient la définition des questions d'évaluation, le choix des méthodes de collectes de données, la sélection des sources d'information, l'utilisation des résultats, etc., et ce, en tout temps de l'évaluation. De telles caractéristiques nous conduisent à qualifier la démarche évaluative d'itérative, particulièrement en ce qui a trait à l'opérationnalisation des décisions d'ordre méthodologique (Figure 5) :

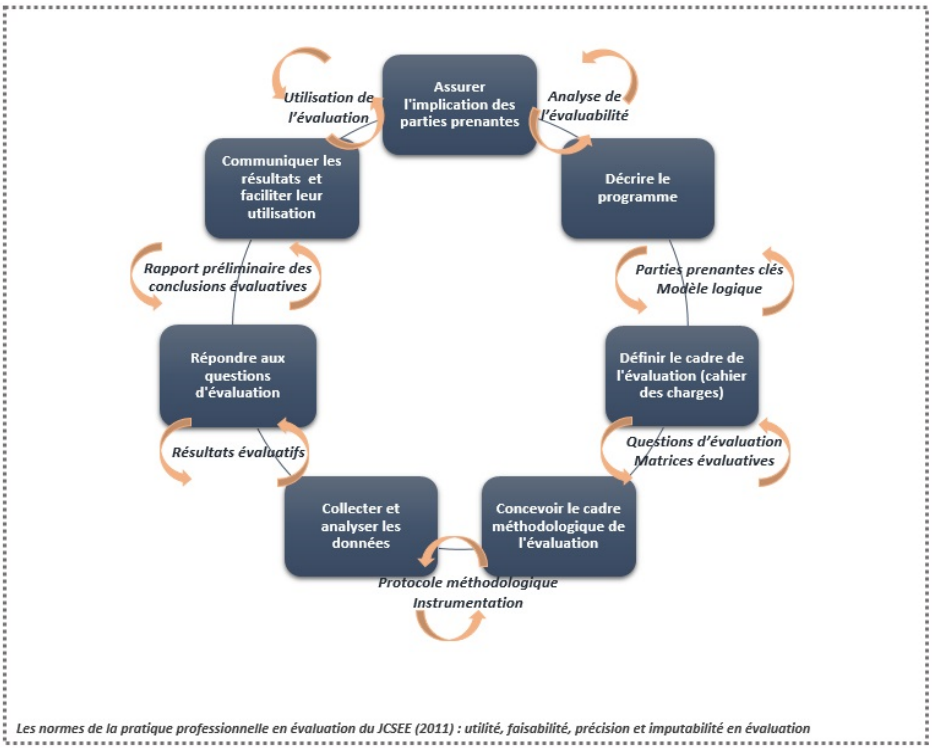


Figure 5. *Le cycle itératif de la démarche d'évaluation de programme*

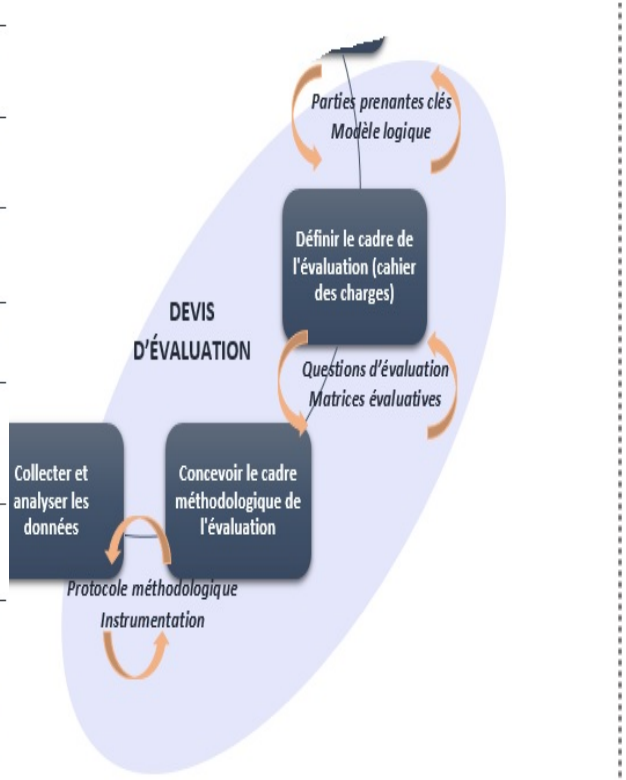
2.1.3 Définitions du devis d'évaluation et des choix dits méthodologiques

Le devis d'évaluation constitue l'empreinte méthodologique de l'évaluation et il doit rendre compte avec transparence, clarté et précision de l'ensemble de la démarche (Rog, 2005). Il

englobe à la fois la définition du cadre d'évaluation et la conception du cadre méthodologique, et il est conçu avec la collaboration des parties prenantes clés. Le *cadre d'évaluation* définit les paramètres de l'évaluation, il fait d'ailleurs office de cahier des charges. Il décrit les bornes de l'évaluation, c'est-à-dire les objectifs de l'évaluation ainsi que de sa fonction, le type d'évaluation et l'approche évaluative sélectionnés ainsi que les questions principales d'évaluation, l'identification des sources données (*Better Evaluation*, 2014; Alkin, 2011, 2009; Stufflebeam, 2004, 1994; Darabi, 2002). Il est essentiel qu'il propose un échéancier et une estimation des ressources matérielles, humaines et financières nécessaires à la réalisation de l'évaluation. Il est souvent opérationnalisé sous la forme d'une matrice d'évaluation où, pour chaque question d'évaluation, sont référencés les indicateurs de mesure, les sources d'information, et les méthodes de collectes de données envisagées qui permettront d'y répondre (McDavid, Huse et Hawthorn, 2019 ; Secrétariat du trésor du Québec, 2018). C'est un excellent outil de planification, de collaboration et de communication du plan d'évaluation avec les clients et c'est un outil indispensable pour la conception du cadre méthodologique. Elles permettent notamment de vérifier la faisabilité de l'évaluation et d'assurer la scientificité de la démarche ainsi que la validité et la fiabilité des inférences faites à partir des données collectées (triangulation des sources, des méthodes, etc.). *Le cadre méthodologique* est une forme de protocole d'évaluation, il offre un plan détaillé et opérationnel des procédures de collecte et d'analyse des données, de communication de l'évaluation et de l'utilisation des résultats (Alkin, 2011; Stufflebeam, 2004, 1999). Il décrit les procédures mobilisées afin d'obtenir des données probantes à partir desquelles l'évaluateur va pouvoir répondre aux questions d'évaluation et émettre un jugement vis-à-vis des composantes du programme évalué et le contexte dans lequel il est implanté. Parmi ces procédures se trouvent la sélection des sources d'information, les procédures de collecte et d'analyse des données, le plan de communication des résultats, etc.

Sachant que le devis d'évaluation constitue la pierre angulaire de la démarche d'évaluation, il parait essentiel d'étayer davantage les différents choix méthodologiques qui s'offrent à l'évaluateur dans le cadre des activités propres à la conception et à la mise en œuvre du devis d'évaluation (Figure 6, p. 47) :

Activités évaluatives		Exemples de choix d'ordre méthodologique
Décisions engagées dans la définition du cadre d'évaluation *(notamment avec l'appui du modèle logique et des parties prenantes clés)*	**1. Déterminer les objectifs principaux de l'évaluation**	(1) *L'utilisation des résultats afin de* contribuer à la compréhension d'un domaine d'intervention donné, étayer la prise de décisions en ce qui a trait à l'amélioration d'un programme, etc.; (2) *l'utilisation du processus afin de* favoriser le développement d'une culture évaluative au sein de l'organisation, assurer l'imputabilité et la transparence de l'organisation, etc.
	2. Déterminer la fonction de l'évaluation, le type d'évaluation et l'approche évaluative	(1) *La fonction* : diagnostic, formative, sommative, etc. ; (2) *le(s) type(s) d'évaluation (objets ou variables)* : besoin, pertinence, processus, efficacité, performance, etc. ; (3) *l'approche évaluative* : développementale, réaliste, participative, axée sur la reddition de compte, etc.
	3. Définir les questions principales d'évaluation	(1) *Générer les questions d'évaluation* : auprès des parties prenantes clés, à partir du modèle logique, etc. ; (2) *déterminer les questions principales* : catégoriser les questions, identifier les plus pressantes, etc. ; (3) *finaliser les questions principales d'évaluation* : valider la liste de questions auprès des parties prenantes clés, vérifier la clarté et la précision des questions, etc.
	4. Développer les matrices d'évaluation	(1) *Les approches méthodologiques envisagées* : étude de cas, devis quasi expérimental, non expérimental, mixte, etc. ; (2) *les indicateurs de mesure préliminaires* : informations que l'on doit obtenir pour répondre aux questions d'évaluation ; (3) *les méthodes de collecte de données envisagées* : instrumentation possible en fonction des (4) *sources d'information envisagées*.
	5. Finaliser le cahier des charges avec les parties prenantes clés	(1) *Obtenir la rétroaction des parties prenantes clés* ; (2) *évaluer la faisabilité* : intrants disponibles, échéancier, accès aux données, etc. ; (3) *finaliser et valider le cadre d'évaluation* : soumission et approbation du cahier des charges, etc.
Décisions engagés dans la conception du cadre méthodologique de l'évaluation *(notamment avec l'appui du modèle logique, des matrices évaluatives et des parties prenantes clés)*	**6. Définir les sources d'information, les échantillons et les procédures de recrutement** *(pour chaque collecte de données)*	(1) *Les sources d'information disponibles* : données secondaires, documentation spécialisée, parties prenantes, clientèle, etc. ; (2) *la population ciblée et les critères de sélection* ; (3) *le type d'échantillon* : probabiliste (aléatoire simple, aléatoire stratifié, etc.), non probabiliste (par quotas, boule de neige, etc.), de convenance (sources disponibles, etc.) ; (4) *les procédures de recrutement* : par l'évaluateur, par l'entremise des gestionnaires, par courriel, par téléphone, par les pairs, etc.
	7. Définir les indicateurs de mesure	(1) *Les mesures de processus* : les composantes du programme, les mécanismes du programme, les résultats intermédiaires attendus, etc. ; (2) *les mesures de rendement* : l'efficacité (l'obtention des résultats escomptés par rapport à la population cible et la portée du programme), l'efficience et l'économie (l'utilisation des extrants, la relation coûts-bénéfices), etc.
	8. Déterminer les méthodes de collecte de données et développer l'instrumentation *(s'assurer que plusieurs méthodes permettent de répondre à une même question d'évaluation)*	(1) *Les procédures de collecte de données* : Où ? Quand ? Pourquoi ? Comment ? Par qui ? (2) *le recours à de multiples méthodes (quantitatives et qualitatives)* : au moment du recueil des données (collectes simultanées ou séquentielles), au moment de la conception du devis d'évaluation (devis mixte intégré ou combinaison de composantes qualitatives et quantitatives),et ce, pour enrichir, pour expliquer, pour trianguler, etc. (3) *la collecte de données quantitatives* est centrée sur l'acquisition de connaissances, de compétences, les changements d'attitudes, les expériences, etc. *Exemple d'instrumentation* : questionnaire, protocole d'observation, épreuves, échelle d'auto-évaluation, etc. ; (4) *la collecte de données qualitatives* est centrée sur les perceptions, les expériences, la compréhension des comportements, la compréhension du programme, etc. *Exemples d'instrumentation* : grille d'observation, carnet de bord, etc. (observation participante), guide d'entrevue, questionnaire (question ouverte), etc. ; (5) *la conception des instruments* : recension des instruments existants (procédures de sélection, d'adaptation, de traduction, etc.), création d'un instrument (sélectionner le type d'items, formuler et répartir les items, etc.), procédés de validation (validation de contenu, étude pilote, etc.), etc.
	9. Déterminer les procédures d'analyse et de synthèse des données	(1) *Le traitement des données quantitatives (numériques)* : analyse descriptive univarié (fréquence, moyenne, médiane, etc.) ou multivariée (corrélation, croisement, etc.) (2) *le traitement des données qualitatives (textuelles)* : analyse de contenu, codage thématique, etc. ; (3) *la synthèse des résultats* : matrices de réponses aux questions d'évaluation, leçons apprises, analyse coûts-bénéfices, etc.
	10. Planifier la communication et l'utilisation des résultats	(1) *Plan de communication* ; (2) *formats de communication* ; (3) *accessibilité* ; (4) *formulation des recommandations* : (5) *soutien de l'utilisation*



uation du JCSEE (2011) : utilité, faisabilité, précision et imputabilité en évaluation

Figure 6. *Étayage des choix méthodologiques propres à la conception et à la mise en œuvre du devis d'évaluation*

2.1.4 Synthèse conceptuelle

Lorsque nous nous intéressons à des problématiques importantes en évaluation, Smith (2008) recommande d'identifier dans quelle mesure celles-ci touchent une ou plusieurs dimensions de l'évaluation (la théorie, la pratique, la méthodologie et la profession). Au regard des éléments abordés jusqu'à présent et des défis relatifs à la praxis, nous avons décidé de nous concentrer, dans le cadre de cette recherche doctorale, sur la dimension « méthodologie » et par extension la dimension « pratique » de l'évaluation. Les choix de conception et d'implantation du devis d'évaluation (cadre de l'évaluation et cadre de la méthodologie évaluative) sont l'« épicentre » de la démarche évaluative. Ils visent à de déterminer l'opérationnalisation des concepts à mesurer et les procédures de collecte de données, et ce, en vue d'assurer la validité, la fiabilité et l'utilité des données recueillies. Les choix dits méthodologiques et la rigueur de la démarche évaluative sont ainsi garants de la crédibilité des résultats, des conclusions et des recommandations produites par l'évaluation de programme (Rog, 2005). Face à de tels enjeux, nous désirons plus particulièrement identifier l'influence que certains facteurs contextuels peuvent exercer sur l'exercice évaluatif et l'opérationnalisation des choix méthodologiques des évaluateurs (Figure 7).

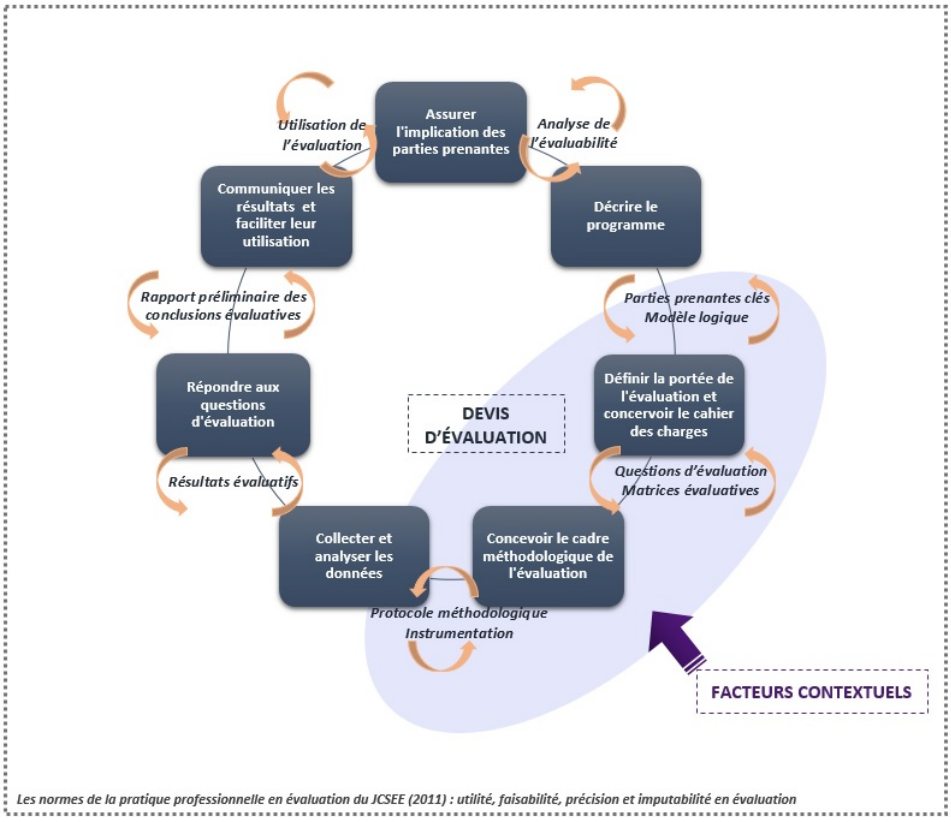


Figure 7. *Synthèse conceptuelle [1/3]*

2.2 La dimension sociopolitique de l'évaluation et la gestion des contraintes ESR, des facteurs contextuels au cœur des enjeux méthodologiques

Maintenant que nous avons introduit la question de l'influence de l'environnement évaluatif sur le déroulement de l'évaluation en général, et la conception du devis d'évaluation et l'opérationnalisation des choix dits méthodologiques en particulier, il est temps de revenir en détail sur certains facteurs et leurs influences sur l'activité évaluative. Une fois que nous définissons le concept de partie prenante, nous abordons : (1) le mouvement de l'utilisation et la raison d'être de l'implication des parties prenantes clés dans l'activité évaluative (2) les enjeux méthodologiques (mésévaluation ou *misevaluation*) et éthiques (mésusage ou *misusage*) (3) les contraintes inhérentes au terrain (l'évaluation en situation réelle [ESR]), éléments contextuels intrinsèquement reliés à la place des parties prenantes au cœur de l'activité évaluative et (4) les enjeux de formation face à la dimension sociopolitique de l'évaluation.

2.2.1 Le rôle des parties prenantes clés et les enjeux méthodologiques face à la dimension sociopolitique de l'évaluation

2.2.1.1 Définition du concept de partie prenante

Le concept de partie prenante (*stakeholder*) fut utilisé pour la première fois au début des années 1960, dans le domaine du conseil en gestion (par l'Institut de recherche de Stanford), pour décrire les individus qui ne sont pas des actionnaires d'une entreprise (*stockholders*), mais qui demeurent indispensables pour son fonctionnement, voir même son existence (Bryson, Patton et Bowman, 2011). Le domaine de l'évaluation l'emprunta par la suite, et son utilisation fut « légitimée » par Freeman (1984) et son approche en gestion centrée sur les parties prenantes. Il considère alors comme partie prenante, tout groupe ou individu qui peuvent exercer une influence sur la réalisation des objectifs d'une organisation (ou être affecté par cette dernière).

Si aujourd'hui les chercheurs et les praticiens en évaluation reconnaissent l'importance de porter une attention particulière aux parties prenantes dans le cadre d'activités évaluatives, ils ne se s'entendent pas nécessairement sur les caractéristiques définissant ces individus ou groupes d'individus. Or, ces caractéristiques ont une influence considérable sur « ce » et « ceux » qu'il importe de prendre en compte pour mener à bien une évaluation (Alkin, Hofstetter et Ai, 1998). En outre, selon les écoles de pensées (Tableau 2), on trouve des définitions plus ou moins

inclusives, selon qu'elles se concentrent sur les parties prenantes du programme (Scriven, 2007; Rossi, Lipsey et Freeman, 2003), ou sur les parties prenantes de l'évaluation (Bryson, Patton et Bowman, 2011 ; Patton, 2008).

Tableau 2. *Le continuum de définitions du concept de partie prenante*

	Partie prenante du programme		**Partie prenante de l'évaluation**
Définitions	« *A stakeholder is someone who has made a significant investment in the program, either financial or psychological*. » (Scriven, 2007, p. 483)	« *People who have a stake or a vested interest in the program, policy or product being evaluated (...) and therefore also have a stake in the evaluation.* » (Greene, 2005b, p. 397)	« *Individuals, groups, or organizations that can affect or are affected by an evaluation process and/or its findings*. » (Bryson, Patton et Bowman, 2011, p. 1)
Individus, groupes ou organismes concernés	Les membres du personnel du programme et les organismes financiers *Remarque : Scriven (1997) fait la distinction entre les clients de l'évaluation, les utilisateurs du programme, les parties prenantes du programme et le public concerné par le programme. Cependant, il reconnait que ces concepts ne s'excluent pas mutuellement.*	Les parties prenantes se répartissent en quatre groupes : - **les personnes ayant une autorité décisionnelle au niveau du programme** incluant d'autres autorités politiques, les financeurs, et les conseils consultatifs - **les personnes ayant une responsabilité directe au niveau du programme** incluant les créateurs du programme, les administrateurs, les gestionnaires et les autres acteurs impliqués dans le programme (le personnel, etc.) - **les bénéficiaires du programme** incluant leurs familles et leurs communautés, - **les personnes désavantagées par le programme**	**+ Les personnes ayant un intérêt direct ou indirect en ce qui concerne l'efficacité du programme**, soit les journalistes, les contribuables et les membres générales du public (Weiss, 1998)

Le recours à des définitions plus inclusives, telles que celles de Greene (2005b) et Bryson, Patton et Bowman (2011), permet à l'évaluateur d'identifier l'ensemble des individus, des groupes ou des organisations concernés par le programme et son évaluation, ce qui constitue une phase initiale importante de l'activité d'évaluation puisqu'elle facilite (entre autres) : (1) la compréhension et la définition de la théorie du programme ; (2) l'identification des différents

groupes de parties prenantes ; (3) l'identification des différentes sources d'information disponibles (et de potentielles contraintes) ; et (4) la sélection des parties prenantes clés. Le dernier point touche un aspect important des pratiques actuelles en évaluation de programme, pratiques qui nécessitent l'implication de certaines parties prenantes dans le processus évaluatif, et ce, qu'on adopte une approche participative ou non. Patton (2012) qualifie ces parties prenantes clés de principaux utilisateurs ciblés (*primary intended users*), soit les partenaires privilégiés de l'évaluateur, qui l'assisteront dans le développement, la mise en œuvre et l'utilisation des résultats de l'évaluation. Ce groupe clé comprend principalement les individus ayant une responsabilité directe au niveau du programme[11] (Greene, 2005b). L'évaluateur doit veiller à ce que l'évaluation répond à leurs besoins (*intended use by intended users*) et en contrepartie, du fait de leur engagement dans l'évaluation et l'utilisation des résultats, les parties prenantes ont la responsabilité de faire les choses différemment pour le bien de leur programme et de leur organisation (prises de décisions éclairées, changements de stratégies, changements politiques, etc.).

Enfin, si nous adoptons une définition inclusive du terme partie prenante en évaluation et si nous reconnaissons l'importance des différents groupes considérés comme telle, nous choisissons, dans le cadre de cette étude doctorale, de nous concentrer sur les **parties prenantes clés** soit les principaux utilisateurs ciblés (Patton, 2012; Greene, 2005b). En effet, nous allons voir à présent que l'implication, dans le processus évaluatif, des individus ayant une responsabilité directe au niveau du programme (créateurs du programme, administrateurs, gestionnaires et autres acteurs impliqués dans le programme), constitue un facteur contextuel majeur dans le cadre d'une évaluation en général, et des activités d'ordre méthodologique en particulier. Aujourd'hui, les parties prenantes clés sont plus présentes que jamais au moment de l'évaluation, ce qui représente

[11] Pour Patton (2012), on peut éventuellement inclure parmi les principaux utilisateurs ciblés, les participants du programme, cela dépend de l'approche évaluative utilisée (participative ou non). Nous avons choisi de ne pas les inclure dans notre définition puisque notre recherche ne porte pas sur les pratiques évaluatives de type participatif et que dans les écrits scientifiques les parties prenantes clés font davantage référence aux individus ou groupes d'individus ayant une responsabilité directe envers le programme.

de nombreux avantages, mais aussi un certain nombre d'enjeux (ou risques), en ce qui concerne la validité, la légitimité et la crédibilité du processus et des résultats évaluatifs, et les habiletés politiques et managériales des évaluateurs.

2.2.1.2 Les parties prenantes clés au cœur du mouvement de l'utilisation de l'évaluation

Les préoccupations relatives aux manques d'utilité et d'utilisation des résultats d'évaluation ont généré une pléthore de recherches sur l'utilisation à partir des années 1970, recherches qui ont conduit les praticiens et les chercheurs en évaluation à reconnaitre et accepter la nécessité de tenir compte des parties prenantes clés dans le cadre d'activités évaluatives, et ce, afin de promouvoir l'utilisation, l'utilité et la qualité de l'évaluation ainsi que le développement d'une culture évaluative au sein des organisations. En soutenant l'engagement des parties prenantes clés tout au long de l'activité évaluative, l'évaluateur leur permet de s'approprier les résultats de l'évaluation, de se sentir responsables de leur utilisation ainsi que de reconnaitre et d'attester de leur crédibilité, validité et légitimité (Patton, 2018, 2008 ; 1988 ; Greene, 1988). Aussi, Patton (2012) caractérise l'utilisation comme suit :

> « *[E]valuations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will affect use. Use concerns how real people in the real world apply evaluation findings and experience the evaluation process.* » (p.365-366).

Lorsqu'on parle d'utilisation, on pense automatiquement à l'utilisation des résultats, cependant, il existe différentes formes d'utilisation de l'évaluation (Alkin et King, 2017 ; Patton, 2015a, 2008; Alkin, 2011 ; Taut et Brauns, 2003; Kirkhart, 2000; Weiss, 1998). Pour notre part, lorsque nous évoquons l'utilisation, nous faisons essentiellement référence à :

- *L'utilisation instrumentale* : les résultats évaluatifs sont utilisés directement pour faciliter les prises de décisions relatives au programme et aux actions à entreprendre à la suite de l'évaluation.
- *L'utilisation du processus* : le processus évaluatif est utilisé afin d'amorcer des changements au niveau du programme. Cette forme d'utilité concerne directement les parties prenantes. Leur implication dans l'évaluation leur permet d'acquérir des connaissances sur la démarche évaluative, les décisions d'ordre méthodologique, etc. Ainsi, elles développent une pensée évaluative qui les conduit à s'approprier

l'évaluation et promouvoir sa rigueur. Tout ceci facilite la gestion et l'anticipation des contraintes et changements organisationnels qui pourraient exercer une influence sur l'évaluation en cours, informations que l'évaluateur pourrait manquer si les parties prenantes ne sont pas impliquées.

En outre, nous verrons plus loin que ces types d'utilisation sont de potentielles cibles de mésusage et de mésévaluation, et que l'*utilisation symbolique* s'avère être une forme de détournement de l'évaluation. En effet, il s'agit d'une utilisation de convenance du processus évaluatif où la seule information recueillie est le fait même de conduire une évaluation (sans nécessairement s'intéresser à ce qu'elle a à offrir), et ce, afin de justifier des décisions déjà prises ou de répondre à une obligation d'évaluation.

Pour résumer les éléments caractéristiques du processus en amont de l'utilisation, Alkin et King (2017) proposent la représentation graphique suivante (*Guttman-scale mapping sentence,* p. 438) :

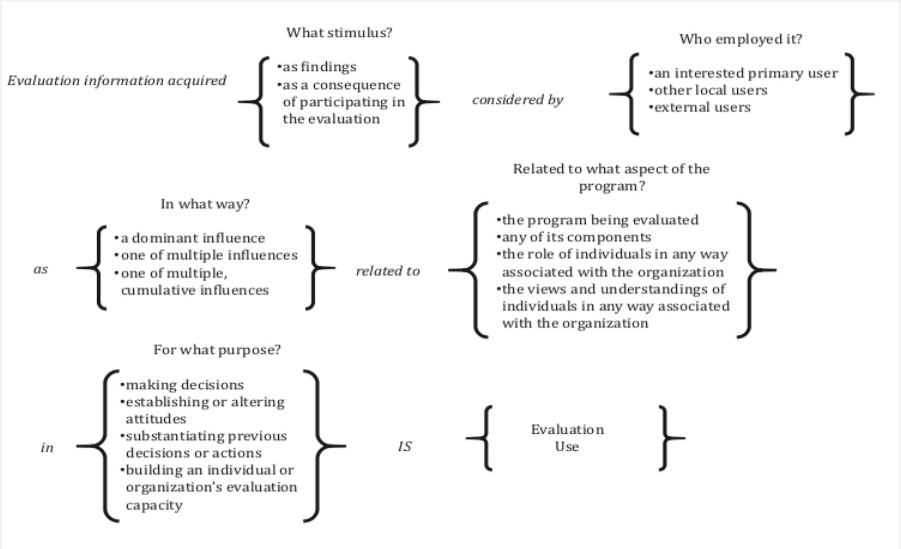


Figure 8. *Conceptualisation de la définition de l'utilisation (Alkin et King 2017)*

De nombreux travaux sur la participation des parties prenantes (Dubois, Bérubé, Lafantaisie, Coutu et Turcotte, 2018 ; Cousins, Whitmore et Shulha, 2013; King, 2007 ; Abma, 2000 ; Cousins et Whitmore, 1998; Cousins et Earl, 1992), ainsi que sur l'utilisation de l'évaluation (Patton, 2018, 2015a, 2008 ; Alkin et King, 2016 ; King et Stevahn, 2013 ; Cousins, 2003 ; Alkin et Coyle, 1988 ; Greene, 1988 ; Cousins et Leithwood, 1986), et la crédibilité de l'évaluation (Marchand, 2020 ; Hurteau et *coll.*, 2012), ont permis la mise en exergue de trois catégories de

facteurs essentiels à la participation des principaux utilisateurs ciblés, à l'utilisation (*intended use by intended users*) ainsi qu'à la validité du processus et des résultats évaluatifs. Ces facteurs sont les suivants : (1) *les caractéristiques et les attitudes des parties prenantes clés (facteurs humains)*[12], soient leurs expériences antérieures avec l'évaluation et leurs potentiels préjugés, leurs attentes envers l'évaluation, leur perception de la crédibilité de l'évaluateur et son capital relationnel, leur anticipation des risques qu'ils encourent à la suite de l'évaluation, etc.; (2) *les caractéristiques du programme évalué et donc de l'environnement évaluatif (facteurs contextuels)*, soient le niveau de maturité du programme, le degré de complexité de la structure organisationnelle et de la hiérarchie des pouvoirs, l'environnement sociopolitique dans lequel le programme est implanté (facteurs contextuels externes), etc.; et (3) *les caractéristiques propres à l'évaluation (facteurs évaluatifs),* soit le devis d'évaluation, la prise en compte des caractéristiques du contexte, la communication des informations au fil du processus évaluatif (le niveau d'importance et la qualité des interactions entre l'évaluateur et les parties prenantes clés), l'utilité et la pertinence des informations évaluatives, les comptes rendus de l'évaluation (fréquence, utilité, format, etc.), etc.

En outre, il est important de noter que ces éléments caractéristiques ont été largement acceptés par la communauté évaluatrice et qu'ils font partie des fondements de la définition des normes du JCSEE (2011)[13], standards que l'évaluateur se doit de respecter. Les enjeux relatifs à l'implication des parties prenantes transparaissent particulièrement au niveau des normes de l'utilité (U) qui veillent à garantir « qu'une évaluation est orientée sur les besoins d'information des utilisatrices et utilisateurs de l'évaluation prévue » (Widmer, Lambert et Bachmann, 2000, p.2). Aussi, nous retenons, par exemple :

U1 – Identification des parties prenantes et des groupes concernés – L'évaluateur doit identifier les personnes qui participent à l'évaluation et les personnes qui sont touchées par cette

[12] Dans certains écrits le facteur humain fait référence aux caractéristiques des parties prenantes clés et des évaluateurs (Fleisher et Christie, 2009; Taut et Alkin, 2003; Alkin, 1985). Cependant ces caractéristiques se basent sur la perception des parties prenantes et non sur leur représentation directe. Pour nous, les caractéristiques de l'évaluateur se trouvent davantage dans la catégorie des facteurs évaluatifs.

[13] Les normes de la JCSEE (2011) sont réparties selon les catégories suivantes : utilité (U), faisabilité (F), précision (P), déontologie (D) et reddition de compte (R).

dernière, et ce, afin de garantir que l'évaluation se déroule dans l'intérêt de ces derniers et qu'elle réponde à leur besoin;

U3 – Crédibilité – L'évaluateur doit montrer qu'il est digne de confiance et qu'il est compétent afin de garantir la faisabilité de l'évaluation ainsi que la crédibilité et l'acceptation des résultats. Pour se faire, il est essentiel qu'il fasse preuve d'intégrité, de sociabilité et d'une excellente capacité d'écoute et de communication;

U8 – Effet de l'évaluation – Le degré d'exploitation des résultats évaluatifs dépend de la perception des parties prenantes clés en ce qui a trait à leur utilité (pour elles et pour le programme). Aussi, dans la lignée des propos de Patton (2008) selon lesquels l'évaluation doit être jugée pour son utilité et par son utilisation, l'évaluateur doit s'assurer d'inclure les utilisateurs ciblés dans la planification et la mise en œuvre de l'évaluation et communiquer régulièrement avec ces derniers.

La prise en compte des attentes et des perceptions des parties prenantes clés en ce qui a trait à l'utilité et à la crédibilité de l'évaluation (et de l'évaluateur) apparait primordiale pour encourager leur implication tout au long de l'activité évaluative; et cette même relation est essentielle pour garantir la faisabilité, la qualité et la validité de la démarche évaluative. En effet, une collaboration plus ou moins étroite, et efficace, avec les parties prenantes clés facilite la compréhension du programme et du contexte évaluatif, les attentes et besoins, l'identification des contraintes de terrain, la conception et la mise en œuvre du devis d'évaluation et la formulation de recommandations utiles et utilisables (Bamberger et Mabry, 2020 ; Chen, 2015 ; Ridde et Dagenais, 2012 ; Bryson, Patton et Bowman, 2011 ; Patton, 2008 ; Morris, 2002 ; Brandon, 1998). Bien entendu, cette liste n'est pas exhaustive, mais ce qu'il est primordial de retenir c'est que la validité et la fiabilité du processus évaluatif et des inférences faites à partir des données collectées ainsi que la représentativité des résultats évaluatifs dépendent de l'évaluateur et de son degré de compréhension du programme et de son environnement. Les parties prenantes clés, expertes du programme évalué, sont des partenaires indispensables pour l'évaluateur, et si ce dernier choisit de ne pas tenir compte de leurs perspectives, leurs conseils, leurs besoins, etc., il met alors en péril son évaluation (Bryson, Patton et Bowman, 2011).

En outre, si la rigueur méthodologique de l'évaluation (indicateur de la crédibilité des résultats) constitue l'un des facteurs les plus importants pour favoriser l'implication des parties prenantes dans le processus évaluatif et l'utilisation des résultats, ce facteur ne semble pas être

nécessairement le plus déterminant. En effet, selon Alkin (2005), les caractéristiques personnelles de l'évaluateur, sa personnalité, ses compétences sociales, son aptitude à faire valoir la parole des parties prenantes clés, à les impliquer dans le processus évaluatif constituent la clé de l'utilisation; et de la crédibilité de l'évaluation (Marchand, 2020 ; Dubois, Bérubé, Lafantaisie, Coutu et Turcotte, 2018 ; Hurteau et *coll.*, 2012 ; Hurteau et Houle, 2008). La qualité, la crédibilité et l'utilité de l'évaluation repose donc également sur la capacité de l'évaluateur de programme à établir et maintenir une dynamique relationnelle constructive avec les parties prenantes clés, un dialogue de qualité, à faire preuve de flexibilité et d'adaptabilité, d'adopter une posture de conseiller. Ceci place sur le devant de la scène le volet de compétence *relations interpersonnelles* chez l'évaluateur qui inclus la collaboration, la communication, la négociation, la gestion et la résolution de conflit (Marchand, 2020 ; Dumaine, 2018 ; Patton, 2018 ; King et Alkin, 2016).

En guise d'introduction de la prochaine section portant sur les revers de l'implication des parties prenantes clés, nous vous présentons les résultats d'une recherche sur les perceptions des parties prenantes en ce qui a trait aux barrières de l'utilisation (ou de la mise en œuvre de pratiques évaluatives efficaces).

La perception des parties prenantes clés en ce qui a trait à l'utilisation

Taut et Alkin (2003, 2002) se sont intéressés aux perceptions des parties prenantes clés quant à la nature des barrières à la mise en œuvre efficace d'évaluations de programme (externes). Pour ce faire, ils se sont inspirés des recherches empiriques sur l'utilisation de l'évaluation et ils ont décidé d'appliquer les facteurs ayant le plus fort degré d'association avec l'utilisation, pour décrire et expliquer les barrières freinant la mise en œuvre d'activités évaluatives. L'étude s'est déroulée dans le cadre de l'évaluation externe d'un programme universitaire d'intervention dans des écoles secondaires. Les auteurs ont réalisé des entretiens semi-dirigés auprès de 18 parties prenantes clés du dit programme[14]. Dans le cadre de cet entretien, les participants ont été invités, dans un premier temps, à faire part de leur réflexion en ce qui a trait aux barrières freinant la mise

[14] Participants de l'étude : quatre professeurs et quatre administrateurs ayant un poste haut placé au sein de l'université, un membre du personnel administratif occupant une position beaucoup moins élevée, six membres responsables de l'implantation du programme et trois employés de la commission scolaire.

en œuvre d'une évaluation, puis dans un deuxième temps, à partager leur point de vue en ce qui a trait aux différents facteurs provenant de la théorie qui pourraient expliquer la manifestation de certaines barrières durant une évaluation externe.

Lorsque les participants ont été invités à faire part, de manière spontanée, de leurs perceptions en ce qui a trait aux barrières à la réalisation efficace d'une évaluation de programme, ils semblaient faire majoritairement référence aux habiletés sociales de l'évaluateur et à sa capacité à mettre en confiance les parties prenantes clés (*facteurs humains*). Cependant, durant la discussion sur leurs expériences aux regards des caractéristiques des trois types de facteurs de l'utilisation, les participants ont cette fois-ci fait davantage référence aux composantes méthodologiques de l'évaluation. Et plus particulièrement, à la prise en compte des parties prenantes afin d'assurer, entre autres, la qualité et l'utilité de l'évaluation (*facteurs évaluatifs*). En outre, ils ont toujours autant fait référence aux facteurs humains, mais cette fois-ci de manière plus éclairée en ce qui a trait aux barrières concrètes de l'utilisation et à la relation entre les différents facteurs.

Les résultats mettent en exergue quatre barrières à la mise en œuvre efficace d'évaluations de programmes externes : *Barrière numéro 1* - les parties prenantes clés accorderaient plus d'importance aux compétences sociales de l'évaluateur et à sa connaissance du contexte évaluatif et du programme qu'à ses compétences techniques. De fait, la construction d'une relation de confiance entre l'évaluateur et les parties prenantes clés est déterminante pour la bonne conduite d'une évaluation. *Barrière numéro 2* - le manque de confiance des parties prenantes représente pour elles un des freins les plus importants pour l'évaluation, puisqu'il génère de la résistance face à l'évaluation qui se traduit par des problèmes d'accès aux données, l'impossibilité de développer une méthodologie de qualité, le recueil de données non pertinentes, la non-utilisation des résultats évaluatifs, etc. *Barrière numéro 3* - le manque de ressource, et principalement le manque de temps et de disponibilité du personnel du programme, est un facteur contextuel important qui a une influence sur la bonne conduite d'une évaluation. *Barrière numéro 4* - que ce soit en raison de problèmes de confiance ou à cause des contraintes inhérentes au terrain, la possibilité de faire des inférences pertinentes et valides à partir des données recueillies a souvent été mentionnée comme un facteur influençant le processus d'évaluation.

Enfin, les chercheurs sont conscients des limites de leur étude qui, soulignons-le, est exploratoire. Les participants faisaient partie d'un programme bien spécifique, auquel ils font référence dans leurs entrevues. Les données ne sont pas généralisables, cependant elles rejoignent

les expériences relatées dans les écrits scientifiques et les éléments clés que nous avons exposés plus tôt dans la section sur l'utilisation. Les points saillants de leur analyse rejoignent les menaces à l'utilité listées par Patton (2008), à savoir, les évaluations qui ne sont pas toujours centrées sur l'utilisation des résultats par les parties prenantes clés, la pauvre compréhension des parties prenantes en ce qui a trait au processus évaluatif et aux résultats, la faible crédibilité des évaluateurs ainsi que l'échec de ce dernier au niveau de l'implication des parties prenantes clés et de la communication des informations dans le courant de l'évaluation. On retrouve également des éléments clés des résultats de l'étude de Marchand (2020) sur la crédibilité de l'évaluation au regard des parties prenantes. Soit autant d'éléments qui ont une influence majeure sur la rigueur de la démarche évaluative ainsi que la validité des procédures méthodologiques et des résultats évaluatifs.

2.2.1.3 Les revers du mouvement de l'utilisation et de l'implication des parties prenantes clés

Dans cette section, nous explorons brièvement des travaux en psychologie de l'évaluation portants sur l'attitude potentiellement réfractaire des parties prenantes clés face à l'évaluation, un signe d'anxiété auquel les évaluateurs de programme sont très souvent confrontés. Ces travaux sont particulièrement intéressants puisqu'ils mettent en exergue un certain nombre de sources et de conséquences liées à ce type de comportement défensif, particulièrement la résistance à l'évaluation. En outre, c'est ce qui nous intéresse le plus ici, puisqu'ils facilitent l'identification de différents indicateurs qui nous permettent plus tard (dans le cadre de l'analyse et de la discussion des résultats) d'observer dans quelle mesure, dans une situation évaluative donnée, certains aspects méthodologiques et certaines décisions évaluatives sont liés à la potentielle résistance des parties prenantes clés. Aussi, nous abordons ensuite les risques de mésévaluation, de non-utilisation et de mésusage du fait notamment de l'implication importante des parties prenantes clés dans l'évaluation et son utilisation. Enfin, nous discutons des enjeux de formation des évaluateurs en raison de la dimension sociopolitique de l'activité évaluative.

2.2.1.3.1 L'attitude réfractaire des parties prenantes clés face à l'évaluation

Par essence, l'anxiété éprouvée face à l'évaluation de programme est inévitable, cependant on observe depuis ces vingt dernières années une croissance au niveau de sa prévalence et de son

importance (Geva-May et Thorngate, 2003; Donaldson, Gooler et Scriven, 2002). Ce phénomène semble coïncider avec l'apparition des approches pragmatistes et l'influence déterminante de ces dernières sur les pratiques évaluatives actuelles (Donaldson et Scriven, 2002). Aujourd'hui, les parties prenantes ne sont plus de simples spectatrices de l'évaluation, certaines d'entre elles ont un rôle déterminant à tenir et elles interagissent beaucoup plus fréquemment avec les évaluateurs de programme. Ainsi, la crainte d'une évaluation négative, autrefois un évènement anxiogène qui se manifestait seulement au moment de la remise du rapport final, fait désormais partie de leur quotidien et cela peut entraîner l'apparition de signes plus ou moins excessifs d'anxiété. Parmi les signes les plus fréquents se trouvent la résistance, la méfiance ainsi que l'embarras, ce qui peut engendrer des complications de différents degrés au niveau du processus évaluatif et constituer un obstacle important à la fiabilité, la validité, l'utilité et l'utilisation des résultats évaluatifs (Geva-May et Thornagate, 2003; Taut et Brauns, 2003; Donaldson, Gooler et Scriven, 2002; Donaldson, 2001).

Dans le cadre de leurs travaux, Donaldson, Gooler et Scriven (2002) se sont intéressés à la psychologie de l'évaluation de programme et ils ont tenté de mettre en exergue les sources, les signes et les conséquences de ce qu'ils qualifient d'anxiété évaluative excessive. De leur côté, Taut et Brauns (2003) se sont plus particulièrement penchés sur la question de la résistance des parties prenantes clés, c'est-à-dire quand ces derniers retardent intentionnellement le processus évaluatif, qu'ils protestent, qu'ils se désengagent ou qu'ils n'utilisent pas les résultats probants de l'évaluation (résistance au changement). La résistance est d'ailleurs l'un des signes d'anxiété mis en lumière par les travaux de Donaldson, Gooler et Scriven (2002), et elle s'avère être l'un des phénomènes les plus récurrents auxquels les évaluateurs sont confrontés. Principalement étudié en termes de manque d'utilisation des résultats évaluatifs, le phénomène de résistance a été rarement observé au regard de l'influence de certains facteurs psychologiques à l'origine de l'apparition de comportements réfractaires, et pourtant, les évaluateurs doivent être en mesure de reconnaitre et de gérer la manifestation de tels comportements afin de collaborer efficacement avec les parties prenantes clés qui sont résistantes (Taut et Brauns, 2003; Donaldson, Gooler et Scriven, 2002; Monsen, 2002; Datta, 2001).

Ainsi, à partir d'écrits scientifiques en évaluation de programme et en psychologie sociale et des organisations, Taut et Brauns (2003) ont formulé des vignettes descriptives dérivées de la théorie : *[construit psychosocial-incidence sur la résistance des parties prenantes]*. Dans le cadre

de leur étude, ils se sont intéressés aux perceptions de 21 évaluateurs experts en ce qui a trait à la résistance des parties prenantes clés face à l'évaluation. Les participants ont été invités à : (1) partager leurs propres explications en ce qui concerne les situations de résistance; (2) lire les vignettes dérivées de la théorie et donner pour chacune d'entre elles un exemple issu de leurs propres expériences; et enfin (3) évaluer la clarté et la pertinence des vignettes au regard de leurs propres expériences. Les expériences relatées par les experts interrogés ainsi que les différentes explications qu'ils ont générées à propos de la résistance des parties prenantes rejoignent particulièrement les concepts de *contrôle*, *d'attitude envers l'évaluation* (expériences antérieures) et de *pouvoir*. Au regard du croisement des résultats de cette étude avec les éléments mis en lumière par les travaux de Donaldson, Gooler et Scriven (2002), voici ce qu'il faut selon nous retenir :

L'attitude envers l'évaluation semble être le facteur le plus déterminant en ce qui a trait à l'apparition de signes défensifs face à l'évaluation. En effet, le manque d'expérience avec l'évaluation de programme, les expériences antérieures négatives, l'adoption des préjugés d'autres individus, la crainte des répercussions négatives de l'évaluation, l'attachement au programme, etc. (facteurs dispositionnels) peuvent conduire les parties prenantes clés à anticiper la manifestation de conséquences négatives, ce qui génère un état émotionnel négatif connu sous le nom de réactance. Parmi ces conséquences anticipées, on trouve : (1) la perte de contrôle (la perception d'une restriction de leur liberté, le sentiment d'insécurité face aux conséquences de l'évaluation); (2) la perte de pouvoir suite (la modification de la structure de pouvoir existante, et par conséquent, la légitimité de leur position et l'étendue de leur influence sur leur environnement est menacée); (3) les effets du jugement négatif (estime de soi [l'appréciation de la performance du programme est intrinsèquement reliée à leur estime], la comparaison sociale et la manière dont ils sont traités par leurs supérieurs). *Les caractéristiques de l'environnement évaluatif (facteurs situationnels)* telles que les normes sociales et la culture organisationnelle du milieu, l'ambiguïté des rôles entre l'évaluateur et les parties prenantes, le manque de reconnaissance des accomplissements du programme, etc. exercent également une influence sur l'attitude réfractaire des parties prenantes clés. Aussi, la Figure 9 suivante présente les potentielles répercussions sur l'évaluation.

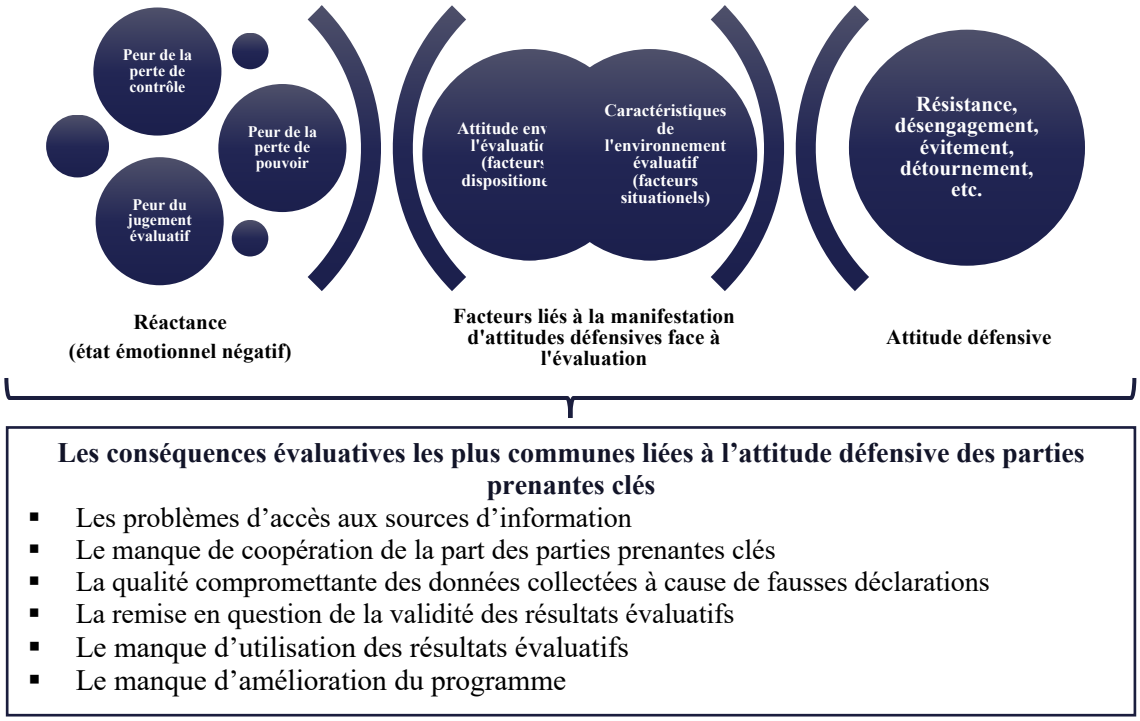


Figure 9. *Les répercussions de l'attitude réfractaire des parties prenantes clés sur l'évaluation*

La place centrale qu'occupe aujourd'hui les parties prenantes au sein de l'évaluation est la résultante des changements au niveau de la pratique de l'évaluation de programme et de la nécessité d'entretenir un climat de collaboration entre les évaluateurs et les utilisateurs ciblés de l'évaluation afin de gérer, entre autres, la résistance du milieu face à l'évaluation. Les parties prenantes clés ont dorénavant la possibilité de communiquer leurs inquiétudes, leurs interrogations, leurs doutes et les évaluateurs se doivent de les écouter, de les considérer et de faire des compromis. Cependant les évaluateurs sont-ils nécessairement équipés pour le faire? Les parties prenantes sont-elles toutes ouvertes à l'évaluation et à la collaboration avec les évaluateurs? Nous l'avons mentionné dans le premier chapitre, l'évaluation de programme est encore loin de faire l'unanimité auprès des organisations et du public en général (EvalAgenda2020). Or, la culture organisationnelle du milieu évaluatif, les expériences passées des parties prenantes clés, les préjugés, le désir de protéger le programme, etc. sont des sources importantes d'anxiété et de résistance face à l'évaluation. Mal gérées, ces situations peuvent se dégrader très rapidement et déclencher une série de complications majeures, particulièrement pour la rigueur méthodologique, la qualité et l'utilité des résultats de l'évaluation.

Enfin, face à l'omniprésence des parties prenantes clés au sein de l'activité évaluative, et des auteurs, tels que Patton (2008), nous avisent que l'augmentation de l'utilisation du processus et des résultats évaluatifs signifie également l'augmentation des risques de mésévaluation (*misevaluation*) et de mésusage (*misuse*). De même, Donaldson, Gooler et Scriven (2002) soulignent que la multiplication de nouvelles approches centrées intensément sur la satisfaction des parties prenantes clés se fait au détriment de la rigueur méthodologique de l'évaluation.

2.2.1.3.2 Les risques de mésévaluation, de non-utilisation et de mésusage

Face à l'abondance des travaux sur l'utilisation, Alkin (2011, 2005 et 1990) a mis en garde la communauté évaluatrice contre le mésusage de l'évaluation. Certes, ces deux concepts diffèrent et existent sur deux continuums distincts, cependant ils demeurent connectés. L'un se concentre sur le degré d'utilisation (*utilisation – non-utilisation*) et l'autre sur la manière dont l'évaluation est utilisée et la dimension éthique (*usage approprié – mésusage*). Avant d'aborder plus en détail les différentes formes de mésusage, il est important d'aborder le concept de mésévaluation (« *misevaluation* ») qui fait référence à une évaluation de mauvaise qualité (remise en question de la validité du processus et des résultats évaluatifs), et qui peut conduire au mésusage. La mésévaluation peut survenir pour plusieurs raisons :

- Le manque de qualité au niveau de la méthodologie évaluative (collectes de données mal menées, analyses statistiques incorrectes, etc.), la mauvaise compréhension (et appréhension) du contexte évaluatif qui conduit l'évaluateur à détourner l'évaluation, le manque de collaboration de l'évaluateur avec les parties prenantes clés qui détiennent les informations dont il a besoin, etc. (Alkin, 1990).
- Le manque de compétence de l'évaluateur, les méfaits de l'évaluateur, la manifestation de circonstances imprévues qui engendrent la détérioration du devis évaluatif, etc. (Cousins, 2004).
- Le recours à des devis méthodologiques inappropriés, ce qui entraîne le recours à des méthodes de collectes de données incorrectes ou l'absence de résultats probants sur lesquels baser des conclusions valides et fiables (Alkin et King, 2017).
- L'évaluateur se fait manipuler par les gestionnaires de programme qui lui demandent d'altérer les résultats évaluatifs négatifs, ou laisse son appréciation (attachement) pour le programme

évalué et le public ciblé influencer ces décisions (Yarbrough, Shulha, Hopson et Caruthers, 2011).

Lorsque la validité et la crédibilité d'une évaluation sont remises en question, cela peut conduire à la *non-utilisation justifiée* (si les utilisateurs jugent de la mauvaise qualité de l'évaluation ou si l'évaluateur reconnait les faiblesses importantes de l'évaluation), à la *mésévaluation* (si les utilisateurs ne sont pas informés ou n'ont pas remarqué les problèmes de validité, etc., et utilisent l'évaluation), ou au *mésusage* (si les utilisateurs sont au courant du manque de validité du processus et des résultats évaluatifs et qu'ils décident de tout de même les utiliser, ou s'ils ont eux-mêmes « saboté » l'évaluation). Le concept de mésusage concerne les manquements éthiques des clients de l'évaluation, soit les utilisateurs de cette dernière (les parties prenantes clés). On parle de mésusage lorsque l'évaluation n'est pas utilisée pour les bonnes raisons ou est utilisée de manière inappropriée, par exemple lorsque les gestionnaires de programme manipulent les résultats ou le processus évaluatif à des fins personnelles (Alkin et King, 2017). La *non-utilisation d'une évaluation de bonne qualité* engendre des préoccupations éthiques et constitue potentiellement un mésusage, notamment lorsque les parties prenantes clés ignorent délibérément des résultats évaluatifs, et ce, sans que leur qualité soit remise en cause ou sans l'appui d'autres sources d'informations.

Le mésusage peut se manifester à différentes étapes de l'évaluation : au moment de la définition du mandat d'évaluation, au moment de la mise de l'évaluation, ou encore au moment de la gestion et de la transmission des résultats d'évaluation (Christie et Alkin, 1999). Afin de faciliter la description du mésusage, nous avons créé une représentation graphique, Figure 10 page suivante (adaptée et librement inspirée des travaux de Alkin et King [2017], Patton [2015a], Alkin et Coyle [1998], Alkin [1990]).

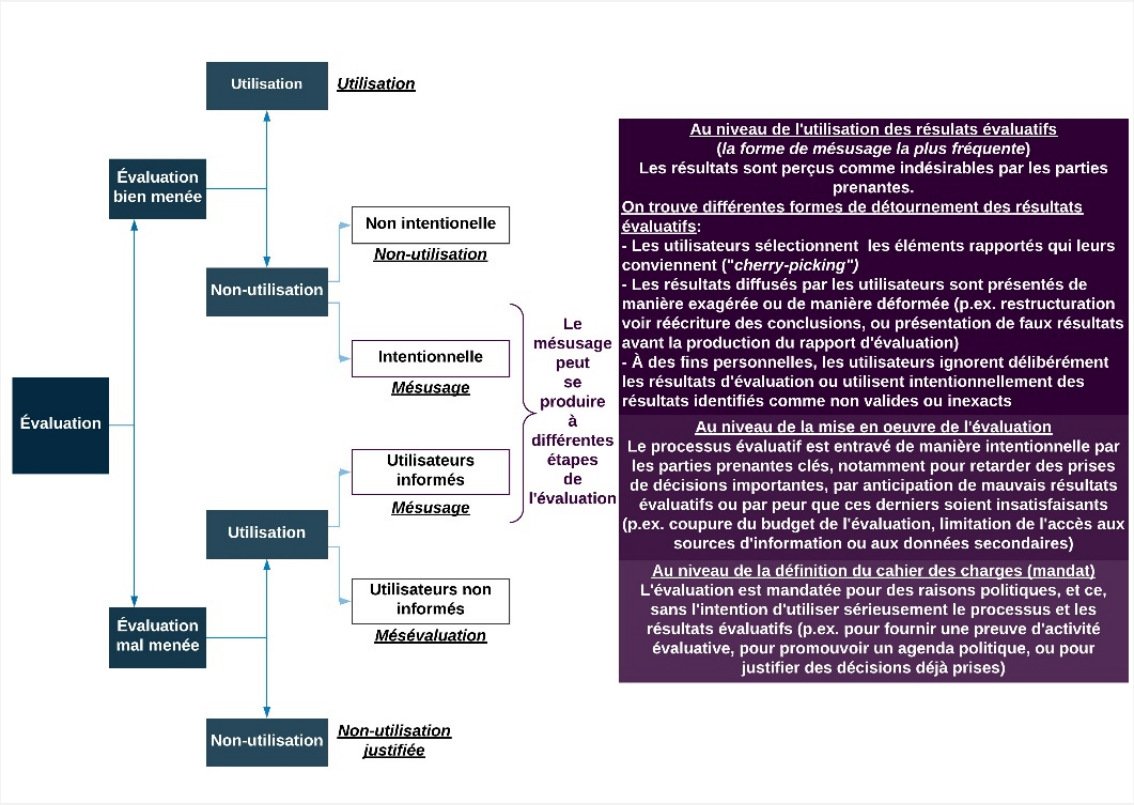


Figure 10. *Représentation graphique du mésusage de l'évaluation*

Les préoccupations relatives à l'utilité et l'utilisation de l'évaluation (son processus et ses résultats) se trouvent au cœur de la pratique en évaluation, faisant de l'utilisation un concept central dans de nombreux modèles et approches en évaluation de programme. Or, « *as use increases, misuse will also increase* » (Patton, 2008, p. 105). En effet, nous l'avons vue plus tôt, en plaçant les parties prenantes clés au centre de l'évaluation pour favoriser l'utilisation, nous ouvrons dans le même temps la porte aux abus. Tout comme l'utilisation, le mésusage est fondamentalement situationnel, et la place de choix des parties prenantes clés génère de nombreuses sources de mésusage telles que les pressions politiques de personnes influentes, la mésévaluation intentionnelle, la falsification des résultats, la priorisation des intérêts personnels, etc.

Pour Patton (2018, 2015a, 2008), les alliés de l'utilisation sont les alliés de l'évaluateur contre le mésusage. En effet, collaborer avec des utilisateurs ciblés (parties prenantes clés) qui comprennent et reconnaissent les mérites de l'évaluation de programme constitue l'un des meilleurs moyens de contenir les risques de détournement de l'évaluation. Plus on les implique, plus ils s'approprient l'évaluation et deviennent des modèles d'éthiques en matière d'utilisation de l'évaluation, voire des gardiens de sa crédibilité, sanctionnant ainsi toutes actions de mésusages.

Une fois de plus, la solution la plus pertinente pour améliorer les pratiques en évaluation est de placer les parties prenantes clés au cœur de l'activité évaluative, de son utilisation et de la protection de sa qualité et de sa crédibilité. Nous nous trouvons pris dans un cycle sans fin, où l'évaluation et son utilisation deviennent plus que jamais des activités politiques, sources de pression et de négociation au sein de l'environnement évaluatif, mais aussi de la société. Et, si l'évaluateur n'a pas toujours la possibilité de contrôler l'intégrité de l'utilisation, il est désormais de son devoir de monitorer, exposer et empêcher les actes de mésusage.

2.2.2 Les contraintes inhérentes au terrain (ou contraintes ESR)

Sur le terrain, l'évaluateur fait face à de nombreux défis liés aux intrants disponibles pour l'évaluation, dans un contexte où les clients commandent des évaluations avec des contraintes temporelles et budgétaires plus ou moins importantes, et où l'accès à des données probantes de qualités peut s'avérer compliqué (Bamberger et Mabry, 2020 ; Bamberger, Rugh et Mabry, 2012). Il doit également adapter sa pratique évaluative à une gamme variée d'influences et de contraintes d'ordre politique, dans un environnement organisationnel habité par de nombreux acteurs avec des attentes différentes, et où le fonctionnement administratif ne facilite pas la mise en œuvre d'évaluations rigoureuses (Bamberger, Rugh, Church et Fort, 2004 ; Bamberger, 2000). L'évaluateur fait souvent face à de fortes pressions méthodologiques, notamment parce que les parties prenantes clés peuvent avoir des idées bien arrêtées sur ce qui est « juste » en ce qui a trait aux ressources nécessaires, aux méthodes d'investigation, aux objets d'étude, aux sources d'information, à la communication des résultats évaluatifs, etc. Nous l'avons vu plus tôt, il doit également gérer la potentielle résistance des parties prenantes clés face à l'activité évaluative, une situation qui peut engendrer des réactions plus ou moins négatives susceptibles de conduire à la mésévaluation et au mésusage. Cependant, il est primordial de reconnaitre que la complexité de la réalité de terrain (et des contraintes qui l'animent) ne doit en aucun cas empêcher l'évaluateur de demeurer respectueux des standards méthodologiques et éthiques de sa profession.

L'approche de l'évaluation en situation réelle propose des stratégies pour accompagner les évaluateurs dans la gestion des contraintes inhérentes au terrain (ou contraintes ESR) et l'utilisation et la réalisation de méthodologies évaluatives adaptées et rigoureuses. Si à l'origine elle était destinée à la réalisation d'évaluation d'impact dans les pays en développement, ses auteurs soulignent sa pertinence pour tout type de mandat évaluatif, et ce, peu importe

l'environnement socio-économique et politique, du programme évalué (Bamberger et Mabry, 2020 ; Bamberger, Rugh et Mabry, 2012 ; Bamberger, Rugh, Church et Fort, 2004). En effet, bon nombre des problématiques liées à la gestion des contraintes ESR s'avèrent courantes et communes pour l'ensemble des milieux évaluatifs, par exemple lorsque : (1) l'évaluation arrive tard dans la vie du programme et qu'aucune base de données comparative n'est disponible ; (2) le temps et le budget alloués à l'évaluation sont inappropriés ; (3) les pressions politiques liées à la manière dont l'évaluation devrait être réalisée et ce que les résultats devraient dire ; ou encore (4) la résistance de parties prenantes face à l'évaluation qu'ils perçoivent comme une menace.

On retrouve ici des éléments caractéristiques notamment des situations évaluatives rencontrées dans le contexte de l'évaluation de programmes publics du gouvernement fédéral canadien. En effet, les résultats de l'évaluation de la Politique d'évaluation de 2009 (Secrétariat du Trésor du Canada, 2015), font ressortir un ensemble de difficultés, sources de contraintes pratiques et méthodologiques importantes, notamment :

1) l'obligation pour les ministères d'évaluer l'ensemble de leurs programmes de subventions et de contributions, tous les cinq ans, et de répondre à l'ensemble des questions fondamentales de la politique évaluative fédérale sur le rendement de leur programme (efficacité, efficience et économie), et ce, en dépit des problèmes de faisabilité (nature et maturité du programme, échéanciers serrés, budgets insuffisants, etc.)[15] ;
2) les lacunes sur le plan de l'existence, l'accessibilité, la qualité et l'exhaustivité des données de mesure du rendement du programme, et ce, malgré l'existence d'une stratégie de mesure de rendement de programme pour la majorité des projets d'évaluation examinés (périodes couvertes 2012-2013, 2013-2014). Un élément qui répond à l'une des exigences de la Politique visant à renforcer la démonstration

[15] N.B. la Politique sur les résultats de 2016 a assoupli certaines exigences, plus précisément pour les programmes dont les dépenses réelles moyennes pour cinq ans sont inférieures à cinq millions de dollars canadien par an.

rigoureuse du rendement des programmes (fondées sur des données probantes complètes et de qualité) ;

3) les données financières sont la plupart du temps indisponibles ou insuffisantes. Et si elles sont disponibles (et justes, et complètes), elles ne sont pas nécessairement structurées en fonction de l'objet d'évaluation – c'est à dire de manière compatible pour analyser le rapport coût-efficacité pour les différentes activités du programme évalué. Aussi, la démonstration de l'efficience du programme repose souvent sur de faibles éléments probants.

Notons que l'ensemble des difficultés que nous venons d'énoncer avaient fait l'objet des observations du Rapport de mise au point sur l'évaluation de l'efficacité des programmes fédéraux du Bureau du vérificateur général du Canada (2013). Ce rapport soulignait que la nature et l'ampleur des lacunes sur le plan des données de mesure de rendement, limitaient de manière plus ou moins importante la détermination et (ou) la portée de l'évaluation des résultats du programme. Et que, quelles que soient les situations rencontrées, les évaluateurs devaient généralement collecter de nouvelles données probantes ou miser sur l'obtention de données plus subjectives (questions de rappels, questions déclaratives dans le cadre d'enquêtes ou d'entrevues auprès d'informateurs clés, par exemple).

Les questions liées à l'efficacité demandent le recours à des devis sophistiqués qui répondent aux normes conventionnelles pour évaluer les relations de cause à effet entre le programme et les résultats observées. Cependant, la réalité de terrain et les contraintes qui l'occupent, particulièrement sur le plan des données probantes, posent de gros problèmes de faisabilité et de fiabilité (Bamberger et Mabry, 2020; McDavid, Huse et Hawthorn, 2019). Dumaine (2018) rapporte que la mesure des résultats ultimes/de l'impact à l'appui de devis expérimentaux ou quasi expérimentaux est très dur à accomplir, et peu entreprise au gouvernement fédéral, particulièrement lors d'évaluations de programmes. Selon lui, si ce type d'étude doit ou peut être réalisé, cela devrait se faire en marge du processus d'évaluation quinquennale des programmes. Dès lors, les résultats constitueront une source de données probantes lors de l'évaluation du programme. De même, Dumaine (2018) souligne que l'analyse systématique du rapport cout-bénéfices voir du rapport cout-efficacité lors de l'évaluation de programmes fédéraux est une attente irréaliste. La mesure de l'efficience et de l'économie devrait faire l'objet d'un exercice indépendant dont les résultats serviront pour l'évaluation du programme.

Lamari, Poulin-Larivière et Lemieux (2016) soulignent que le calibrage inadéquat de l'évaluation au regard des ressources disponibles et mobilisables est un problème caractéristique de l'évaluation de programmes publics. Aussi, l'exigence de livrer une évaluation fiable et fondée sur suffisamment de données probantes, avec un échéancier souvent trop serré et un budget limité face aux contraintes importantes d'accès aux données, constitue pour les auteurs un phénomène paradoxal qui « entraine un nivellement par le bas des évaluations et limite la collecte de données requises pour répondre aux exigences de rigueur et de validité des évaluations produites. » (p.25).

Ainsi, qu'on se réfère ou non à l'approche ESR, les stratégies adaptatives d'ordre méthodologique proposées par cette dernière font échos avec la réalité professionnelle de l'ensemble de la communauté évaluatrice. Ici, les travaux de Bamberger et Mabry (2020) nous permettent de mettre en exergue les processus décisionnels engagés sur le terrain lorsque les évaluateurs font face à une multitude de contraintes. Les stratégies qu'ils décrivent rendent comptent des actions couramment entreprises par les évaluateurs sur le terrain, actions qui au demeurant sont réalisées de manière plus ou moins rigoureuse. En effet, dans un contexte évaluatif contraignant, ces derniers sont amenés à évaluer et à ajuster constamment leurs choix méthodologiques, entachant parfois la rigueur scientifique de la démarche évaluative. Le tableau suivant (Tableau 3) rend compte des différents choix méthodologiques mis en œuvre par les évaluateurs de programme en réponse aux contraintes ESR :

Tableau 3. *Adaptation des choix méthodologiques en réponse aux contraintes ESR (Bamberger et Mabry, 2020)*

Contraintes ESR	Stratégies adaptatives d'ordre méthodologique
Contraintes budgétaires	1. Simplifier le devis d'évaluation : proposition d'un devis non expérimental (pas de groupe de comparaison, de mesures pré intervention, etc.) 2. Discuter avec les utilisateurs ciblés de la réduction de la quantité d'information à recueillir dans la cadre de la collecte de données 3. Optimiser/privilégier l'utilisation de données secondaires fiables 4. Réduire la taille des échantillons 5. Diminuer le coût de la collecte de données en optant pour des méthodes moins coûteuses (concevoir les questionnaires en ligne, réduire la longueur et la complexité de l'instrument, privilégier les groupes de discussion et les entrevues d'informateurs clés, recourir à de multiples méthodes et à la triangulation des sources d'information, etc.)

Contraintes temporelles	[*En complément des stratégies adaptatives face aux contraintes budgétaires*] 6. Accroître le budget de l'évaluation (si possible) 7. Agrandir l'équipe d'évaluation ou solliciter des professionnels plus expérimentés pour réaliser et (ou) superviser la collecte et l'analyse des données 8. Privilégier l'utilisation de technologies informatiques pour un gain de temps pendant la collecte, la transcription et l'analyse des données
Contraintes d'accès aux données	*En l'absence de données de base (prétest) et de groupes témoins* : 1. Reconstruire le corpus de données de base après coup (données de programme, données secondaires, entrevues d'informateurs clés, items de rappel de situation pré intervention, etc.) 2. Reconstruire un groupe de comparaison pour comparaison « post », ou recourir à la comparaison de groupes non équivalents 3. Trianguler les méthodes, les sources et les évaluateurs *Recueillir des informations sur des sujets délicats ou sur des groupes vulnérables (difficiles à atteindre)* : 4. Privilégier l'observation participante, les enquêtes sociométriques, l'accès aux sources d'information par l'échantillonnage de type boule de neige
Contexte / Contraintes politiques et organisationnelles	1. Répondre aux exigences des organismes commanditaires, aux besoins des utilisateurs ciblés en ce qui concerne la conception de l'évaluation 2. Tenir compte des préférences méthodologiques et modalités de communication de l'évaluation des organismes commanditaires/utilisateurs ciblés 3. Reconnaitre les enjeux organisationnels et les risques qu'ils représentent pour la rigueur méthodologique (complexité du fonctionnement institutionnel, coopération des gestionnaires de programme et de données, etc.)

Enfin, il est essentiel de retenir que la gestion des contraintes ESR exige des compromis importants entre la rigueur des choix méthodologiques de l'évaluateur et les attentes des parties prenantes clés en ce qui a trait au contenu du devis d'évaluation (et ce, souvent en dépit des contraintes temporelles et budgétaires), car la manière dont l'évaluateur va répondre aux contraintes de l'environnement constitue une menace à la validité des résultats et des recommandations de l'évaluation. Revoir le niveau de sophistication du devis d'évaluation « à la baisse » en réponse aux contraintes ESR, ne dois pas justifier le manque de rigueur scientifique, bien au contraire. Selon nous, choisir par exemple un devis non expérimental et le combiner avec l'utilisation de multiples sources de données (méthodes et sources d'information), c'est une option plus rigoureuse, que de tenter de reconstruire un groupe contrôle après coup, alors que les données de base pour les participants de l'intervention ne sont pas non plus disponibles. Le choix de parler, ici, de multiples méthodes et non de méthodes mixtes n'est pas anodin, car si le recours aux méthodes mixtes est la stratégie idéale de l'évaluation en situation réelle et la méthode de choix du mouvement pragmatiste, des problèmes semblent persister sur le terrain. En effet, en milieu gouvernemental, le recours à des évaluations de devis mixtes est à peine plus élevé que l'utilisation

des méthodologies qualitatives, et leurs significations statistiques s'avèrent souvent limitées (Christie et Fleischer, 2010). Bamberger et Mabry (2020) expliquent d'ailleurs travailler sur la proposition d'un modèle et de lignes directrices pour une utilisation réussie des approches mixtes en évaluation en situation réelle. En effet, au début de leurs travaux, les auteurs pensaient que les principales contraintes ESR étaient liées au temps, au budget et à l'accès aux données. Cependant, ils ont rapidement réalisé que les plus gros défis provenaient de la nécessité de tenir compte (et de satisfaire) des différentes attentes et exigences organisationnelles, dans un système administratif souvent rigide et inadapté pour accommoder les stratégies méthodologiques envisagées en réponse aux contraintes ESR (Bamberger, Rugh, Church et Fort, 2004). Or, les approches mixtes préconisent le recours à une démarche prescrite, mobilisant un processus méthodologique spécifique et rigoureux, et ce, dès la définition de la portée de l'évaluation. Aussi, la manifestation constante de pressions sociopolitiques et de contraintes ESR au fil de l'évaluation expliquerait, pour nous, pourquoi en fin de compte l'usage des méthodes mixtes signifie en réalité, l'utilisation de multiples méthodes. Lorsqu'un environnement évaluatif se montre très rapidement contraignant, mobiliser une approche multiméthodes, miser sur la complémentarité des méthodes quantitatives et qualitatives, s'appuyer sur le jugement de l'évaluateur pour déterminer les relations de causalité, et recourir à la triangulation (méthodes, sources, et évaluateurs), sont les stratégies privilégiées pour garantir la validité et la fiabilité des résultats évaluatifs (McDavid, Huse et Hawthorn, 2019 ; Greene, 2015, 2005c). Cela permet, entre autres, de répondre au mieux à ce que Bamberger et Mabry (2020) appellent de « *major headaches* » pour les évaluateurs, à savoir :

- La pression de la part de leurs collègues de ne pas faire de vagues en étant trop négatif ou en soulevant des questions sensibles, des points litigieux.
- La pression de la part des utilisateurs ciblés d'écarter certains problèmes, de ne pas interroger certains groupes qui pourraient décrire le programme sous un angle négatif.
- La difficulté d'obtenir des informations (existence, ampleur et accessibilité des données probantes, accès à la documentation et aux données de programme, faible taux de réponse aux questionnaires, etc.) ou de planifier des entrevues d'informateurs clés avec des parties prenantes clés du programme (gestionnaires, responsables d'intervention, personnel de première ligne, etc.) et autres partenaires institutionnels.
- Etc.

2.2.3 Les enjeux de formation face à la dimension sociopolitique de l'évaluation

En 2009, Fleischer et Christie ont publié les résultats d'une enquête menée auprès des membres américains de la Société américaine d'évaluation afin d'appréhender leurs attitudes, leurs perceptions et leurs expériences en ce qui a trait à l'utilisation de l'évaluation. 1 140 membres ont répondu à cette enquête[16]. Le corps du questionnaire était composé d'items touchant à cinq dimensions clés de l'utilisation développés à partir d'une recension d'écrits scientifiques. La lecture des résultats nous a rapidement permis d'identifier des enjeux de formation liés à la sociopolitisation de l'activité évaluative et à la gestion des contraintes de terrain. De fait, nous proposons une discussion des résultats au regard des éléments mis en exergue jusqu'à présent dans ce chapitre et des résultats d'une étude portant sur les écarts entre la théorie et la pratique dans le cadre de l'expérience de terrain d'évaluateurs en formation (étudiants-évaluateurs) (Chouinard et *coll.*, 2017). Nous sommes conscients que l'étude de Fleischer et Christie date de 2009, mais ses résultats demeurent pertinents et ils font échos aux enjeux contemporains en évaluation de programme, particulièrement en ce qui a trait à la variable « méthodologie » :

(1) *parmi les 15 facteurs déclarés comme ayant une influence importante ou extrêmement importante sur l'utilisation (n=897)*, l'adhésion à des standards de rigueur méthodologique élevés (65%) et la prévention du mésusage des résultats évaluatifs (61%) se trouvent dans le dernier tiers du tableau de résultats. Si ces résultats se situent tout de même au-dessus de la moyenne, ils restent très surprenants compte tenu du fait que la crédibilité de l'évaluation (jugé essentiellement au regard de la qualité méthodologique) constitue l'un des facteurs les plus importants pour

[16] Éléments caractéristiques de l'échantillon : 1) 96% des participants ont déclaré que la pratique de l'évaluation de programme faisait partie de leur identité professionnelle (i.e. primaire ou secondaire), parmi eux 92% conduisaient une évaluation au moment de l'enquête (37% se sont identifiés comme évaluateur externe, 22% comme évaluateur interne et 33% comme évaluateur à la fois interne et externe); 2) 87% des participants se sont classés dans le niveau intermédiaire ou avancé en termes de connaissances et d'expériences en évaluation. 57% ont déclaré qu'ils pratiquent l'évaluation de programme depuis une dizaine d'années ou un peu moins ; 3) Parmi les 96% qui se sont définis comme évaluateur, environ 50% ont rapporté qu'ils travaillaient à la fois sur des évaluations à petites et à grandes échelles, 34% sur des évaluations à petites échelles, et 14% sur des évaluations à grandes échelles. 39% réalisent en moyenne entre une et trois évaluations par an, 30% entre quatre et six par an, et 26% sept (et plus) par an.

promouvoir l'utilisation et apaiser l'anxiété évaluative. En contrepartie, les facteurs ayant trait à la prise en compte des parties prenantes se trouvent en haut du tableau. Bien entendu, nous l'avons vu, ces facteurs influencent autant l'utilisation que la qualité de l'évaluation. Cependant, au regard des autres résultats, il semblerait que les préoccupations méthodologiques aient tendance à être relayées au second plan, ce qui rejoint les propos de Donaldson, Gooler et Scriven (2002) selon lesquels le renforcement de l'adhésion et de la satisfaction des parties prenantes clés se fait souvent au détriment de la rigueur méthodologique et de la qualité des résultats évaluatifs. De même, l'étude de Chouinard et *coll.* (2017) a montré que les étudiants-évaluateurs remarquaient que la priorité était donnée à la dynamique de groupe entre les personnes impliquées dans l'évaluation, la gestion des attentes, la charge de travail, etc., plutôt qu'à la qualité et au processus de l'évaluation. Ils ont particulièrement souligné le caractère désordonné de l'activité évaluative et le fait que des contraintes sociopolitiques les empêchaient de suivre le devis d'évaluation à la lettre.

(2) *En ce qui concerne les sept constats sur l'implication des parties prenantes et leur importance aux yeux de l'évaluateur (n=954),* 74% des répondants ont reconnu que l'implication des parties prenantes est beaucoup plus importante pour définir les bornes de l'évaluation, alors que seulement 48% ont jugé qu'il était important d'impliquer les parties prenantes dans toutes les étapes du processus évaluatif. La dernière donnée nous parait étrange puisque pour 78% d'entre eux, impliquer de multiples parties prenantes augmente l'utilisation des résultats, et que les travaux de recherche dans le domaine montrent que consulter les parties prenantes seulement au moment de la définition du mandat et des questions d'évaluation ne constituait pas une garantie suffisante pour l'utilisation. Selon nous, ceci explique pourquoi selon 86% des répondants certaines parties prenantes clés rejettent les conclusions de l'évaluation sur la base de leurs croyances et valeurs personnelles plutôt que sur l'obtention de nouvelles informations. En effet, s'il n'est pas nécessaire que les parties prenantes participent à toutes les étapes, il est cependant primordial qu'elles soient tenues informées tout au long du processus évaluatif et qu'elles aient la possibilité de donner leur rétroaction au moment du développement de la méthodologie, etc. La qualité des données collectées et la rigueur méthodologique en dépendent. D'ailleurs, les travaux de Bamberger et Mabry (2020) ont montré que les contraintes politiques représentent les défis les plus importants dans la mise en œuvre d'évaluation rigoureuse du fait notamment des exigences méthodologiques des parties prenantes clés et de la rigidité des organismes gouvernementaux concernant les approches méthodologiques standards. Aussi, la nécessité de faciliter un processus de négociation

des décisions d'ordre méthodologique entre les exigences des parties prenantes clés, la gestion des contraintes de temps, d'argent et d'accès aux données ainsi que la protection de la rigueur scientifique de l'évaluation, il est primordial de communiquer régulièrement avec les parties prenantes clés, et ce, tout au long du processus évaluatif.

(3) *D'ailleurs, parmi les dix rôles de l'évaluateur identifiés comme importants ou très importants (n=1047),* impliquer les parties prenantes dans le processus évaluatif (98%), faciliter l'apprentissage organisationnel (89%) et maximiser l'utilisation (86%) représentent les activités les plus reconnues par les répondants. Aussi, les praticiens reconnaissent le lien entre l'évaluation et le développement des capacités évaluatives des organisations. Cependant, compte tenu des résultats précédents, nous émettons l'hypothèse que les évaluateurs interrogés ne détenaient peut-être pas les connaissances et (ou) les habiletés suffisantes associées au développement de stratégies qui facilitent l'implication efficace des parties prenantes clés (ni trop ni pas assez), afin de favoriser l'utilisation, et ce, sans négliger la rigueur méthodologique. En effet, 68% des évaluateurs ont reconnu la non-utilisation des résultats évaluatifs comme un problème majeur, mais les problèmes de mésusage intentionnel et non intentionnel sont pour eux assez minimes. Or, pour les répondants, le rejet des conclusions de l'évaluation par des parties prenantes clés sur la base de convictions ou d'intérêts personnels est l'élément le plus important en ce qui concerne l'implication de ces dernières. Ce qui semble coïncider avec le problème important de non-utilisation. Cependant, ce type de non-utilisation sans la remise en question de la qualité de l'évaluation est une forme de mésusage intentionnelle. De même, pour reprendre les propos de Patton (2008), avec l'augmentation de l'utilisation augmente le mésusage, et le mésusage n'est pas nécessairement flagrant, il est souvent le résultat d'une accumulation de mauvaises influences tout au long de l'évaluation. En outre, il peut se manifester de manière anodine, sans que l'évaluateur le perçoive comme tel, ou encore l'évaluateur n'est tout simplement pas témoin des manipulations en début, pendant ou après l'évaluation. Aussi, le fait que les risques de mésusage ne semblent pas préoccuper les évaluateurs est pour nous un élément inquiétant.

(5) *Pour terminer, si le risque de mésévaluation ne constituait pas un élément de l'étude de Fleischer et Christie (2009),* il s'avère un enjeu important pour les pratiques méthodologiques des évaluateurs de programme, particulièrement pour les évaluateurs novices. En effet, l'étude de Chouinard et *coll*. (2017) a fait émerger une distinction essentielle entre l'évaluation en tant qu'entreprise d'investigation très technique et l'évaluation en tant qu'activité socioculturelle et

politique. Les résultats ont montré que si la formation des jeunes évaluateurs tente de les équiper pour mieux réaliser des évaluations rigoureuses (connaissances et compétences méthodologiques), ces derniers n'ont pas suffisamment de connaissance et d'expérience pour s'adapter et gérer les dimensions sociale et politique de l'activité d'évaluation. Et, le fait que la volonté de collaborer des parties prenantes clés repose sur leurs perceptions des habiletés sociales et de la crédibilité de l'évaluateur renforce ce manque de formation des jeunes évaluateurs. En effet, sur le terrain, les étudiants-évaluateurs reconnaissent avoir eu du mal à établir et à maintenir ce qu'ils appellent « les exigences idéales pour être un évaluateur efficace ». Ils expliquent qu'ils n'ont pas été capables de faire suffisamment preuve d'autorité et de crédibilités, et d'accéder aux parties prenantes clés (Chouinard et *coll.*, 2017). Aussi, un tel constat les a conduits à identifier la gestion des relations au sein et à l'extérieure de l'équipe évaluative comme l'un des rôles les plus importants de l'évaluateur, car la manière dont les parties prenantes clés le perçoit détermine leurs degrés d'intérêt et d'engagement envers l'évaluation. Certains évaluateurs-étudiants ont d'ailleurs expliqué que les préjugés et les préoccupations des parties prenantes clés envers leur statut de « novice » ont engendré un désintérêt des clients, l'annulation de rendez-vous de travail, l'absence de communication, la difficulté de planifier la collecte de données, etc., ce qui entache la qualité de l'activité évaluative et augmente les risques de mésévaluation. Par ailleurs, si les jeunes évaluateurs doivent constamment céder la place aux évaluateurs plus expérimentés, cela freine leurs apprentissages expérientiels, leur adaptation à la complexité du terrain et leur développement professionnel. Le fait que la formation en évaluation priorise l'aspect technique de l'activité évaluative, et que l'expérience de terrain est censée favoriser les apprentissages relatifs à l'adaptation de l'activité évaluative en réponse aux facteurs socioculturels et politiques et autres contraintes ESR, pose un problème majeur en termes de pratique méthodologique. Les questions techniques relatives aux choix méthodologiques (le recours à un devis expérimental ou non, le recours à un devis méthodologique qualitatif, quantitatif ou mixte, la sélection des sources d'information, la sélection des questions d'entrevues, etc.) peuvent provoquer de vifs débats, encore faut-il que les jeunes évaluateurs soient en mesure de les gérer de manière efficace et de délivrer une évaluation rigoureuse.

2.2.4 Synthèse conceptuelle

L'évolution de l'évaluation de programme a amené de nombreux changements nécessaires, tant sur le plan méthodologique que sur le plan humain, faisant de l'évaluation contemporaine une activité à la fois technique et sociopolitique. Aujourd'hui, l'évaluateur doit être en mesure de produire une évaluation rigoureuse en s'assurant à la fois d'impliquer les parties prenantes clés tout au long de l'activité évaluative, d'adapter ses choix méthodologiques en fonction des contraintes politico-organisationnelles, budgétaires, temporelles, et d'accès aux données ainsi qu'en veillant à prévenir les risques de mésusage et de mésévaluation. En effet, la validité, la pertinence et la fiabilité d'une évaluation reposent sur la capacité de l'évaluateur à adapter efficacement ses choix méthodologiques face aux différentes pressions exercées par l'environnement socio-économique, politique et culturel dans lequel l'évaluation se déroule. La manifestation de différents facteurs tels que la résistance des parties prenantes clés, le mésusage, la limite des intrants disponibles pour l'évaluation, les normes sociopolitiques et culturelles de l'environnement évaluatif, etc., peuvent affecter : 1) la pertinence et la faisabilité du devis d'évaluation; 2) l'accès à des sources d'information et des données pertinentes et fiables; 3) la validité des inférences faites à partir des données collectées (validité des conclusions); et 4) la pertinence et la fiabilité de l'équipe d'évaluation en termes d'aptitudes méthodologique, administrative, économique et sociale, et éthique.

Enfin, les enjeux de formation et de développement professionnels face à la dimension sociopolitique de l'activité évaluative et la gestion des contraintes inhérentes au terrain nous conduisent à nous questionner à présent sur les caractéristiques de la pratique des évaluateurs selon leur niveau d'expertise, particulièrement en ce qui a trait à la variable « méthodologie » et l'adaptation de la pratique des évaluateurs dans un environnement évaluatif complexe et contraignant (Figure 11, p. 76).

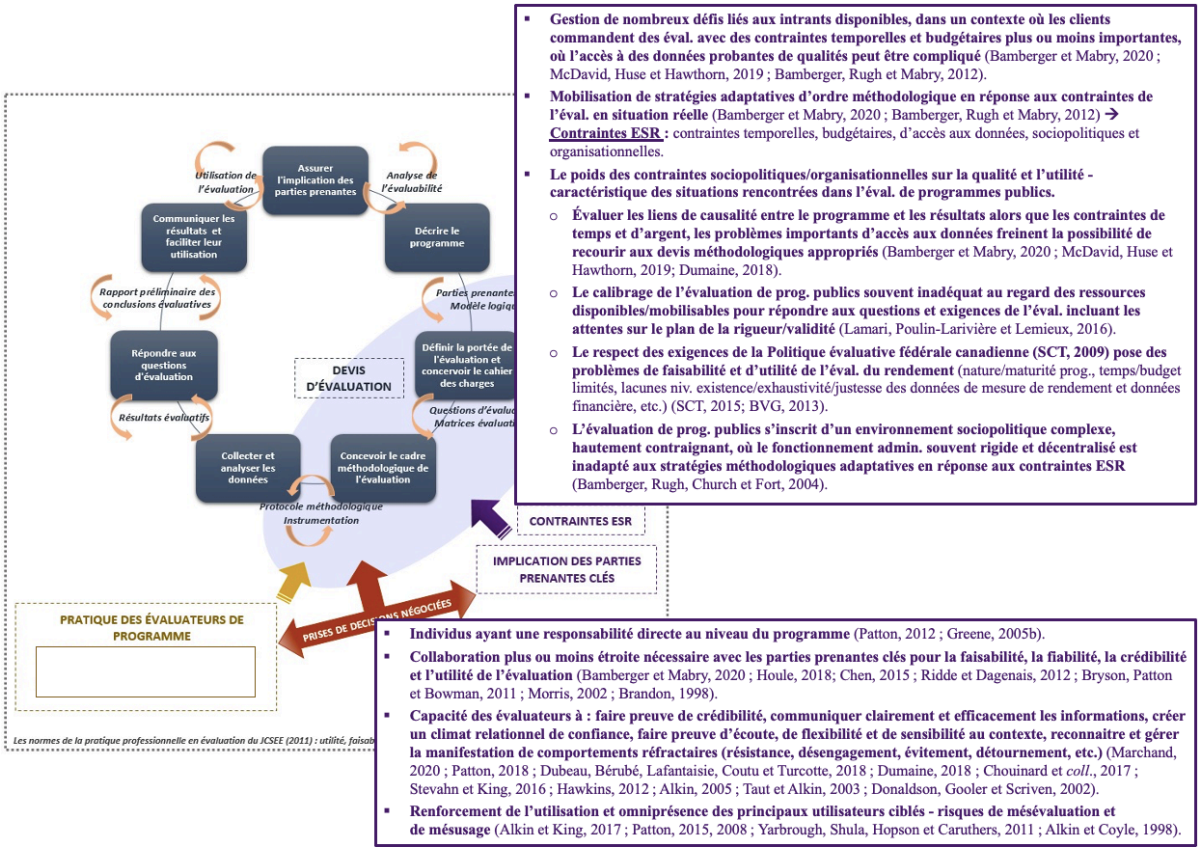


Figure 11. *Synthèse conceptuelle [2/3]*

2.3 Les caractéristiques de la pratique des évaluateurs de programme en fonction de leur niveau d'expertise

Dans cette section, il s'agit d'identifier les caractéristiques de la pratique des évaluateurs novices et de la pratique des évaluateurs experts. Aussi nous traitons, entre autres, des questions suivantes : qu'est-ce qu'un expert ? En quoi la pratique des évaluateurs novices se distingue-t-elle de celle des évaluateurs experts ?

2.3.1 L'expertise et le développement professionnel

2.3.1.1 Qu'est-ce que l'expertise et comment se développe-t-elle?

Nous retrouvons deux courants au sein desquels évoluent, ou ont évolué les théories et les études sur l'expertise (Daley, 1999). Le premier courant d'étude se concentrait sur l'élaboration de modèles, de concepts et de théories sur l'expertise et la résolution de problème, applicable sur une kyrielle de domaines professionnels (Simon et Newell, 1971). Le second courant d'étude, au

contraire, visait le développement d'études portant sur des domaines spécifiques et les processus de développement professionnels qui leur sont propres. On trouve, par exemple, les travaux fondateurs de Simon et Chase (1973) portant sur la pratique des joueurs d'échecs professionnels et la mémoire d'experts. Ainsi, l'analyse de la performance des experts a permis de mettre de l'avant que le stockage en mémoire de toutes les connaissances nécessaires à la pratique ne suffit pas, et qu'il est crucial que les informations pertinentes soient bien organisées et accessibles quand elles sont nécessaires pour la réalisation de l'activité en cours (Ericsson, Patel et Kintsch, 2000). En effet, un stockage en mémoire, efficace et fiable, des informations pertinentes est particulièrement important pour les experts, quand ils planifient et mettent en œuvre des raisonnements complexes qui caractérisent leur performance de niveau supérieur. Aussi, durant la planification, les experts doivent mentalement comparer de nombreuses possibilités de séquences d'action ce qui maximise le taux d'informations disponible dans la mémoire de travail (Simon et Chase, 1973). C'est pourquoi les débutants aux échecs ne génèrent pas de longs plans d'action, et cela prend de nombreuses années d'étude et de pratique des échecs avant qu'un joueur expert soit capable de mobiliser de longs et fiables plans. De surcroit, les maitres aux échecs, quant à eux, ont développé une mémoire telle, qu'ils sont capables de jouer aux échecs les yeux bandés, car ils ont une représentation de toutes les pièces sur l'échiquier en mémoire, et ce, durant la partie entière.

Dans le second courant d'étude on retrouve également les recherches de Dreyfus et Dreyfus (1980) qui portent sur la pratique des pilotes d'avion, et à la suite desquelles les chercheurs ont proposé un modèle du développement de l'expertise, que Benner et Tanner (1987, 1982) appliqueront dans l'analyse de la pratique des infirmières. Ce modèle démontre que les professionnels évoluent sur un continuum développemental (au fil des expériences de terrain), leur permettant de progresser du statut de novice à celui d'expert, et ce, en cinq stades (Tableau 4).

Tableau 4. *Les cinq stades du développement professionnel (Dreyfus et Dreyfus, 1980)*

Novice (stade 1)
- Ils ont très peu ou pas d'expérience en situation réelle. - Ils doivent s'appuyer sur les règles apprises durant leur formation professionnelle pour déterminer leurs actions.
Débutant avancé (stade 2)
- Ils réalisent des performances acceptables. - Ils ont été confrontés à suffisamment de situations réelles pour être en mesure de classifier des composantes contextuelles et situationnelles significatives et récurrentes (Benner, 1982). - Ils commencent à différencier les situations entre elles, mais ils ont toujours des difficultés à faire la distinction entre ce qui est important et ce qui ne l'est pas.

Professionnel qualifié (stade 3)
- Ils pratiquent en moyenne depuis 3 à 5 ans. - Ils peuvent planifier et organiser des activités. - Ils maîtrisent chaque étape de leur plan d'action et ils estiment avoir l'habileté à gérer et à s'adapter à des situations imprévisibles et potentiellement contraignantes.
Professionnel très compétent (stade 4)
- Ils commencent à développer une approche holistique de la pratique de leur profession. - Ils ont été exposés à une grande variété d'ensembles de situations typiques. - Ils sont à présent en mesure de déterminer l'importance (ou non) des différents éléments situationnels ou contextuels, et les faits saillants, pertinents, en ce qui a trait à l'atteinte des objectifs à long terme. - L'activité cognitive (traitement de l'information) en corrélation avec l'expérience du praticien (point de vue particulier) face à une situation dans son ensemble, est organisée et stockée afin de fournir une base de données pour reconnaitre de futures situations similaires perçues d'un point de vue similaire.
Expert (stade 5)
- Ils ont développé un répertoire de situations vécues suffisamment vaste pour que chaque situation spécifique détermine une action immédiate, intuitive, et appropriée. - Ils peuvent avoir recours à leurs intuitions, chaque type de situation est à présent associé à une réponse spécifique (processus développé au stade 4). - Ils ont développé un sens de priorité en ce qui a trait aux différents types d'activités relatives à leur plan d'action, d'intervention. Aussi celles-ci n'ont pas le même ordre d'importance. - Ils ont développé une rationalité délibérée ou un réseau approfondi de perspectives, qui les amènent à voir une situation en termes de situations passées. L'expert a appris à s'attendre à certains événements, voir même à certains aspects de la situation, et ce, de manière sélective (Benner et Tanner, 1987).

Enfin, l'évolution des professionnels sur le continuum novices-experts s'effectuerait sous forme de trois « basculements » au niveau de la performance des praticiens (Daley, 1999). Le premier se réalise au niveau du paradigme de travail du professionnel (soit de la dépendance aux principes abstraits, à la mobilisation d'expériences concrètes passées). Le deuxième au niveau des perceptions du professionnel (soit de la perception des situations comme discrètes et indépendantes, à la compréhension de celles-ci comme faisant partie d'un tout). Enfin, le troisième se réalise au niveau de la posture du professionnel (soit d'un observateur détaché, à un acteur engagé). De surcroit, il nous parait important de revenir sur le terme expérience mobilisé dans cette modélisation. En effet, l'expérience est comprise et utilisée dans l'acquisition de savoir-faire, et a une définition particulière qui doit être clarifiée. Aussi, comme il est décrit dans le modèle Dreyfus et Dreyfus (1980), l'expérience n'est pas le simple passage du temps ou la longévité ; elle est le raffinement des idées préconçues et de la théorie, à la suite de la rencontre de nombreuses situations pratiques réelles, et ce, dans le cadre du développement professionnel (Benner, 1982).

2.3.1.2 Le développement professionnel (savoir agir avec compétence)

Selon Le Boterf (2002), « une personne compétente est une personne qui sait agir avec pertinence dans un contexte particulier, en choisissant et en mobilisant un double équipement de ressources : ressources personnelles (connaissances théorique et pratique, savoir-faire, qualités, culture, ressources émotionnelles...) et ressources de réseaux (banques de données, réseaux documentaires, réseaux d'expertise, etc.). » (p. 46). Ainsi, savoir agir avec compétence signifierait être capable de (1) construire et d'activer des combinatoires pertinentes de ressources incorporées ou internes (personnelles) et de ressources externes ou de réseaux (2) pour gérer une gamme de « situation(s) professionnelle(s) » étant décrite avec une activité ou une mission à laquelle sont associés des « critères de réalisation » contextualisés (3) en vue d'atteindre des « résultats » (produits, services) auxquels sont associés des « échelles de réussites » et adressés à un « destinataire » (clients, usagers, patients...), afin de (4) répondre à des « enjeux ». Selon l'auteur, il ne s'agirait donc pas de posséder des compétences, et des ressources pour faire preuve de compétence dans le cadre de sa pratique professionnelle. En effet, il apparait nécessaire d'être en mesure de les « organiser et de les mobiliser en combinatoires pertinentes » et ainsi être capable de gérer des situations professionnelles en tenant compte des critères de réalisation et de contextualisation qui leur sont propres. Enfin, l'équipement incorporé dans lequel le praticien peut puiser pour agir avec compétence comporte divers types de ressources (Le Boterf, 2002), que nous avons regroupées dans le Tableau 5 suivant et qui faciliterait selon nous le cheminement au travers des différents niveaux de professionnalisme ainsi que le développement de l'expertise en cinq stades de Dreyfus et Dreyfus (1980), que nous avions présentés précédemment.

Tableau 5. *L'équipement incorporé dans lequel le praticien peut puiser pour agir avec compétences (Le Boterf, 2002)*

Savoirs théoriques et procéduraux
- Des connaissances générales : concepts, savoirs disciplinaires, etc. permettant de comprendre un phénomène, une situation. - Des connaissances spécifiques à l'environnement professionnel : connaissances sur le contexte de travail telles que l'équipement, les règles de gestion, la culture organisationnelle, l'organisation de l'entreprise ou de l'unité, etc. Ainsi permettent-elles d'agir sur mesure. - Des connaissances procédurales : elles permettent de décrire « comment il faut faire », « comment s'y prendre pour ». Ainsi décrivent-elles des procédures, des méthodes, des modes opératoires.
Savoirs pratiques et expérientiels

- Des savoirs opérationnels : il s'agit des démarches, des méthodes, des instruments dont la personne maîtrise l'application pratique. Ainsi permettent-ils de savoir opérer.
- Des connaissances et des savoirs expérientiels : ils sont issus de l'expérience, de l'action. Les savoir-faire sont difficilement exprimables, on parle alors de « connaissances tacites », de « façon de faire », d'« astuces », etc.
- Des savoir-faire relationnels : soit des capacités qui permettent de coopérer facilement avec autrui (capacité d'écoute, de négociation, de travail en équipe, de travail en réseau).
- Des savoir-faire cognitifs : il s'agit des opérations intellectuelles nécessaires à l'analyse et à la résolution de problèmes, à la conceptualisation et à la réalisation de projets, à la prise de décision, à l'invention (induction, déduction, abstraction réfléchissante, raisonnement par analogie, production d'hypothèse et généralisation). Ainsi s'agit-il de créer des informations nouvelles à partir d'informations existantes.

Enfin, nous avons vu un peu plus tôt que les résultats de l'étude de Taut et Alkin (2003) sur la perception des parties prenantes en ce qui a trait à la pratique des évaluateurs de programmes, identifient la compétence de l'évaluateur comme un facteur d'influence important. Si les parties prenantes évoquent la compétence globale de l'évaluateur, trois volets spécifiques de dessinent dans leur discours, soit : i. *ses compétences sociales*, incluant notamment ses habiletés interpersonnelles; ii. *ses connaissances sur le programme évalué* et ses caractéristiques contextuelles; et iii. *ses compétences techniques/méthodologiques*, c'est-à-dire sa capacité à concevoir, mettre en œuvre et gérer l'évaluation. Aussi, on retrouve lesdites compétences dans la *Taxonomie des compétences essentielles* développée par Stevahn, King, Ghere et Minnema (publié en 2001, et révisé en 2005), qui fait référence aux connaissances, aptitudes, et attitudes requises pour que l'évaluateur réalise des évaluations solides, qui reflètent les normes, les principes et lignes directrices approuvées par les principales associations d'évaluation nord-américaine. Ce référentiel est composé de 61 compétences organisées en six domaines ou volets de compétences (Annexe 4) :

1. *Pratique professionnelle* – volet de compétences qui se concentrent sur les normes et les valeurs fondamentales qui sous-tendent la pratique de l'évaluation, comme le respect des normes évaluatives et de l'éthique.
2. *Enquête systématique* – volet de compétences qui se concentrent sur les aspects plus techniques de la pratique de l'évaluation, tels que la conception, la collecte de données, l'analyse, l'interprétation et la production de rapports.
3. *Analyse situationnelle* – volet de compétences qui se concentrent sur l'analyse et la réponse aux intérêts, problèmes et circonstances contextuelles uniques d'évaluation donnée.
4. *Gestion de projet* – volet de compétences qui se concentrent sur les rouages de la conduite d'une évaluation, comme la budgétisation, la coordination des ressources et la supervision des procédures.

5. *Relations interpersonnelles* – volet qui se concentrent sur se concentrent sur les compétences interpersonnelles utilisées dans la conduite d'études d'évaluation, telles que la communication, la négociation, la gestion/résolution de conflit, la collaboration et les compétences interculturelles.
6. *Pratique réflexive* – volet de compétences qui se concentrent sur la prise de conscience de l'expertise en évaluation et des besoins en termes de croissance professionnelle (y compris la connaissance de soi en qu'évaluateur, l'identification de ses propres besoins pour améliorer sa pratique et l'engagement dans le développement professionnel vers cet objectif).

En outre, cette à partir de la Taxonomie de Stevahn, King, Ghere et Minnema (2005), que la Société canadienne d'évaluation (SCÉ) a entrepris, dans le cadre de son Programme de désignation professionnelle (processus optionnel d'accréditation en sol canadien), la conception de son propre référentiel de compétences requises à l'exercice de l'évaluation. La démarche a consisté en la conduite de diverses recherches, consultations des membres de la SCÉ et procédures de validation auprès d'experts canadiens en 2008 et 2009. Puis, en 2017-2018, le Groupe de travail sur l'examen des compétences a réalisé une mise à jour du référentiel afin de répondre à l'évolution du cadre de la profession, et il a simplifié et précisé les énoncés relatifs aux compétences. Ce référentiel répartit, aujourd'hui, 36 compétences au sein de cinq volets de compétences que vous nous présentons dans le Tableau 6 (p. 82).

Tableau 6. *Référentiel des compétences requises à l'exercice de l'évaluation de programme au Canada (SCÉ, 2018)*

Volet 1. Pratique réflexive – Ces compétences réfèrent aux connaissances portant sur la théorie et la pratique de l'évaluation ; sur le recours aux normes et lignes de conduite en éthique. Elles incluent aussi à la conscience de soi, ce qui réfère à une réflexion sur sa pratique professionnelle et la nécessité d'une formation continue et d'un développement professionnel.
1.1 Connaître les théories, modèles, méthodes et les outils associés à l'évaluation et demeurer au fait des nouveaux courants de pensée et des meilleures pratiques.
1.2 Intégrer les normes de pratique professionnelle en évaluation de programme qui émanent du comité mixte Canada/É.U.
1.3 Intégrer au sein de la pratique professionnelle les lignes directrices en éthique élaborées par la Société canadienne d'évaluation et maintenir une préoccupation éthique tout au long de la démarche évaluative.
1.4 Prendre en considération le bien-être des personnes et des systèmes naturels dans le cadre de la pratique évaluative.
1.5 Fournir une perspective indépendante et équilibrée tout au long de l'évaluation.
1.6 S'engager à la transparence tout au long de l'évaluation.
1.7 Avoir recours à la conscience de soi et à la pensée réflexive pour améliorer sa pratique sur une base continue.
1.8 Participer à des activités de réseautage professionnel et avoir un apport à la profession de l'évaluation ainsi qu'à sa communauté de pratique.
Volet 2. Pratique technique – Ces compétences mettent l'accent sur les décisions stratégiques, méthodologiques et interprétatives nécessaires pour mener une évaluation.
2.1 Clarifier le but et l'étendue de l'évaluation.
2.2 Évaluer la faisabilité de l'évaluation.
2.3 Clarifier la théorie de programme.
2.4 Cerner les thèmes et les questions qui seront abordés dans le cadre de l'évaluation.
2.5 Planifier une évaluation.
2.6 Avoir recours à des méthodes d'évaluation appropriées
2.7 Identifier les données, les sources, l'échantillonnage et les outils de collecte de données requis.
2.8 Recueillir, analyser et interpréter les données en utilisant des méthodes appropriées.
2.9 Utiliser les résultats pour répondre aux questions à l'origine de la démarche évaluative et, le cas échéant, pour formuler des recommandations.
2.10 Générer des rapports d'évaluation complets et équilibrés dans le but de faciliter la prise de décision et l'apprentissage.
Volet 3. Pratique contextuelle – Ces compétences visent avant toute chose à comprendre, analyser et inclure les nombreuses circonstances qui rendent chaque évaluation unique, compte tenu de la culture, des parties prenantes et du contexte.
3.1 Examiner et prendre en considération les multiples contextes humains et naturels dans lesquels le programme s'inscrit.
3.2 Identifier les besoins des parties prenantes ainsi que leur capacité à participer, tout en reconnaissant, respectant, et étant sensible à la diversité.
3.3 Respecter toutes les parties prenantes et s'efforcer de créer et de maintenir une relation de confiance.
3.4 Encourager et favoriser l'utilisation du processus de l'évaluation et des résultats qui en découlent.
3.5 Identifier, s'adapter aux changements observés au sein du contexte du programme et envisager les éventuels effets positifs et négatifs de l'évaluation.
3.6 Assurer la mise en place de processus de réciprocité entre l'évaluateur et les parties prenantes afin de favoriser le partage des connaissances et de l'expertise et améliorer leur compétence en évaluation.
3.7 Avoir recours aux processus et aux pratiques évaluatifs favorisant la réconciliation et le renforcement les relations entre les peuples autochtones et non autochtones.
Volet 4. Pratique de gestion – Ces compétences mettent l'accent sur l'application de solides aptitudes en gestion de projet tout au long du projet d'évaluation.
4.1 Agir comme leader dans le cadre d'un projet d'évaluation.
4.2 Définir les paramètres de réalisation, les plans et les ententes concernant l'évaluation.
4.3 Identifier et utiliser efficacement les ressources humaines, financières et techniques requises.
4.4 Coordonner le travail des autres membres de l'équipe.
4.5 Avoir recours à des pratiques de gestion et de communication de groupe.
4.6 Faire part de l'avancement du projet à toutes les personnes concernées.
Volet 5. Pratique concernant les relations interpersonnelles – Ces compétences mettent l'accent sur les habiletés sociales et personnelles qui sont requises pour communiquer et interagir efficacement avec toutes les parties prenantes.
5.1 Utiliser des stratégies de communication adaptées au contexte culturel, linguistique, social et politique.
5.2 Faire preuve d'une compétence adéquate et efficace en matière de communication écrite et visuelle.
5.3 Faire preuve d'habiletés efficaces, appropriées et respectueuses en matière de communication verbale et non verbale.
5.4 Utiliser une variété de mécanismes favorisant des ententes mutuellement négociées, une compréhension commune et la concertation.
5.5 Bâtir des partenariats dans le contexte de l'évaluation.

2.3.2 La pratique évaluative ou les choix des évaluateurs dans le cadre de leur pratique selon qu'ils soient novices ou experts

Certes, il y a les connaissances relatives aux théories en évaluation, cependant les choix des évaluateurs en ce qui concerne la pratique évaluative, dépendent de ce qui est le plus approprié de mobiliser en situation réelle. Ainsi, l'excellence en évaluation serait basée sur les choix et les méthodes appropriées au contexte (Stone 1984, cité par Nadeau, 1988). De nombreux auteurs soulignent que les choix des évaluateurs dépendent en grande partie des facteurs contextuels, ce qui impliquerait l'existence de connaissances pratiques par rapport à des connaissances dites formelles relatives aux théories en évaluation. Ces connaissances pratiques les aideraient à interpréter la situation rencontrée afin d'effectuer les choix les plus appropriés, en fonction des objectifs de l'évaluation, des attentes des parties prenantes, des possibilités d'accès aux données, et des contraintes et des caractéristiques du milieu (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019 ; Fitzpatrick, Christie et Mark, 2008; Greene, Lipsey, Schwandt, Smith et Tharp, 2007; Greene, 2006; King, 2003; Worthen, Sanders et Fitzpatrick, 1997).

Les choix effectués par les évaluateurs seraient alors relatifs à leurs propres théories implicites, aucune méthode ne semblant être appropriée en tout temps et en toutes situations (Shadish, Cook et Leviton, 1991). Théoriciens naissants, les évaluateurs recourent à des jugements considérés appropriés vis-à-vis des méthodes à mobiliser, et ce, en fonction des caractéristiques de chaque situation évaluative rencontrée, tout en envisageant les avantages et inconvénients des choix auxquels ils font face (McPherson, 2015). Ainsi apprennent-ils de leurs succès, mais aussi des échecs qu'ils ont rencontrés au cours d'évaluations passées (Tourmen, 2009).

2.3.2.1 Les caractéristiques générales de la pratique des professionnels de l'évaluation de programme

Selon Kundin (2010), lorsqu'on s'intéresse à la question des processus décisionnels engagés par les évaluateurs de programme sur le terrain, on trouve des études sur l'alignement entre le travail des évaluateurs et les théories ou théories de l'évaluation, ou les travaux qui suggèrent que les praticiens de l'évaluation ont recours à leurs intuitions, qu'ils mobilisent leur expertise, leurs valeurs et leur jugement afin d'être en mesure de répondre le plus efficacement à une situation. Aussi, si pour l'auteur les écrits scientifiques en évaluation augmentent, ceux relatifs à des études portant spécifiquement sur la pratique des évaluateurs sont peu nombreux et

sporadiques. Ainsi, parmi les rares recherches qui abordent concrètement le processus décisionnel mobilisé par les évaluateurs dans leur pratique de tous les jours, les ajustements qu'ils réalisent afin de faire face aux contraintes de terrains, etc., nous retrouvons les thèses doctorales de Tourmen (2007) et McPherson (2015).

Dans le cadre de sa recherche doctorale sur les compétences des évaluateurs de politiques publiques, Tourmen (2007) s'est concentrée sur les activités de phase de conception d'une démarche d'évaluation. L'échantillon était composé de neuf professionnels français (cinq expérimentés et quatre débutants). L'analyse des activités de conception de l'évaluation a été réalisée en croisant les résultats de trois méthodes de collecte de données :

- Des entretiens de retour sur activité auprès de 3 professionnels;
- L'observation de 2 séquences réelles mêlant des évaluateurs expérimentés et des débutants (construction d'un diagramme des impacts attendus et construction d'un référentiel d'évaluation);
- Une étude de cas de rédaction d'un cahier des charges d'évaluation (conception d'un modèle logique et d'un plan d'évaluation), avec des entretiens d'explicitation (démarche entreprise, choix méthodologiques, difficultés rencontrées, etc.), qu'elle a soumis à quatre praticiens de l'évaluation, soit deux évaluateurs expérimentés (responsable d'évaluation pour le ministère et le conseil régional et au moins cinq ans d'expérience) et deux évaluateurs débutants (l'un ayant une formation en évaluation et un an d'expérience en stage, et l'autre, ayant suivant quatre mois de formation en évaluation, sans expérience).

Parmi les contributions des résultats de l'étude de Tourmen (2007), on trouve une description de l'activité des débutants et son organisation, et de l'activité des professionnels expérimentés et son organisation (Tableau 7).

Tableau 7. *Caractéristiques générales des professionnels de l'évaluation (débutants et expérimentés) (Tourmen, 2007)*

	Caractéristiques générales des professionnels
Évaluateurs débutants	- Diagnostic limité de la situation, et faible analyse des indices de situation volontairement disséminés dans le cas lors de sa construction. - Interprétation limitée de la situation pour agir. - Réalisation des étapes du plan d'action selon les règles prescrites pour s'assurer du respect des buts et des façons de faire. - Conscientisation de l'importance des critères de faisabilité qu'ils n'arrivent cependant pas à bien intégrer dans leurs choix méthodologiques.

Évaluateurs expérimentés	- Diagnostic approfondi de la situation, et mise en évidence du manque d'informations relatives au cas (manque de données et particulièrement d'indices informels sur les attentes des acteurs du programme cas). - Questionnement des potentiels conflits et de la participation possible des acteurs tout au long l'exercice. - Interprétation de la situation et réalisation d'un pronostic de son évaluation pour agir. - Assertions et raisonnements d'anticipation afin de prédire les effets de la politique, les attitudes et la réaction des acteurs face aux conclusions possibles de l'évaluation. - Établissement de lien entre les différentes étapes et réajustement (si besoin est) afin d'assurer la bonne réalisation de l'évaluation, tout en prenant en compte les dimensions liées à la faisabilité.

Nous sommes conscients de la limite que constitue le nombre restreint de participants à l'étude de cas, et le fait que les résultats rendaient peut-être davantage compte de ce que les évaluateurs pensent qu'ils « devraient » faire plutôt que ce qu'ils « font réellement » en réponse à des situations réelles. Néanmoins, l'étude de Tourmen (2007) représente, selon nous, une réelle contribution sur la question de la démonstration des processus décisionnels engagés dans la pratique des évaluateurs de programme situations. En effet, les résultats montrent :

- *Les évaluateurs expérimentés* ont tendance à être centrés sur les résultats et l'adaptation à chaque situation particulière. Ils sont attentifs à l'environnement, au contexte, aux contraintes. Les choix méthodologiques opérés s'inscrivent dans un diagnostic approfondi d'une situation évaluation donnée ainsi qu'un pronostic de son évolution.
- *Les évaluateurs débutants* sont davantage centrés sur la méthode et les étapes prescrites. Ils sont collés aux concepts étape par étape dans leur démarche d'évaluation. Leur pratique s'inscrit dans une logique d'imitation, et ils sont déroutés par les variations d'une situation.

Nous élaborons davantage les mécanismes décisionnels entrepris par les évaluateurs expérimentés, dans la section suivante qui traite de l'apport de l'expérience pour apprendre à évaluer un programme efficacement.

L'étude doctorale de McPherson (2015) visait quant à elle à explorer et décrire la façon dont les évaluateurs experts utilisent, et s'appuient sur des théories implicites dans leur cadre de leur pratique évaluative. Pour ce faire, le chercheur a administré deux tâches dans le cadre d'un entretien mené auprès de onze évaluateurs exerçant dans des cabinets de conseil et en milieu universitaire. La première tâche portait sur la réalisation d'une activité où avec une photo en main et un scénario dans un contexte spécifique, le participant devait faire part (à l'écrit, et à l'orale) de trois impressions face à la photo et sept éléments sur lesquels ils aimeraient obtenir des

informations pour mener son évaluation (passation répétée avec une nouvelle association photo-scénario). La deuxième tâche était une entrevue semi-dirigée articulée autour des expériences évaluatives des participants et les cinq composantes de la Théorie de l'évaluation de Shadish, Cook et Levinton (1991), soit la programmation sociale, la construction des connaissances, la valorisation, l'utilisation, et la pratique de l'évaluation. De manière générale, les résultats montrent que si les caractéristiques situationnelles exercent une influence certaine sur l'approche évaluative, ce sont les construits théoriques propres à l'évaluateur conçu à l'appui des savoirs pratiques en évaluation de programme, d'expériences antérieures, et du milieu de pratique qui vont ultimement façonner les réponses de l'évaluateur aux perturbations contextuelles. En effet, les données suggèrent que les évaluateurs ne s'appuient pas entièrement sur une seule théorie établie (publiée) en évaluation de programme, mais tiennent plutôt des hypothèses théoriques implicites globales qui guident leur adaptation et leur mise en œuvre de leurs propres théories d'évaluation à des opportunités d'évaluation uniques. Enfin, précisons qu'en dépit du fait que la taille de l'échantillon demeure restreinte, et que l'étude s'est limitée à des descriptions narratives de comportements qui n'ont peut-être pas pleinement révélé, dans les situations évaluatives rapportées, où l'activation des schèmes et le recours aux théories implicites se sont exactement produits (McPherson, 2015), la tentative de modélisation du processus cognitif engagé dans la formulation des théories implicites et du jugement évaluatif est une contribution très intéressante (Figure 12, p. 88). D'ailleurs, selon Kundin (2010), « *[t]he empirical knowledge gained from studying how evaluators use their ''know-how'' is essential for understanding the nature of everyday practice* » (p. 358).

2.3.2.2 Les apports de l'expérience pour apprendre à évaluer efficacement

L'analyse de l'activité des professionnels expérimentés a permis à Tourmen (2009, 2008, 2007) de montrer que ces évaluateurs ont développé un ensemble de ressources cognitives qui leur permet de résoudre les difficultés auxquelles ils sont confrontés, et de développer différents types de stratégies. Ainsi l'étude de Tourmen (2007) montre que les professionnels expérimentés ont développé des *stratégies de diagnostic*, soit la mobilisation de « réseau d'indices auxquels ils sont attentifs et qu'ils cherchent à diagnostiquer, un réseau de connaissances et de croyances [...], afin de qualifier l'état d'une situation et de prédire son évolution » (p. 306) ; des *stratégies d'action*, soit la mobilisation de « réseaux de règles d'action qui permettent d'adapter leurs choix aux diagnostics effectués » (p. 306) ; et des *stratégies d'arbitrage et de compromis* entre les options

possibles, soit « les représentations des buts redéfinies en situation, et des concepts-en-acte qui leur permettent de trancher entre différentes options et d'être souples dans les choix des méthodes. » (p. 306).

Compte tenu de ces résultats, Tourmen (2009, 2007) émet l'hypothèse selon laquelle les ressources développées et mobilisées par les évaluateurs expérimentés se sont développées dans et par l'expérience par des processus de conceptualisation en action, et des processus de pragmatisation des ressources transmises. Il s'agit de systèmes de représentation en actes composés de « propositions tenues pour vraies sur le réel », et de raisonnements conditionnels de type « si..., alors... ». Dans le contexte de la pratique des évaluateurs de programmes gouvernementaux, ces systèmes de représentation (ou théories spontanées) sont centrés sur la faisabilité et la légitimité de l'évaluation, et portent sur trois éléments clés qui permettent de diagnostiquer le contexte, de déduire l'enchainement des actions et d'en anticiper les conséquences possibles : i. les caractéristiques de l'objet évalué ; ii. les stratégies des acteurs autour de l'évaluation (exigences politico-organisationnelles sur le plan de l'évaluation, des attentes, des approches, des méthodes, etc.) ; et iii. les moyens méthodologiques envisagés et les ressources disponibles pour évaluer (Tourmen et Droyer, 2013 ; Tourmen, 2007).

Aussi, la modélisation des processus cognitifs aboutissant à la formulation de théories implicites, et de jugements évaluatifs qui a été développée dans le cadre de l'étude de McPherson (2015, Figure 12 à la page 89), offre une contribution intéressante relativement au processus de conceptualisation en action.

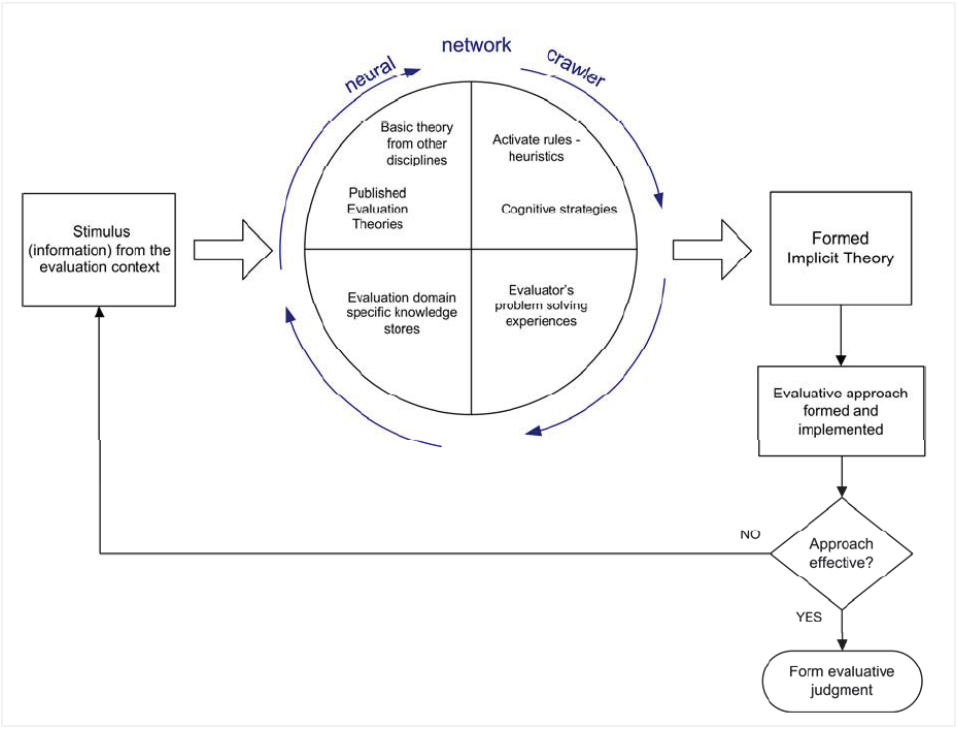


Figure 12. *Processus cognitif aboutissant à la formation de théories implicites (McPherson, 2015).*

Il existe diverses conditions qui permettent le développement de ces ressources cognitives, plus particulièrement le rôle de l'expérience pour apprendre à évaluer. Précisons que la réflexivité joue un rôle crucial dans le cadre l'apprentissage expérientiel, et il importe que l'évaluateur puisse tirer des leçons de ses expériences et leurs résultats afin de développer des représentations de ce qui fonctionne efficacement selon les situations, de même que des raisonnements d'anticipation et des conceptualisations opératoires des situations d'évaluation (Tourmen, 2009, 2007). Dans le contexte du développement professionnel des évaluateurs débutants, il importe que ces derniers puissent assister, et ce, à multiples reprises, à des projets d'évaluation du début à la fin, et qu'ils soient confrontés à des situations variées. Tourmen et Droyer (2013) parlent du rôle de la transmission directe des ressources cognitives sur les lieux de travail avec notamment l'explicitation des théories implicites mobilisés par les évaluateurs expérimentés pour prendre différentes décisions méthodologiques à la lumière des caractéristiques de la situation évaluative rencontrée (transmission orale), ou indirecte (ouvrages, conférences, ateliers, etc.). McPherson (2015) souligne lui aussi le rôle essentiel du mentorat, de la transmission des théories implicites et des schèmes opératoires des évaluateurs experts aux évaluateurs novices ou moins expérimentés.

Dans l'ensemble, on rejoint d'ailleurs ici les résultats de nombreux travaux qui citent parmi les stratégies pédagogiques les plus efficaces pour le développement des compétences des jeunes évaluateurs, le travail de terrain, le mentorat et les activités d'analyse réflexive, et ce, idéalement associés en différentes combinaisons (Holbrook et Chen, 2017 ; Dillman, 2012 ; Kelly et Kaczynski, 2008 ; Christie et Rose, 2003 ; Levin-Rozalis et Rosenstein, 2003).

2.3.3 Synthèse conceptuelle

Les caractéristiques du profil des évaluateurs de programme, plus particulièrement son niveau d'expertise semblent jouer un rôle déterminant dans la pratique professionnelle de ces derniers et leurs capacités d'adaptation à un environnement évaluatif donné. Aussi, dans le cadre de notre étude, nous nous intéresserons aux choix méthodologiques de l'évaluateur, selon qu'il soit novice ou expert, et de l'influence que les facteurs contextuels tels que l'influence des parties prenantes clés et la gestion des contraintes inhérentes au terrain peuvent avoir sur l'opérationnalisation desdits choix.

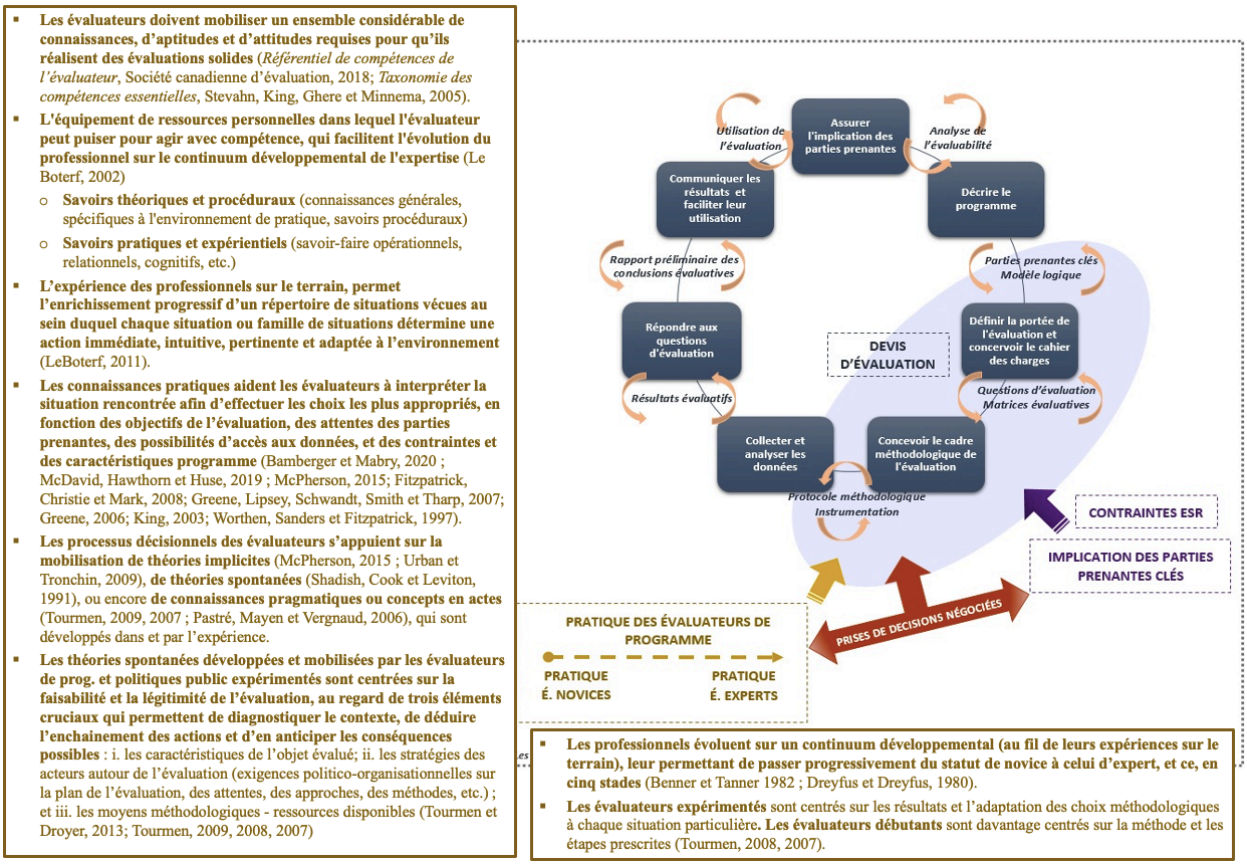


Figure 13. *Synthèse conceptuelle (3/3)*

2.4 Synthèse générale du cadre conceptuel

À travers ce deuxième chapitre, nous voulions explorer les enjeux méthodologiques (et pratiques) associés aux dimensions sociale et politique de l'activité d'évaluation de programme. Un certain nombre de facteurs contextuels sont susceptibles d'avoir des répercussions plus ou moins considérables sur le développement et la mise en œuvre de la méthodologie évaluative ainsi que sur la qualité et l'utilité des résultats. Nous avons montré qu'aujourd'hui l'évaluateur doit être en mesure de produire une évaluation rigoureuse en s'assurant à la fois de collaborer efficacement avec les parties prenantes clés tout au long de l'activité évaluative, d'adapter ses choix méthodologiques en fonction des contraintes sociopolitiques, budgétaires, temporelles et d'accès aux données. La qualité de l'évaluation est dès lors tributaire de la capacité de l'évaluateur à mobiliser des savoirs pratiques en situation réelle, son expertise, et son jugement afin d'entreprendre les choix méthodologiques les plus appropriés en contexte. Un raisonnement méthodologique qui s'appuie, entre autres, sur la mobilisation de théories implicites ou spontanées acquises dans et par l'expérience de l'évaluateur.

À la lumière des éléments qui ont été abordés, notre recherche doctorale vise à explorer et à décrire les choix méthodologiques opérés par les évaluateurs de programme qu'ils soient novices ou experts. Elle se concentre plus particulièrement sur les processus décisionnels engagés dans le cadre de la conception et de la mise en œuvre du devis d'évaluation, ainsi que sur les facteurs contextuels qui ont une influence sur lesdits processus (les contraintes inhérentes au terrain incluant les normes sociopolitiques de l'environnement évaluatif, la collaboration avec les parties prenantes clés et les risques de mésévaluation et de mésusage).

Précisons qu'afin d'être en mesure de comprendre et de décrire pleinement les pratiques évaluatives ainsi que l'influence des facteurs contextuels sur lesdites pratiques et les stratégies mise en œuvre pour y faire face, nous avons décidé : 1) d'examiner à la fois les perceptions des

évaluateurs quant aux « bonnes » pratiques évaluatives (ou pratiques recommandées) ; 2) les explications fournies quant aux choix méthodologiques opérés (pratiques réelles)[17].

Enfin, pour conclure ce chapitre, nous proposons une synthèse des concepts centraux de notre recherche et de leurs interactions, et ce, à l'appui des points saillants des écrits recensés, qui rendent compte de leurs caractéristiques et de leur relation entre eux.

<u>Les choix dits méthodologiques</u>

- Le devis d'évaluation constitue l'empreinte méthodologique de l'évaluation et il doit rendre compte avec transparence, clarté et précision de l'ensemble de la démarche (Rog, 2005). Il est constitué du cadre d'évaluation et du cadre méthodologique.
  - Le *cadre d'évaluation* définit les paramètres de l'évaluation. Il fait d'ailleurs office de cahier des charges. Il décrit les bornes de l'évaluation, c'est-à-dire les objectifs de l'évaluation ainsi que de sa fonction, le type d'évaluation et l'approche évaluative sélectionnés ainsi que les questions principales d'évaluation, l'identification des sources de données, etc. (*Better Evaluation*, 2014; Alkin, 2011, 2009; Stufflebeam, 2004 ; Darabi, 2002). Il est essentiel qu'il propose un échéancier et une estimation des ressources matérielles, humaines et financières nécessaires à la réalisation de l'évaluation. Le principal extrant est la matrice d'évaluation où, pour chaque question d'évaluation, sont référencés les indicateurs de mesure, les sources d'information, et les méthodes de collectes de données envisagées qui permettront d'y répondre (McDavid, Huse et Hawthorn, 2019 ; Secrétariat du trésor du Québec, 2018). C'est un excellent outil de planification, de collaboration et de communication avec les clients et c'est un outil indispensable pour la conception du cadre méthodologique.
  - *Le cadre méthodologique* est une forme de protocole d'évaluation. Il offre un plan détaillé et opérationnel des procédures de collecte et d'analyse des données, de communication de

[17] Nous élaborons davantage sur notre conception des termes « bonnes pratiques » (ou pratiques recommandées) et « pratiques réelles » dans le cadre méthodologique de cette thèse doctorale (Chapitre 3).

l'évaluation et de l'utilisation des résultats, ainsi que l'ensemble de l'instrumentation (Alkin, 2011 ; Stufflebeam, 2004).

- Les choix dits méthodologiques entrepris dans le cadre de la conception de l'évaluation ont trait aux activités suivantes :

Décisions méthodologiques engagées dans la définition du cadre d'évaluation	1. Déterminer les objectifs principaux de l'évaluation 2. Déterminer la fonction de l'évaluation, le type d'évaluation et l'approche évaluative 3. Définir les questions principales d'évaluation 4. Développer la matrice d'évaluation 5. Finaliser le cahier des charges *(rétroaction parties prenantes clés, analyse de la faisabilité, approbation)*
Décisions méthodologiques engagées dans la conception du cadre méthodologique de l'évaluation	6. Définir les sources d'information, les échantillons et les procédures de recrutement *(pour chaque collecte de données)* 7. Définir les indicateurs de mesure et d'observation 8. Déterminer les méthodes de collecte de données et développer l'instrumentation, et de synthèse des résultats 9. Déterminer les procédures d'analyse et de synthèse des données 10. Planifier la communication et l'utilisation des résultats

→ La Figure 6, p. 47 offre un étayage des choix méthodologiques entrepris pour chacune des activités énoncées ci-dessus.

- La démarche d'évaluation et l'ensemble des processus de sa mise en œuvre s'inscrivent dans le respect des normes de pratiques professionnelles du Joint Committee *Joint Commitee on Standards for Educational Evaluatio*n (2011) : utilité, faisabilité, convenance, précision-justesse et imputabilité.
- Les choix méthodologiques s'inscrivent dans un processus itératif susceptible de changer et d'évoluer au fil de la conception et du déroulement de l'évaluation. Les conditions environnementales, la nécessité de s'adapter au contexte politique et organisationnel, la gestion des contraintes de terrain, entre autres, conduisent très souvent l'amendement devis en cours de route, à improviser pendant la collecte de données afin de s'assurer que les résultats évaluatifs soient crédibles et défendables (McDavid, Hawthorn et Huse, 2019 ; Rossi, Lipsey et Freeman, 2019 ; Darabi, 2002).

La relation/l'implication des parties prenantes clés

- L'implication, dans le processus évaluatif, des individus ayant une responsabilité directe au niveau du programme (créateurs du programme, administrateurs, gestionnaires et autres acteurs impliqués dans le programme), constitue un facteur contextuel majeur. Ces individus sont les parties prenantes clés ou principaux utilisateurs ciblés (Patton, 2012 ; Greene, 2005b).
- Une collaboration plus ou moins étroite avec les parties prenantes clés facilite la compréhension du programme, du contexte évaluatif et des attentes/besoins des utilisateurs ciblés, l'identification et la gestion des contraintes de terrain, la conception et la mise en œuvre du devis d'évaluation et la formulation de recommandations utiles et utilisables (Bamberger et Mabry, 2020 ; Chen, 2015 ; Ridde et Dagenais, 2012 ; Bryson, Patton et Bowman, 2011 ; Morris, 2002 ; Brandon, 1998).
- La faisabilité, la rigueur et l'utilité de l'évaluation sont étroitement liées au caractère efficace de la collaboration avec les parties prenantes clés (à la qualité des interactions), et sont donc tributaires de la capacité des évaluateurs de programme à : faire preuve de crédibilité, communiquer clairement et efficacement les informations, créer un climat relationnel de confiance, faire preuve d'écoute, de flexibilité et de sensibilité au contexte, reconnaitre et gérer la manifestation de comportements réfractaires (résistance, désengagement, évitement, détournement, etc.) (Marchand, 2020 ; Dubeau, Bérubé, Lafantaisie, Coutu et Turcotte, 2018 ; Stevahn et King, 2016 ; Dumaine, 2018 ; Patton, 2018 ; Hawkins, 2012 ; Alkin, 2005 ; Taut et Alkin, 2003 ; Donaldson, Gooler et Scriven, 2002).
  - L'attitude des parties prenantes clés envers l'évaluation est un facteur déterminant en ce qui a trait à l'apparition de signes défensifs. Le manque d'expérience avec l'évaluation de programme, les expériences antérieures négatives, le manque de confiance envers l'évaluation et l'évaluateur, l'adoption des préjugés d'autres individus, la crainte de répercussions négatives, l'attachement au programme, entre autres, peuvent les conduire à anticiper la manifestation de conséquences négatives, ce qui génère un état émotionnel négatif (peur de la perte de contrôle, peur du jugement évaluatif, peur de la perte de pouvoir) (Taut et Alkin, 2003 ; Donaldson, Gooler et Scriven, 2002).

- Les caractéristiques de l'environnement évaluatif telles que les normes sociales et la culture organisationnelle du milieu, l'ambiguïté des rôles entre l'évaluateur et les parties prenantes, le manque de reconnaissance des accomplissements du programme, etc. exercent également une influence sur l'attitude réfractaire, la manifestation de signes défensifs chez les parties prenantes clés (Bamberger et Mabry, 2020 ; Donaldson, Gooler et Scriven, 2002).
- Trois volets de compétences spécifiques de l'évaluateur de programme se trouvent au cœur des préoccupations des parties prenantes clés, de leur perception de la crédibilité de l'évaluation : ses compétences sociales, ses connaissances sur le programme évalué et ses caractéristiques contextuelles, ainsi que ses compétences méthodologiques, c'est-à-dire sa capacité à concevoir, mettre en œuvre et gérer l'évaluation (Marchand, 2020 ; Taut et Alkin, 2003).
- La manière dont les parties prenantes perçoivent l'évaluateur influence leurs degrés d'intérêt et d'engagement envers l'évaluation. Sur le terrain les étudiants-évaluateurs peuvent être mis en difficulté et avoir du mal à établir et maintenir ce qu'ils appellent les exigences idéales pour être un évaluateur efficace, car ils ne sont pas capables de faire suffisamment preuve d'autorité et de crédibilité, et d'accéder aux parties prenantes (Chouinard et *coll*., 2017).
- La capacité de l'évaluateur de programme à établir et maintenir une dynamique relationnelle constructive avec les parties prenantes clés, un dialogue de qualité, à faire preuve de flexibilité et d'adaptabilité, à agir comme conseiller, montre le caractère essentiel du volet de compétence *relations interpersonnelles* chez l'évaluateur, soit la collaboration, la communication, la négociation, la gestion et la résolution de conflit (Marchand, 2020 ; Dumaine, 2018 ; Patton, 2018 ; King et Alkin, 2016).

- Le renforcement de l'utilisation de l'évaluation et l'omniprésence des principaux utilisateurs ciblés sont accompagnés de risques de mésévaluation (*misevaluation*) et de mésusage (*misuse*). Des risques qui touchent principalement la qualité des opérations d'évaluation, l'utilité et l'utilisation du processus et des résultats évaluatifs (Alkin et King, 2017 ; Patton, 2015a, 2008 ; Alkin et Coyle, 1998). → Voir notre modélisation des concepts de mésévaluation et de mésusage à la page 65 (Figure 10).

Les contraintes inhérentes au terrain ou les contraintes de l'évaluation en situation réelle

- Les évaluateurs doivent gérer de nombreux défis liés aux intrants disponibles pour l'évaluation, dans un contexte où les clients commandent des évaluations avec des contraintes temporelles et budgétaires plus ou moins importantes, et où l'accès à des données probantes de qualités peut s'avérer compliqué (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019 ; Bamberger, Rugh et Mabry, 2012).
- Les évaluateurs doivent mobiliser des stratégies adaptatives d'ordre méthodologique en réponse aux contraintes de l'évaluation en situation réelle (Bamberger et Mabry, 2020) → Contraintes ESR : contraintes temporelles, contraintes budgétaires, contraintes d'accès aux données, contraintes sociopolitiques et organisationnelles (voir notre synthèse des stratégies adaptatives en réponse aux contraintes ESR dans le Tableau 3, p. 68-69).
- Le poids des contraintes sociopolitiques et organisationnelles sur la qualité et l'utilité de l'évaluation est un élément caractéristique des situations évaluatives rencontrées dans le contexte de l'évaluation de programmes publics.
  - Les attentes associées au mandat des politiques évaluatives gouvernementales (évaluation fondée sur les données probantes) réactivent les débats autour de la question des méthodes appropriées pour inférer le lien de causalité entre le programme et les résultats observés, et du maintien comme *gold standard* du devis expérimental ou quasi expérimental pour l'évaluation de l'efficacité (McDavid, Huse et Hawthorn, 2019 ; Christie et Fleisher, 2010 ; Cook, Scriven, Coryn et Evergreen, 2010 ; Donaldson et Christie, 2005).
  - Les questions liées à l'efficacité demandent le recours à des devis sophistiqués qui répondent aux normes conventionnelles pour évaluer les relations de cause à effet entre le programme et ses résultats, mais la réalité de terrain et les contraintes qui l'occupent, particulièrement sur le plan des données probantes, posent de gros problèmes de faisabilité et de fiabilité (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019).

    Une approche pragmatiste est recommandée, soit le recours à un devis d'évaluation qui mobilise une approche multiméthodes, mise sur la complémentarité des méthodes quantitatives et qualitatives, s'appuie sur le jugement de l'évaluateur pour déterminer les relations de causalité, et assure un processus rigoureux et systématique de triangulation (McDavid, Huse et Hawthorn, 2019 ; Greene, 2015 ; Stufflebeam et Coryn, 2014).

- Le calibrage de l'évaluation de programmes publics est souvent inadéquat au regard des ressources disponibles et mobilisables pour répondre au mandat d'évaluation. Malgré un échéancier trop serré et un budget insuffisant, il est attendu de livrer une évaluation fiable, fondée sur des données probantes pour pallier efficacement les difficultés importantes d'accès aux données (Lamari, Poulin-Larivière et Lemieux, 2016).
- L'obligation pour les ministères du gouvernement fédéral canadien d'évaluer l'ensemble de leurs programmes tous les cinq ans, et de répondre à l'ensemble des questions fondamentales sur leur rendement (efficacité, efficience et économie) semble poser des problèmes sur le plan de la faisabilité et de l'utilité de l'évaluation (nature et maturité du programme, échéanciers serrés, budgets insuffisants, problèmes d'accès aux données de rendement et aux données financières, etc.) (Secrétariat du Trésor du Canada, 2015).
- Les lacunes souvent observées semblent concerner l'accessibilité, la qualité et l'exhaustivité des données probantes quant à l'efficacité des programmes publics fédéraux (Secrétariat du Trésor du Canada, 2015 ; Bureau du vérificateur général du Canada, 2013).
- Les questions relatives à la mesure de l'atteinte des résultats ultimes/de l'impact (attribution chiffrée, recours à un devis expérimental ou quasi expérimental), et de l'efficience et l'économie (analyse systématique de la relation coût-bénéfices, coût-efficacité) de programmes fédéraux posent de gros problèmes de faisabilité, et constituent des attentes irréalistes dans le contexte de la réalisation de l'évaluation quinquennale des programmes fédéraux (Dumaine, 2018).
- Les évaluateurs doivent adapter leurs pratiques évaluatives face à une variété d'influences et de contraintes dans un environnement organisationnel habité par de nombreux acteurs et où le fonctionnement administratif est souvent rigide et inadapté pour accommoder les stratégies méthodologiques envisagées en réponse aux contraintes (Bamberger, Rugh, Church et Fort, 2004 ; Bamberger, 2000).

- « L'évaluateur met toujours en jeu sa compétence à analyser avec rigueur, cohérence et justesse une action ou un projet, étant donné les contraintes inhérentes au milieu, c'est-à-dire les attitudes parfois réfractaires du client ou encore les conditions défavorables (délai imposé, limite budgétaire) dans l'exercice de son mandat. » (Beaudoin, Lefrançois et Ouellet, 1986, p.192).

L'expertise et l'apport de l'expérience pour réaliser des évaluations pertinentes et de qualité

- Les évaluateurs doivent mobiliser un ensemble considérable de connaissances, d'aptitudes et d'attitudes requises pour qu'ils réalisent des évaluations solides, qui reflètent les normes, les principes et lignes directrices approuvés par les principales associations d'évaluation nord-américaines (*Taxonomie des compétences essentielles*, Stevahn, King, Ghere et Minnema, 2005). Le référentiel réparti 61 compétences sur six volets ou domaines de compétences : (1) pratique professionnelle ; (2) enquête systématique ; (3) analyse situationnelle ; (4) gestion de projet ; (5) relations interpersonnelles ; et (6) pratique réflexive. → Voir le référentiel de la Société canadienne d'évaluation (2018) adaptée de celui de Stevahn, King, Ghere et Minnema (2005) à la page 83 (Tableau 6).
    - De nouveaux évaluateurs entrent constamment sur le terrain, soit par l'entremise de diplôme universitaire en évaluation de programme, soit *« accidentellement »* dans le cadre de leur poste, en assumant des responsabilités d'évaluation avec peu ou pas de préparation (Ghere, King, Stevahn et Minnema, 2006). Dans un tel contexte, certains évaluateurs sont conscients qu'il existe des normes et des lignes directrices en évaluation de programme, mais d'autres ne le sont pas (King et *coll.*, 2001).
- Les professionnels évoluent sur un continuum développemental (au fil de leurs expériences sur le terrain), leur permettant de passer progressivement du statut de novice à celui d'expert, et ce, en cinq stades (Benner et Tanner 1982 ; Dreyfus et Dreyfus, 1980). → Voir notre modélisation dans le Tableau 4, pages 77.
- « Une personne compétente est une personne qui sait agir avec pertinence dans un contexte particulier, en choisissant et en mobilisant un double équipement de ressources : ressources personnelles (connaissances théorique et pratique, savoir-faire, qualités, culture, ressources émotionnelles...) et ressources de réseaux (banques de données, réseaux documentaires, réseaux d'expertise, etc.). » (Le Boterf, 2002, p. 46). → Voir notre description de l'équipement de ressources personnelles dans lequel le professionnel peut puiser pour agir avec compétence, et progresser sur le continuum du développement de l'expertise dans le Tableau 5, pages 79.
    - Savoir agir avec compétence c'est être capable de (1) construire et d'activer des combinatoires pertinentes de ressources incorporées ou internes (personnelles) et de ressources externes ou de réseaux (2) pour gérer une gamme de « situation(s)

professionnelle(s) » étant décrite avec une activité ou une mission à laquelle sont associés des « critères de réalisation » contextualisés (3) en vue d'atteindre des « résultats » (produits, services) auxquels sont associés des « échelles de réussites » et adressés à un « destinataire » (clients, usagers, patients...), afin de (4) répondre à des « enjeux » (Le Boterf, 2011, 2002).

- La pratique professionnelle des évaluateurs de programme évolue sur un continuum développemental, au fil de l'expérience des professionnels sur le terrain, permettant l'enrichissement progressif d'un répertoire de situations vécues au sein duquel chaque situation ou famille de situations détermine une action immédiate, intuitive, pertinente et adaptée à l'environnement.
  - Les connaissances pratiques aident les évaluateurs à interpréter la situation dans laquelle il se trouve afin d'effectuer les choix les plus appropriés, en fonction des objectifs de l'évaluation, des attentes des parties prenantes, des possibilités d'accès aux données, et des contraintes et des caractéristiques du milieu (Bamberger et Mabry, 2020 ; McDavid, Hawthorn et Huse, 2019 ; McPherson, 2015 ; Fitzpatrick, Christie et Mark, 2008 ; Greene, Lipsey, Schwandt, Smith et Tharp, 2007 ; Greene, 2006 ; King, 2003 ; Worthen, Sanders et Fitzpatrick, 1997 ; Stone, 1984 [cité par Nadeau, 1988]).
  - *Les évaluateurs expérimentés* ont tendance à être centrés sur les résultats et l'adaptation à chaque situation particulière (ils sont attentifs à l'environnement, au contexte, aux contraintes ; les choix méthodologiques opérés s'inscrivent dans un diagnostic approfondi d'une situation évaluative donnée ainsi qu'un pronostic de son évolution). *Les évaluateurs débutants,* quant à eux, sont davantage centrés sur la méthode et les étapes prescrites (ils sont collés aux concepts étape par étape dans leur démarche d'évaluation ; leur pratique s'inscrit dans une logique d'imitation, et ils sont déroutés par les variations d'une situation) (Tourmen, 2008, 2007).
  - Les processus décisionnels des évaluateurs s'appuient sur la mobilisation de théories implicites (McPherson, 2015 ; Urban et Tronchin, 2009) ou spontanées (Shadish, Cook et Leviton, 1991), ou encore de connaissances pragmatiques/concepts en actes (Tourmen, 2009, 2007 ; Pastré, Mayen et Vergnaud, 2006), développées dans et par l'expérience.

- Les théories spontanées sont des systèmes de représentation en actes composés de « propositions tenues pour vraies sur le réel », et de raisonnements conditionnels de type « si..., alors... » ou de règles d'action qui permettent l'adaptation, le diagnostic et l'anticipation (Tourmen et Droyer, 2013).
- Dans le contexte de la pratique des évaluateurs de programmes et politiques publics, les théories spontanées développées et mobilisées par les évaluateurs expérimentés sont centrées sur la faisabilité et la légitimité de l'évaluation, et portent sur trois éléments cruciaux qui permettent de diagnostiquer le contexte, de déduire l'enchainement des actions et d'en anticiper les conséquences possibles : i. les caractéristiques de l'objet évalué ; ii. les stratégies des acteurs autour de l'évaluation (exigences politico-organisationnelles sur la plan de l'évaluation, des attentes, des approches, des méthodes, etc.) ; et iii. les moyens méthodologiques sélectionnés (pertinence du choix des méthodes, des indicateurs, etc.) - ressources disponibles (Tourmen et Droyer, 2013 ; Tourmen, 2009, 2008, 2007).

CHAPITRE 3. POSTURE ÉPISTÉMOLOGIQUE ET CADRE MÉTHODOLOGIQUE

Dans le cadre de ce troisième chapitre, nous décrivons le cadre méthodologique mobilisé pour mener cette recherche doctorale et ainsi répondre à nos questions de recherche. Premièrement, avant d'entrer dans le vif du sujet de la méthodologie, nous introduisons le paradigme de recherche que nous avons adoptée et qui oriente les présuppositions épistémologiques relatives à l'objet de recherche, ainsi que la sélection du devis de recherche, l'instrumentation employée, le choix des participants, l'analyse et l'interprétation des données (Denzin et Lincoln, 2011). Deuxièmement, nous présentons le devis de recherche sélectionné, à savoir l'étude de cas multiples. Troisièmement, nous revenons sur les critères et les procédures de sélection des participants de la recherche. Enfin, nous offrons une description détaillée du déroulement de la recherche. En outre, nous avons organisé cette partie par collectes de données (qui sont au nombre de deux), chacune décrivant l'instrumentation et les procédures de collecte et d'analyse de données.

3.1 Positionnement épistémologique

Par la présente recherche, nous proposons de décrire et de mieux comprendre les processus d'opérationnalisation des choix méthodologiques des évaluateurs dans le cadre de leurs opérations d'évaluation de programme et l'influence que certains facteurs contextuels peuvent avoir sur ces choix. Selon Denzin et Lincoln (2011), la formulation d'un tel objectif évoque le recours aux méthodes de recherche dites qualitatives, méthodes qui produiront des données empiriques de type qualitatives dédiées à la description et à l'interprétation de l'expérience des acteurs de la recherche, dans un contexte spécifique. De surcroit, le caractère complexe de notre problématique ainsi que notre volonté de décrire un phénomène particulier propre à l'évaluation, observé sur le terrain, mais peu décrit par la recherche en évaluation, nous a conduits à entreprendre une recherche exploratoire, par extension descriptive. En effet, selon Gauthier (2008), une recherche de nature exploratoire se caractérise essentiellement par les thèmes peu connus qu'elle vise, qui ont fait état de peu d'analyse et dont le chercheur ne parvient pas à réaliser un portrait à partir des connaissances existantes. La démarche d'une telle recherche s'associe avec une approche descriptive qui se concentre sur les mécanismes et sur les acteurs (soit le « comment », et le « qui » des phénomènes) de la problématique à l'étude.

Au regard de la perspective interprétative/qualitative de notre recherche, nous avons adopté une posture paradigmatique interprétative. Une telle épistémologie conduit le chercheur interprétatif à adopter une vision relativiste du monde social, un monde constitué de multiples réalités subjectives tout aussi valides et valables (Schwandt, 2000). Ainsi, l'objectif du chercheur n'est pas d'expliquer la réalité, mais de la comprendre au travers des interprétations que les acteurs font de cette dernière. Au regard de l'interdépendance entre l'objet (le phénomène), les acteurs et le chercheur, la réalité est donc objectivée et intersubjectivement construite.

Plus concrètement, dans le cadre d'une recherche interprétative, le chercheur s'intéresse à comprendre les significations que les acteurs attribuent à la situation à l'étude (Savoie-Zajc, 2018 ; Anadòn et Savoie-Zajc, 2009). En effet, le sens que les acteurs donnent à leurs conduites constitue la base des connaissances scientifiques de la recherche interprétative. C'est à partir de ces connaissances que le chercheur peut recréer, se représenter, interpréter et comprendre les actions des acteurs dans un contexte spécifique bien déterminé (Anadòn, 2006). En outre, ce dernier doit veiller à prendre en considération (et accepter) un ensemble d'éléments subjectifs propre à l'acteur (ses intentions, ses motivations, ses intérêts politiques et sociaux, etc.) qui se concentre davantage sur les pratiques que sur les faits et qui oriente les interprétations des acteurs (Anadòn, 2006 ; Pourtois et Desmet, 1988). Ainsi les réalités subjectives et intersubjectives sont-elles considérées, par le chercheur interprétatif, comme des données incontournables pour la construction des savoirs, et non comme des obstacles à la production de connaissances (Anadòn et Guillemette, 2007).

Enfin, la dimension empathique de la recherche interprétative joue un rôle important dans la démarche interprétative et la production de connaissances scientifiques valables en ce qui a trait au contexte spécifique de la situation à l'étude (connaissances non généralisables). En effet, selon Perret et Séville (2003), pour comprendre les significations et les interprétations produites par les acteurs vis-à-vis de la situation à l'étude ainsi que les réalités sociales vécues et perçues par ces derniers, le chercheur interprétatif doit s'approprier le langage des acteurs et développer auprès d'eux et avec eux une certaine proximité. En tant que chercheur interprétatif, il est alors important d'allier à la fois subjectivité et objectivité, c'est-à-dire de porter notre regard sur l'objet d'étude, de se rapprocher de l'acteur et du terrain d'étude, tout en se distanciant de celui-ci afin d'être en mesure d'objectiver nos données (Savoie-Zajc, 2018).

Dans le cadre de cette recherche, le recueil et l'interprétation des significations que les évaluateurs de programme attribuent à leurs actions dans le cadre de leurs pratiques évaluatives nous permettent de comprendre de « l'intérieur » les mécanismes sous-jacents aux prises de décisions dites méthodologiques. Aussi sommes-nous en mesure d'explorer et de décrire en profondeur l'influence que les facteurs contextuels, soit le profil des évaluateurs, la relation avec les parties prenantes, les contraintes inhérentes au terrain, peuvent avoir sur l'opérationnalisation des choix méthodologiques des évaluateurs dans le cadre de leurs opérations d'évaluation.

L'orientation exploratoire et descriptive de cette recherche ainsi que l'adoption d'une posture interprétative (et d'une méthode qualitative), doit permettre l'émergence de pistes de recherche et une meilleure compréhension des choix méthodologiques en évaluation et des facteurs susceptibles de les influencer. Des retombées pratiques, telles que des propositions méthodologiques, pourront être offertes aux praticiens, ce qui permettra notamment de réduire les écarts entre la théorie et la pratique, et de favoriser la réalisation d'évaluations de programme efficientes et de qualité.

3.2 Devis de recherche : l'étude de cas

Compte tenu de la nature exploratoire (et descriptive) de notre recherche, de notre posture épistémologique interprétative et de notre problématique de recherche (qui se concentre sur le « comment » et les acteurs du phénomène à l'étude), nous avons sélectionné l'étude de cas comme devis de recherche. Le choix de ce devis nous parait pertinent puisqu'il s'appuie sur les recommandations de Yin (2003) qui stipulent que le recours à l'étude de cas est pertinent si l'on s'intéresse au « comment » et au « pourquoi » des phénomènes propres à une situation où les chercheurs ont peu de contrôle sur les évènements observés. L'auteur nous montre ainsi que cette méthode de recherche qualitative permet de rapporter une situation réelle (prise dans son contexte) ainsi que son analyse, afin d'appréhender la manifestation et l'évolution d'un phénomène (et ses processus particuliers) auquel le chercheur s'intéresse. De plus, rappelons que l'objectif de notre recherche est de décrire les pratiques des évaluateurs de programme, novices et experts, et plus particulièrement leurs choix méthodologiques et l'influence possible de facteurs contextuels sur ces processus décisionnels lors de leurs opérations d'évaluation. Cet objectif rejoint celui de Baxter et Jack (2008) qui avait eu recours à l'étude de cas comme méthode de recherche dans le cadre de

leur étude sur le processus de prise de décision des étudiantes infirmières afin de déterminer le type de décisions prises par ces futures praticiennes et les facteurs qui influencent ces prises de décision ; ce qui renforce notre choix méthodologique.

Au regard des différentes typologies de l'étude de cas en fonction de l'objet de recherche et des propriétés du cas, notre étude de cas est :

- **À visée descriptive et exploratoire (Gagnon, 2011 ; Yin, 2003, 1994) :** l'objectif est de décrire un phénomène (l'opérationnalisation des choix méthodologiques des évaluateurs de programme) et le contexte réel dans lequel il s'est produit (facteurs contextuels d'une situation évaluative précise). Selon l'auteur, l'apport de la recherche repose sur la description complète et profonde du phénomène dans son contexte, phénomène dont certains aspects demeurent très peu explorés.
- **De type collectif ou multicas (Karsenti et Demers, 2018 ; Stake, 1995) :** l'étude est réalisée auprès de plusieurs cas qui, par leur synthèse, permettront de comprendre le phénomène d'intérêt (la multiplicité des cas permet de faire émerger des caractéristiques communes et divergentes relatives à un phénomène); tout contribuant à la mise en exergue des particularités de chaque cas. La dimension *novice-expert* de notre objectif de recherche nous a conduits à développer un devis d'étude de cas multiples, soit de cas de « novices » et de cas « d'expert ». Puisqu'il s'agit d'une recherche interprétative, les évaluateurs de programme sélectionnés exercent dans un même domaine. Enfin, l'étude de cas multiples renvoie à une approche de recherche flexible qui recourt à la triangulation de sources de données multiples.

D'après les travaux de Roy (2009), Yin (2003, 1994) et Stake (1995), l'inscription du phénomène à l'étude dans son contexte réel, la délimitation des frontières du cas (objet d'étude, acteurs et facteurs) par le cadre conceptuel, et le recours à de multiples cas favorisent la rigueur de l'étude de cas et la scientificité des résultats obtenus. En effet, la description exhaustive de plusieurs cas propres au phénomène d'étude est comprise dans un système intégré et délimité qui permet la mise au jour du phénomène et de ses caractéristiques complexes (Merriam, 1998 ; Stake, 1995). Ainsi, les connaissances scientifiques produites sont concrètes et contextuelles (Merriam, 1998), limitées à un milieu donné, et les interprétations recueillies sont la manifestation d'une expérience (Patton, 2015b ; Savoie-Zajc, 2009). Ceci constitue pour Merriam (1998) la force de la

validité externe de l'étude de cas. En effet, rappelons que les cas individuels (qui constituent les groupes de cas de l'étude de cas multiples) sont sélectionnés afin de comprendre en profondeur une situation d'étude particulière et non pour déterminer une vérité générale. Aussi, l'observation de la construction de la réalité par les acteurs constitue la force de la scientificité de la méthode. On se concentre sur la compréhension des significations et des expériences des acteurs, la découverte de la complexité du phénomène à l'étude et l'interprétation holistique des connaissances émergentes (Merriam, 1998). La construction de la réalité locale repose sur la formulation d'analyse interne à chacun des cas individuels puis de l'analyse externe en regroupant les classes de cas et en étudiant les différences et les similitudes entre ces derniers (Verhoeven, 2010).

En résumé, notre recherche exploratoire s'inscrit dans une perspective interprétative, de type étude de cas contrastants, soit de cas « d'experts » et de cas de « novices », et notre protocole méthodologique[18] s'appuie sur :

- **La définition des frontières des cas (Stake, 1995) :** les choix méthodologiques des évaluateurs, novices et experts, qui exercent dans un même domaine et l'influence des facteurs contextuels (la relation avec les parties prenantes et les contraintes inhérentes au terrain) sur ces choix dans le cadre de leurs opérations d'évaluation de programme.
- **La démarche d'analyse déductive et le processus de généralisation analytique de Yin (2003, 1994) :** si l'étude de cas mobilise un nombre limité de sujets, la collection d'un grand nombre d'informations sur chacun d'eux ainsi que sur le contexte permet d'examiner de multiples variables ainsi que les liens qui les unissent (Karsenti et Demers, 2018). Selon Mucchieli (2007), le processus de généralisation s'inscrit dans la recherche des processus constitutifs du phénomène à l'étude. Aussi, notre processus de généralisation se déroule comme suit :

[18] Vous trouverez en conclusion de ce chapitre (partie 3.5), une représentation graphique du protocole de l'étude de cas multiples réalisée dans le cadre de cette recherche doctorale (Figure 17, p. 138).

(1) le classement méthodique des unités d'informations recueillies dans des matrices conçues à partir des principes et des concepts clés du cadre conceptuel (Miles, Huberman et Saldaña, 2014 ; Miles et Huberman, 1994/2003 ; Huberman et Miles, 1994). Ce premier niveau d'analyse permet de construire des portraits exhaustifs de la pratique de chacun des évaluateurs novices et experts relatifs au processus décisionnel en ce qui concerne l'élaboration et la mise en œuvre du protocole méthodologique ;

(2) la progression à un niveau d'analyse et de généralité plus élevé, avec la réalisation d'un examen intracas à partir des portraits individuels. Cette étape de l'analyse permet l'émergence d'une modélisation « cas de novices » et d'une modélisation « cas d'experts » ;

(3) l'atteinte d'un niveau généralisation supérieure, avec la comparaison intercas à partir des concepts clés et des questions de recherche et l'émergence des similitudes et des différences.

- **La nature des savoirs produits de Merriam (1998) :** la collecte et l'analyse des données ainsi que l'interprétation des résultats permettent la production de savoirs qui enrichissent notre cadre conceptuel par la proposition de modélisation(s) théorique(s) du phénomène étudié (démarche inductive). Aussi, nous sommes en mesure de proposer une description des liens entre les processus décisionnels relatifs à la méthodologie évaluative et les facteurs contextuels suivants : (1) le niveau d'expertise des évaluateurs (2) la relation avec les parties prenantes et (3) les contraintes inhérentes au terrain.

3.3 Participants et processus de recrutement

Avant d'entamer le processus de recrutement des participants, nous avons déposé des demandes d'approbation déontologique auprès du bureau d'éthique et d'intégrité de la recherche de l'Université d'Ottawa et d'une Université du Québec (Annexe 5). Aussi avions-nous préparé, en français et en anglais, les formulaires de consentement de participation à la recherche ainsi que les différentes lettres de présentation de la recherche et les courriels de recrutement (Annexe 6).

Dans le cadre de cette section, nous présentons les procédures de recrutement et de sélection des cas. Nous verrons que le processus de recrutement a été quelque peu laborieux et que nous avons dû mettre en œuvre un certain nombre de scénarios de recrutement pour obtenir notre échantillon de cas. Pour bien rendre compte de ce processus, nous abordons les différentes procédures de recrutement entreprises (ou scénarios) ainsi que l'évolution des critères de sélection des participants, critères qui ont dû être redéfinis pour répondre à la réalité de terrain, notamment en ce qui a trait à la *formation en évaluation de programme*. Enfin, nous présentons les participants qui ont été sélectionnés.

3.3.1 Procédures de recrutement et de sélection des cas

En raison de la nature de notre étude, notre échantillon est de type intentionnel et vise la sélection de cas contrastants (cas novices versus cas experts). Si l'on prend appui sur la recherche doctorale de Tourmen (2007), celle-ci avait étudié la pratique de six évaluateurs débutants et de six évaluateurs expérimentés, au sein d'un centre d'évaluation de programmes et de politiques publiques. Gagnon (2011) recommande qu'une étude de cas multiple comprenne de quatre à dix cas parce qu'il est ainsi possible de « fournir une riche description du contexte dans lequel les évènements se déroulent et de mettre au jour la structure profonde des comportements sociaux » (p. 41). En outre, afin de nous prémunir contre la possibilité de mortalité échantillonnale en cours d'étude, l'auteur préconise que l'échantillon comporte au moins un à deux cas de plus que le nombre requis. De ce fait, en tenant compte des risques de mortalité de l'échantillon, nous avons tenté d'obtenir dix à douze participants *a priori*; de cette manière, si certains d'entre eux n'ont pas été en mesure de compléter l'étude, nous avons tout de même pu la mener à terme avec les cas restants.

Comme nous le verrons dans le cadre de la présentation des différents scénarios de recrutement que nous avons mis en œuvre, nous désirions recruter des évaluateurs de programme novices et experts exerçant dans un même domaine. Au départ, nous avions privilégié le recrutement des participants au sein de programmes de formation universitaire en évaluation de programme et d'un centre de recherche universitaire spécialisé en évaluation de programme. À partir du cadre conceptuel, nous avions établi les critères de sélection des cas suivants : (1) les évaluateurs novices devaient faire partie de la dernière cohorte d'étudiants francophones en

formation en évaluation de programme, et avoir un an (ou moins) d'expérience en évaluation de programme (stage, assistanat de recherche ou réalisation d'une évaluation de programme dans le cadre de leur formation) (2) les évaluateurs experts devaient détenir au moins un diplôme en évaluation de programme, pratiquer depuis au moins dix ans, et potentiellement contribuer au domaine de l'évaluation de programme (participation à des colloques scientifiques et professionnels, publication d'articles, enseignement, etc.)[19]. Cependant, comme nous l'annoncions en début de section, nous avons dû entreprendre la mise en œuvre d'autres scénarios de recrutement, procédures de recrutement alternatives qui nous ont d'abord conduits à adapter, puis à revoir nos critères de sélection des cas de cette recherche.

En ce qui a trait à la sélection des participants[20], novices et experts, nous avons travaillé de concert avec les directeurs ou autres membres de la direction des différents sites de recrutement. Ainsi, ces derniers nous ont recommandé des évaluateurs qui répondent à nos critères de sélection et ils ont facilité notre entrée en contact avec ces derniers. Selon Gagnon (2011), cela a contribué à les mettre en confiance, et nous a permis de plus « facilement » les convaincre de souscrire à l'étude. De plus, ce travail étroit avec les gestionnaires nous a permis d'avoir accès à l'ensemble des informations dont nous avions besoin, notamment les rapports d'évaluation de programme auxquels les évaluateurs participants ont contribué et que nous utilisons dans le cadre de la deuxième collecte de données.

3.3.1.1 Scénarios de recrutement en milieu universitaire : procédures et critères de sélection des cas

Pour les procédures de recrutement, celles-ci ont commencé par l'envoi d'un courriel d'invitation à participer à notre recherche doctorale au sein duquel nous nous présentions

[19] Dans le cadre de la recherche de Tourmen (2007), il s'agissait d'évaluateurs expérimentés, possédant un diplôme en évaluation avec au moins cinq ans d'expérience, ce qui les plaçait au niveau du stade 4 du développement professionnel de Dreyfus et Dreyfus (1980).

[20] Processus de sélection similaire dans la recherche doctorale de Tourmen (2007).

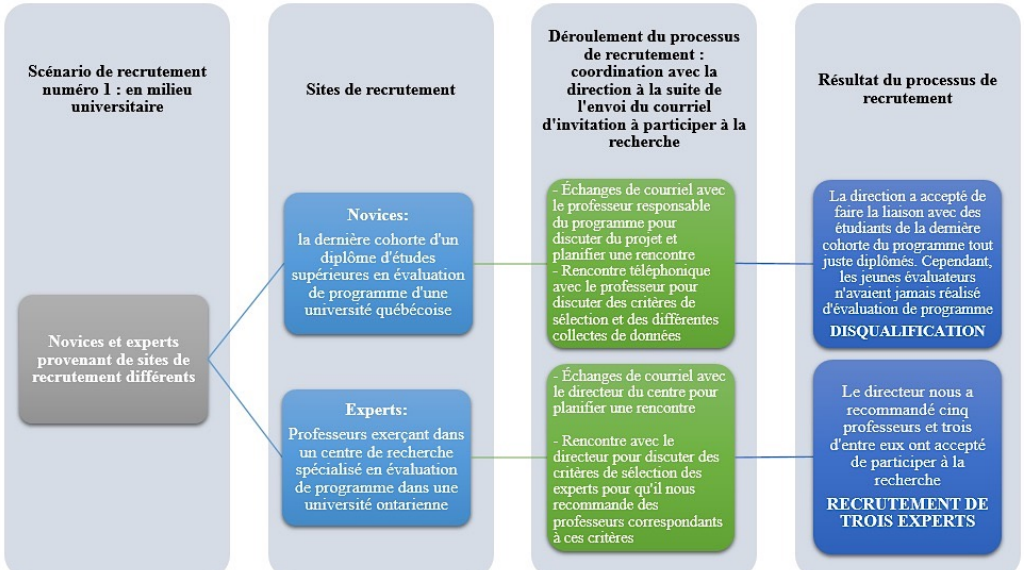
brièvement et nous faisions part des objectifs de notre recherche et du type de participants recherché. Nous avions également inclus un document qui décrit la recherche, les critères de sélection des participants, les bienfaits liés à la participation à notre recherche ainsi que la collecte de données (Annexe 6).

Pour le recrutement des participants en milieu universitaire, nous avions opté pour deux scénarios. Dans le cadre du premier scénario, les évaluateurs novices et les évaluateurs experts devaient provenir de deux sites de recrutement différents : les novices d'une université québécoise et les experts d'une université ontarienne (Encadré 2). Dans le cadre du second, l'ensemble des participants faisait partie du même site de recrutement, une université ontarienne (Encadré 3, p.109). Les critères de sélection des participants correspondaient aux critères de sélection initiaux, nous les avions juste adaptés au milieu universitaire en question. Nous présentons ci-dessous les critères de sélection des participants ainsi qu'une schématisation succincte du processus mis en œuvre dans le cadre du premier scénario de recrutement (Encadré 2) :

Encadré 2. *Procédures du 1er scénario de recrutement en milieu universitaire*

<u>Évaluateurs novices</u> :
- Être diplômé de la dernière cohorte du programme d'études supérieures en évaluation de programme ou ancien étudiant ayant obtenu son diplôme en évaluation de programme l'année précédente.
- Avoir un an d'expérience (ou moins) en évaluation de programme (stage, réalisation d'une évaluation de programme dans le cadre de cours du diplôme ou réalisation d'une à deux évaluations de programme).
- Parler français.

<u>Évaluateurs experts</u> :
- Détenir un ou plusieurs diplômes en évaluation de programme.
- Avoir au moins dix ans d'expérience en évaluation de programme.
- Démontrer un certain degré de contribution au domaine ***(optionnel)***.
- Parler français.



Nous avons dû disqualifier les potentiels participants du groupe novices, car la seconde collecte de données nécessite l'accès à un rapport d'évaluation produit par le participant. Or, la directrice du programme nous a mentionné que les étudiants ne réalisent pas d'évaluation dans le cadre de leur programme de formation en évaluation de programme. Ce fut pour nous un élément surprenant puisque d'autres programmes de formation de ce type ont une composante « pratique de terrain ». La disqualification des participants novices potentiels nous a conduits à opter pour le second scénario de recrutement, qui implique le recrutement d'évaluateurs novices sur le même site que les évaluateurs experts recrutés dans le cadre de la mise en œuvre du premier scénario. Aussi, les critères de sélection des évaluateurs novices ont dû être adaptés pour correspondre au nouveau milieu universitaire et nous avons mis en œuvre le second scénario (Encadré 3) :

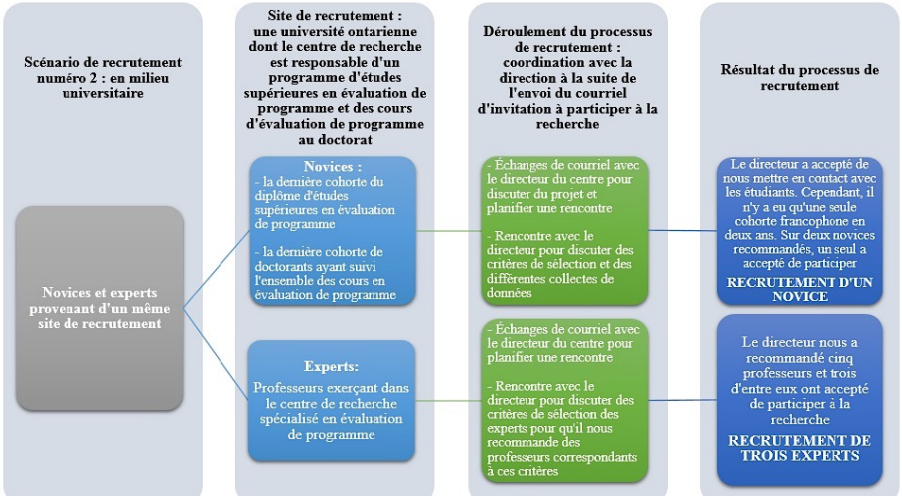
Encadré 3. *Procédures du 2e scénario de recrutement en milieu universitaire*

<u>Évaluateurs novices</u> :

- Être diplômé de la dernière cohorte francophone du programme d'études supérieures en évaluation de programme.
- Être étudiant francophone au doctorat en psychologie et avoir suivi les cours :
  - PSYX01 : évaluation de programme, soit le cours dans le cadre duquel les étudiants ont élaboré un devis d'évaluation ; **<u>ET</u>** PSYX02 : recherche appliquée aux interventions sociales et communautaires, soit le cours avancé dans le cadre duquel les étudiants réalisent une évaluation de programme et produisent un rapport d'évaluation.
- Ou/Et être assistant de recherche francophone au centre de recherche X spécialisé en évaluation **<u>ET</u>** avoir suivi des cours en évaluation de programme.
- Avoir un an ou moins d'expérience en évaluation de programme (stage, réalisation d'une évaluation dans le cadre des cours du programme ou réalisation d'une à deux évaluations durant un assistanat au centre de recherche X spécialisé en évaluation).

<u>Évaluateurs experts</u> :

- Détenir un ou plusieurs diplômes en évaluation de programme.
- Avoir au moins dix ans d'expérience en évaluation de programme.
- Démontrer un certain degré de contribution au domaine ***(optionnel)***.
- Parler français



Le faible nombre de cas novices recrutés nous a conduits à ne pas poursuivre notre étude en milieu universitaire ainsi qu'à renoncer aux cas recrutés jusque-là. Nous avons dès lors entrepris des procédures de recrutement auprès d'un département d'évaluation de programme au sein d'un ministère du gouvernement fédéral du Canada, département qui offre une formation en évaluation de programme à leurs nouveaux employés.

3.3.1.2 Scénario de recrutement au sein du gouvernement fédéral : procédures et critères de sélection des cas

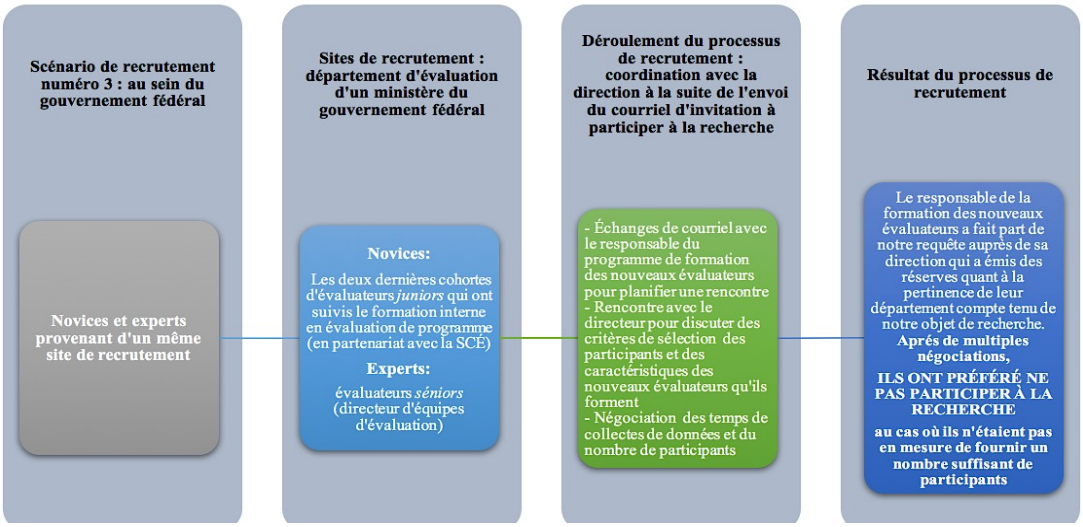
À la suite de notre participation à une conférence en évaluation, nous avons appris qu'un ministère du gouvernement fédéral du Canada offrait une formation en évaluation de programme à leurs nouveaux évaluateurs de programme (évaluateurs *juniors*) en collaboration avec la Société canadienne d'évaluation de programme (SCÉ). Aussi, sommes-nous entrés en contact avec le responsable de ce programme de formation au sein du ministère en question. Ce dernier a démontré un fort intérêt pour notre recherche et nous avons été en mesure d'amorcer le processus de recrutement. Comme ce fut le cas précédemment, au regard de nos premiers échanges le responsable du programme de formation interne, les critères de sélection des participants ont dû être amendés pour correspondre à cette nouvelle réalité de terrain (procédures du scénario de recrutement au gouvernement fédéral, Encadré 4, p. 111).

Après avoir consulté la haute direction du département d'évaluation, le responsable du programme de formation des évaluateurs *juniors* nous a fait part de leurs réserves quant aux types d'évaluation qu'ils réalisent. En effet, ces dernières sont majoritairement orientées sur la mesure de performance et la reddition de compte, ce qui pour eux constituait une limite pour notre recherche. Après de nouvelles négociations, ils nous ont également expliqué qu'ils ne seraient pas en mesure de nous recommander un nombre suffisant de participants et qu'ils préféraient ne pas s'engager dans la recherche. À la suite de leur désistement, notre personne-ressource nous a mis en contact avec un gestionnaire du département d'évaluation d'un autre ministère. Après de multiples échanges de courriel, la direction a finalement décidé que le climat organisationnel dans lequel se trouvait le département au moment du recrutement n'était pas propice pour la participation de leurs employés à notre recherche.

Encadré 4. *Procédures du scénario de recrutement au sein du gouvernement fédéral*

Évaluateurs juniors (*novices*) :
- Détenir un diplôme en évaluation de programme ou avoir suivi la formation interne en évaluation de programme.
- Avoir entre un et deux ans d'expérience en évaluation de programme.
- Parler français.

Évaluateurs séniors (*experts*) :
- Être détenteur d'un ou plusieurs diplômes en évaluation de programme.
- Avoir au moins dix ans d'expérience en évaluation de programme.
- Démontrer un certain degré de contribution au domaine ***(optionnel)***.
- Parler français



3.3.1.3 Scénario de recrutement au sein de cabinets de consultants : procédures et critères de sélection des cas

Par le biais d'un professeur qui donne des cours d'évaluation de programme dans une faculté d'éducation d'une université de l'Ontario, nous avons été mis en contact avec l'un des directeurs d'un cabinet de consultants en gestion et évaluation dans les secteurs publics fédéral et provincial (cabinet B). Dès notre première rencontre téléphonique, il a accepté de participer à notre recherche, à la condition d'attendre quelques mois pour commencer la recherche. Entre temps, ce dernier nous a référé un second cabinet d'expert-conseil qui exerce dans les mêmes secteurs (Cabinet A), et dont l'un des associés donne, tout comme nous, un cours d'études supérieures en évaluation de programme dans une université ontarienne. Ce point commun et la référence par un cabinet concurrent avec qui il collabore souvent a facilité notre entrée en contact et son désir de participer également à la recherche.

Cependant, à la suite d'échange de nombreux courriels et d'une rencontre téléphonique avec le cabinet A, nous avons dû revoir nos critères de sélection pour répondre à leur réalité de terrain, mais aussi pour être en mesure de recruter le plus de participants possible. En effet, nous en étions à notre cinquième concrète chance de recruter des participants (sans compter les tentatives qui ont échoué dès l'envoi du courriel d'invitation à participer à la recherche) et nous avons donc dû prendre un certain nombre de décisions afin de garantir la conduite de notre recherche doctorale. Aussi, les critères relatifs à la langue de communication des participants (ils travaillent majoritairement en anglais qu'ils soient anglophones ou francophones), au nombre d'années d'expérience des évaluateurs novices qui sont qualifiés de *consultants* (*juniors*) et à la formation en évaluation de programme ont dû être modifiés.

Au sein du cabinet A, l'évolution professionnelle des évaluateurs *consultants* (*juniors*) se fait sur environ trois ans avant qu'ils puissent candidater pour une position plus avancée. Trois ans au cours desquels ils font partie d'équipes d'évaluation sous la supervision d'un évaluateur sénior, effectuant différentes tâches relatives aux collectes de données d'évaluation de programme en cours, sans être impliqués dans l'ensemble dans la conception du devis ou la rédaction du rapport d'évaluation. Aussi, si le critère initial était d'avoir un an d'expérience (ou moins) et d'avoir réalisé une évaluation de programme (ou d'avoir participé à l'ensemble des étapes d'une évaluation), les évaluateurs *consultants* ne sont pas impliqués dans l'ensemble du processus évaluatif, ils participent juste à un ensemble de tâches méthodologiques. Après avoir discuté avec l'évaluateur associé, nous avons déterminé que les évaluateurs qui ont trois ans d'expérience et moins avaient des caractéristiques qui coïncidaient avec les critères relatifs aux évaluateurs novices, voire peut-être débutants avancés pour certains employés. De plus, puisque la deuxième collecte de données exigeait d'avoir réalisé (ou participé) à une évaluation de programme dont le rapport était accessible, cela nécessitait de sélectionner des participants qui ont entre une et trois années d'expérience. En effet, les cabinets de consultants prennent un maximum de précautions dans la communication de leurs évaluations, celles-ci doivent être publiées sur le site internet du ministère concerné pour qu'on puisse accéder aux rapports d'évaluation finaux. Nous verrons que cette contrainte de terrain a entrainé la disqualification de certains participants après la première collecte de données, car ils n'étaient plus en mesure de communiquer l'unique rapport d'évaluation qu'ils pouvaient nous fournir.

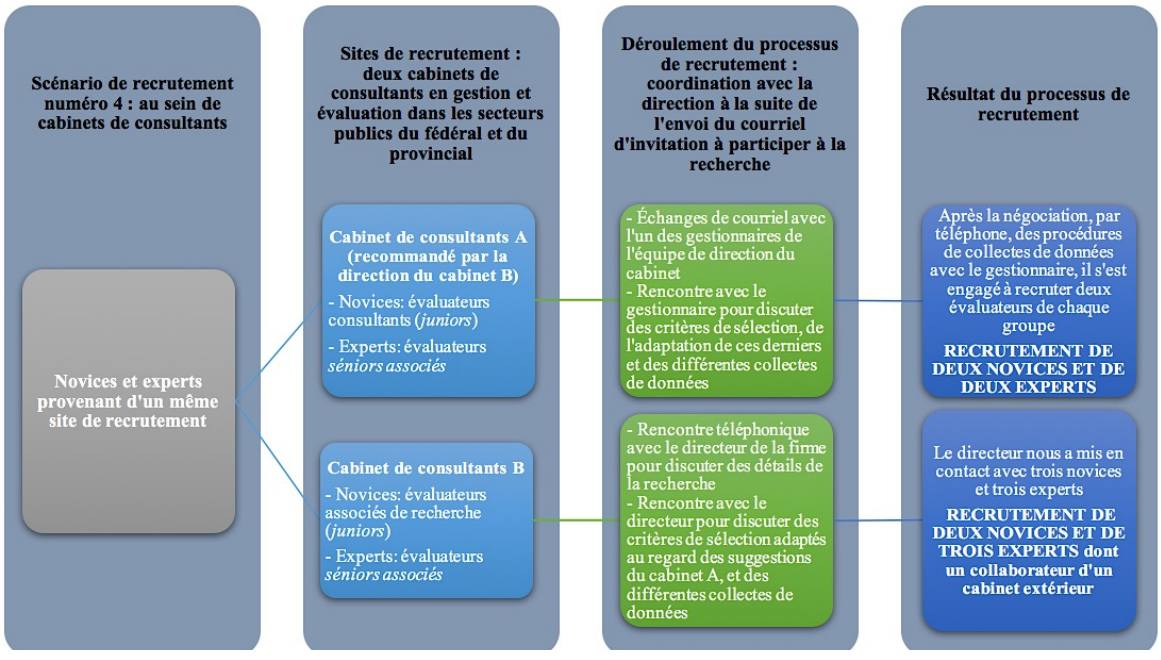
En ce qui a trait aux experts nous nous sommes concentrés sur les évaluateurs *séniors associés*. Ces derniers ne détiennent pas nécessairement de diplôme(s) en évaluation de programme, mais ils ont obtenu un ou plusieurs diplômes d'études supérieures avec une composante « recherche », en sciences humaines et sociales, en économie, en droit, etc. (maitrise ou doctorat). Ils sont associés du cabinet et ils sont accrédités par la société canadienne d'évaluation de programme. Enfin, certains d'entre eux donnent des cours ou des formations en évaluation de programme et participent régulièrement à des conférences organisées par différentes associations professionnelles en évaluation de programme.

Aussi, voici les critères de sélection des évaluateurs des deux cabinets de consultants (critères que nous avons « validés » avec le cadre conceptuel et auprès des gestionnaires des cabinets participants) ainsi que le processus de recrutement mis en œuvre (Encadré 5) :

Encadré 5. *Procédures du scénario de recrutement au sein de cabinets de consultants*

Évaluateurs *juniors* (novices) :
- Détenir un diplôme en évaluation de programme ou avoir suivi une formation interne en évaluation de programme.
- Avoir entre une et trois années d'expérience (ou moins) en évaluation de programme.
- Parler français et/ou anglais

Évaluateurs *séniors associés* (experts) :
- Détenir un ou plusieurs diplômes en évaluation de programme, en recherche en sciences sociales, etc.
- Avoir au moins dix ans d'expérience en évaluation de programme.
- Être un associé du cabinet
- Démontrer un certain degré de contribution au domaine ***(optionnel)***
- Parler français et/ou anglais



Nous avons ajouté le critère « être un associé du cabinet » afin de nous assurer de recruter les évaluateurs *séniors* qui détiennent le plus haut niveau d'expertise dans leur cabinet de consultation. De plus, si nous nous sommes assurés que les critères de classification des évaluateurs au sein des deux cabinets étaient similaires, il est possible que certains évaluateurs *juniors* aient commencé à participer à des tâches de gestion de projet. Aussi, en recrutant des évaluateurs *séniors associés* nous étions sûrs d'avoir le plus grand écart possible entre ces derniers et le groupe d'évaluateurs *juniors*.

3.3.2 Participants sélectionnés

À la suite de la période de recrutement, nous avons donc sélectionné quatre évaluateurs *juniors* (novices) et quatre évaluateurs *séniors associés* (experts) qui exercent dans des cabinets de cabinet de consultants en gestion et en évaluation dans les secteurs publics fédéral et provincial. Le Tableau 8 résume les caractéristiques des évaluateurs qui ont été sélectionnés pour participer à cette recherche, nous donnons plus de détails sur le profil des participants dans le chapitre qui portent sur les résultats :

Tableau 8. *Présentation des caractéristiques des participants de la recherche*

Évaluateurs *juniors* (novices)		**Évaluateurs *séniors associés* (experts)**	
'J1'	- Doctorat en psychologie expérimentale incluant une formation initiale en évaluation de programme - Deux ans d'expérience - Anglophone (unilingue)	**'S1'**	- Maitrise avec thèse en sciences politiques et en philosophie - Avocat en droit civil et criminel spécialisé en litige - Avocat en droit civil - Accréditation par la Société canadienne d'évaluation - 16 ans d'expérience en évaluation de programme - Associé du cabinet de consultant - Membre du conseil d'administration d'une association professionnelle en évaluation de programme - Anglophone (unilingue)

‘J2’	- Doctorat en psychologie expérimentale - Moins d’un an d’expérience en évaluation de programme - Associé de recherche au cabinet de consultants - Anglophone (unilingue)	**‘S2’**	- Doctorat en économie - Accréditation par la Société canadienne d’évaluation de programme - Enseigne des cours en évaluation de programme dans le cadre de programme d’études supérieures d’une université ontarienne - 16 ans d’expérience en évaluation de programme - Associé du cabinet de consultants - Francophone (bilingue français anglais)
‘J3’	- Maitrise en économie - Évaluateur *junior* pendant 4 ans avec quelques expériences de gestion de projets d’évaluation principalement en analyse documentaire et financière, avant de retourner à l’université pour faire un doctorat en économie - Aujourd’hui, il demeure évaluateur *junior* dans le cadre d’équipe d’évaluation sous la direction d’évaluateurs séniors (il participe à différentes tâches relatives à la collecte et à l’analyse de donnée), mais il ne gère pas de projet	**‘S3’**	- Maitrise avec thèse en sociologie - Accréditation par la Société canadienne d’évaluation de programme - 20 ans d’expérience en évaluation de programme - Associé du cabinet de consultants - Anglophone (unilingue)
‘J4’	- Maitrise avec thèse en affaires internationales - Diplôme d’études supérieures en évaluation de programme en administration publique - 2 ans d’expérience - Francophone (bilingue français-anglais)	**‘S4’**	- Maitrise avec thèse en sciences politiques - Avocat en droit civil - Accréditation par la Société canadienne d’évaluation - 17 ans d’expérience en évaluation de programme - Directeur du cabinet de consultants - Ancien président d’une association professionnelle en évaluation de programme - Francophone (bilingue français-anglais)

Enfin, avant de débuter la recherche, nous nous sommes assurés que les participants signent et nous renvoient les formulaires de consentement de participation à la recherche. De plus, au début de chaque entretien, nous avons pris le soin de réitérer les procédures mises en place pour garantir l’anonymat des participants, des cabinets de consultants et des programmes qu’ils ont évalués.

3.4 Déroulement de la recherche

Afin de mener à bien cette étude doctorale, nous avons réalisé deux collectes de données.

Dans le cadre de la première collecte de données (section 3.5.1), il s'agissait de documenter les bonnes pratiques des évaluateurs de programme, et de comprendre plus en profondeur leurs perceptions quant à l'influence potentielle de facteurs contextuels (tels que la relation avec les parties prenantes et les contraintes inhérentes au terrain) sur l'opérationnalisation de leurs choix méthodologiques. Pour ce faire, nous avons réalisé un entretien métaévaluatif dans le cadre duquel les participants devaient nous faire part de leur métaévaluation d'un rapport d'évaluation de programme (activité métaévaluative qu'ils avaient préparée avant la passation de l'entretien).

Dans le cadre de la deuxième collecte de données (section 3.5.2), nous avions pour objectif de documenter les pratiques réelles des évaluateurs de programme, et de comprendre le processus décisionnel relatif à l'élaboration et la mise en œuvre du protocole méthodologique dans le cadre de leurs opérations d'évaluation. Pour rendre compte de ces pratiques, nous avons eu recours à l'entretien d'explicitation. Aussi, selon les recommandations de Vermersch (2004, 1994, 1991), ce type d'entretien nécessite une collecte de données préliminaires qui porte sur l'analyse de l'action qui fait l'objet de l'explicitation. En outre, dans le cadre de cette recherche, il s'agit de l'explicitation d'une activité évaluative passée et dont le rapport fourni par les participants constitue la trace à analyser en vue de la passation de l'entretien. Nous avons donc conçu une grille d'analyse de rapport d'évaluation centrée sur le devis d'évaluation, que nous avons utilisée comme référentiel d'analyse de l'activité évaluative de nos participants afin de préparer l'entretien.

3.4.1 Première collecte de données : l'entretien métaévaluatif

Dans le cadre de cette première collecte de données, les participants devaient adopter une posture critique et réaliser la métaévaluation d'un rapport d'évaluation dans le cadre d'un entretien métaévaluatif. Afin d'être en mesure de réaliser l'entretien en question, la collecte de données s'est déroulée en plusieurs étapes, étapes qui feront l'objet des sections qui suivent. Dans un premier temps, nous avons préparé l'activité métaévaluative (choix du rapport d'évaluation, choix des consignes et des procédures de passation de l'activité). Dans un deuxième temps, nous avons créé un référentiel d'analyse d'activité (Tourmen, 2007), afin de préparer le guide d'entretien métaévaluatif (ou entretien de retour sur l'activité) ainsi que faciliter la prise de note durant

l'entretien. Dans un troisième temps, les participants ont préparé l'activité de métaévaluation d'un rapport d'évaluation et ils nous ont présenté leur critique dans le cadre d'un entretien métaévaluatif.

Les objectifs pour cette première collecte de données étaient les suivants :

1. Comprendre ce qui constitue, pour les participants, une bonne pratique évaluative.
2. Explorer avec eux le processus en amont d'une prise de décision en ce qui a trait à la méthodologie élaborée (idéale) et la méthodologie mise en œuvre.
3. Comprendre le rôle que peuvent jouer les contraintes inhérentes au terrain (telles que l'implication des parties prenantes ainsi que la manière dont ils appréhendent leur relation avec ces dernières) et l'influence que cela peut avoir sur le processus évaluatif.

3.4.1.1 Préparation de l'activité de métaévaluation

Avant tout, il est nécessaire de mentionner qu'une activité de métaévaluation[21] peut être très chronophage. Aussi Scriven (2005) suggère-t-il à titre de méthode alternative de se centrer sur la critique du devis d'évaluation employé, lequel inclut la méthodologie, suggestion que nous avons décidé de suivre. Ainsi, dans le cadre de la préparation de cette activité métaévaluative, nous avons dû prendre un certain nombre de décisions relatives :

- **À la sélection du rapport d'évaluation qui fait l'objet de l'activité** : initialement, nous devions utiliser un article scientifique publié en français, qui rapporte une évaluation de programme. Cependant, la sélection de participants francophones et anglophones nous a conduits à opter pour une autre solution qui nous permettait de recourir à une même évaluation pour l'ensemble des participants. Ainsi avons-nous décidé d'utiliser un rapport d'évaluation de programme assez court, publié en anglais et en français. De surcroit, nous avons également

[21] La métaévaluation est une pratique d'évaluation systématique d'une évaluation de programme afin de déterminer la qualité du processus et des résultats évaluatifs (JCSEE, 2011 ; Greene, 1992) : elle est formative lorsqu'elle sert à améliorer le processus évaluatif au fil de l'évaluation et elle est sommative lorsqu'elle fournit aux clients, aux publics concernés par le programme ainsi qu'à l'évaluateur des informations en ce qui a trait aux forces et aux faiblesses de l'évaluation (Scriven, 2005, 2001 ; Stufflebeam, 2001).

dû prendre en considération le domaine de pratique de nos évaluateurs participants, à savoir l'évaluation de programmes publics au niveau du gouvernement fédéral canadien. Aux regards de ces critères, nous avons consulté les banques de rapports d'évaluation accessibles sur les sites internet de différents ministères. Nous avons fait part d'une liste de rapports à un expert en évaluation de programme pour le gouvernement fédéral afin de choisir un rapport d'évaluation avec suffisamment de matières à discuter et pertinent pour notre échantillon. Une fois le rapport sélectionné[22], nous avons effectué une validation du contenu de la version française de ce dernier auprès d'un spécialiste en traduction et en didactique du français.

- **À la négociation des procédures de réalisation de l'activité :** initialement, les participants devaient préparer l'activité métaévaluative en début d'entretien. Nous devions leur accorder 60 minutes pour préparer une analyse critique du devis d'évaluation présenté dans le rapport d'évaluation distribué, afin qu'ils participent par la suite à l'entretien métaévaluatif. L'activité devait durer entre 120 et 190 minutes. Cette exigence en termes de temps a posé des difficultés puisque les évaluateurs devaient participer à cette activité dans le cadre de leurs heures de travail. En effet, les gestionnaires et les évaluateurs participants ne semblaient pas en mesure de libérer un temps de travail aussi conséquent. En outre, nous avons donc dû négocier les conditions de réalisation de l'activité et opté pour la réalisation de celle-ci en deux temps : (1) nous leur avons envoyé, la veille de l'entretien, le rapport d'évaluation qui fait l'objet de la métaévaluation ainsi que les consignes de préparation de l'activité ; et (2) le jour de l'entretien les participants nous ont présenté l'analyse qu'ils avaient préparée et nous avons discuté du contenu du rapport et de leurs commentaires (et annotations).
- **Et au choix des consignes :** comme nous le verrons dans la partie qui porte sur la deuxième collecte de données, l'exercice métaévaluatif se réalise généralement à l'aide d'indicateurs qui proviennent de différentes sources telles que les associations professionnelles en évaluation de programme et le *Joint Committee on Standards for Educational Evaluation* (JCSEE).

[22] Rapport d'évaluation d'un programme de prévention du ministère des Affaires autochtones et du Nord Canada, publié en 2014, en anglais (58 pages) et en français (67 pages).

Cependant, afin de recueillir les bonnes pratiques des évaluateurs novices et experts, nous avons décidé de ne pas leur donner de liste d'indicateurs. De plus, puisque nous avons convenu que les participants préparent l'activité métaévaluative la veille de l'entretien, nous nous sommes également assurés de ne pas mentionner les termes métaévaluation, activité métaévaluative ou encore analyse critique (que ce soit dans le formulaire de consentement ou dans les consignes de l'activité). En effet, nous voulions éviter l'usage de tels concepts clés afin de limiter les chances que les participants aillent se référer aux écrits scientifiques. Enfin, nous avons donc envoyé la consigne suivante :

> *« Dans le cadre de l'entretien de recherche de demain, vous allez lire le rapport d'évaluation procuré en pièce jointe et discuter du contenu de ce dernier avec la chercheuse (moi-même). Aussi, je vous invite à prendre connaissance du rapport dès aujourd'hui, particulièrement le devis d'évaluation (incluant la méthodologie) : je vous suggère d'annoter les forces et les faiblesses de l'évaluation ainsi que vos suggestions pour l'amélioration de cette dernière. Vous bénéficierez d'un temps de préparation additionnel demain, au début de l'entrevue (au besoin). L'entretien devrait durer entre 60 et 90 minutes ».*

3.4.1.2 Instrumentation : référentiel d'analyse d'activité et guide d'entretien métaévaluatif

Le recours à l'entretien comme méthode de collecte de données répond à trois arguments proposés par Poupart (1997). Premièrement, l'argument épistémologique selon lequel l'entretien permet d'explorer en profondeur les perceptions et les représentations des sujets de la recherche au regard de l'objet d'étude. Deuxièmement, l'argument éthique et politique qui stipule que l'entretien ouvre à la compréhension et au savoir de l'intérieur d'un phénomène et des enjeux que celui-ci implique pour les sujets de la recherche. Enfin, l'argument méthodologique qui met de l'avant que l'entretien donne un accès privilégié à l'expérience des sujets de l'étude.

Ainsi, en favorisant une interaction directe entre le chercheur et le sujet de sa recherche, l'entretien constitue la méthode privilégiée afin de saisir le point de vue des acteurs du phénomène à l'étude, et de recueillir les significations et les interprétations qu'ils attribuent à leurs propres expériences (Patton, 2015b ; Baribeau et Royer, 2012). Qu'il soit plus ou moins directif, l'entretien permet au chercheur de colliger les données recueillies qui portent sur le cadre personnel de référence des sujets (jugements, perceptions, significations, etc.) vis-à-vis du sens qu'ils attribuent à la situation à l'étude. Il peut ensuite objectiver les significations et les interprétations des acteurs

et faire émerger des connaissances scientifiques qui permettent de comprendre et de décrire le phénomène étudié (Van der Maren, 2010).

Enfin, lorsque l'entretien est semi-directif (comme c'est le cas pour l'entretien métaévaluatif), l'entretien est évolutif puisque l'intervieweur ne peut pas prévoir de manière certaine la totalité du contenu de ce dernier, même s'il définit les thèmes et les sous-thèmes qu'il désire aborder dans le cadre de l'entretien (Poupart, 2012). Aussi, l'entretien débute de manière assez ouverte avec une consigne ou question de départ plutôt large (suivis de sous-consignes) afin de faire parler l'interviewé tout en évitant de trop préstructurer le champ de réponse. Il est essentiel de permettre au sujet interrogé de traiter des thèmes et des sous-thèmes de l'étude selon son propre schéma de réponse (ses thèmes et ses catégories), pour ne pas lui imposer une manière de penser. De surcroit, l'intervieweur doit faire preuve d'une grande capacité d'écoute afin de formuler des relances pertinentes et d'étudier plus en profondeur les dimensions de l'expérience de l'interviewé, tout en lui permettant de faire preuve d'initiative dans sa propre auto-exploration.

Compte tenu des objectifs de cette première collecte de données, la conduite des entretiens métaévaluatifs dans le cadre desquels chaque participant réalisait à l'oral sa métaévaluation fut dans un premier temps non directive, puis dans un deuxième temps davantage semi-directive selon les besoins. En effet, il était essentiel qu'ils aient l'occasion d'étayer davantage leur analyse critique et d'aller plus loin que les éléments qu'ils avaient relevés dans le cadre de la préparation de l'activité. Afin de faciliter la conception du guide d'entretien métaévaluatif et sa schématisation, nous avons élaboré un référentiel d'analyse d'activité à partir des objectifs de la collecte de données ainsi que des concepts clés du cadre conceptuel (Annexe 7). Cet outil a également été conçu pour faciliter la prise de note durant l'entretien, puisque le participant nous faisait d'abord part de sa métaévaluation avant que nous ne revenions ensemble en détail sur les éléments clés de celle-ci. Par ailleurs, l'utilisation du guide d'entretien (Annexe 7), mais aussi du référentiel, fut souple puisqu'il s'agissait de privilégier la parole du participant et qu'il était essentiel de formuler des relances uniquement lorsque c'était nécessaire (afin d'être en mesure de répondre aux objectifs présentés plus tôt).

Enfin, nous avons réalisé une préexpérimentation de l'entretien métaévaluatif (préparation de l'activité métaévaluative et entretien de retour sur activité) auprès d'un doctorant en psychologie qui a suivi l'ensemble des cours offerts en évaluation de programme. Il a également effectué plusieurs évaluations de programmes publics du gouvernement fédéral dans le cadre d'un

internat dans un cabinet de consultants en recherche et évaluation. Ainsi, la préexpérimentation des procédures de passation de l'entretien métaévaluatif nous a permis de :

- Vérifier la clarté des consignes de préparation de l'activité métaévaluative.
- Confirmer notre choix de rapport d'évaluation.
- Nous familiariser avec l'utilisation du référentiel d'analyse de l'activité dans le cadre de la conduite de l'entretien et de la prise de note.
- Pratiquer l'utilisation du guide d'entretien et apporter les ajustements nécessaires au schéma d'entretien métaévaluatif afin d'assurer la fluidité de la passation et d'accéder plus facilement à ce qui constitue, pour l'interviewé, de bonnes pratiques méthodologiques en évaluation.

3.4.1.3 Procédures de passation de l'entretien métaévaluatif

Au regard des conditions de passation de l'entretien métaévaluatif convenues avec les gestionnaires des cabinets de consultants où exercent nos participants (mentionné plus tôt), cette collecte de données s'est déroulée en plusieurs étapes. D'abord, nous avons convenu d'une date de rencontre avec les participants pour réaliser l'entretien évaluatif, date qui devait coïncider avec la possibilité pour ces derniers de réaliser dans le cadre de leurs horaires de travail, l'activité métaévaluative ainsi que l'entretien métaévaluatif. Ensuite, 24 h avant l'entretien, nous leur avons envoyé par courriel les consignes de l'activité métaévaluative ainsi que le rapport d'évaluation dans la langue de leur choix. À partir de là, les participants avaient la responsabilité de préparer leur métaévaluation avant l'entretien. Enfin, nous nous rendions le lendemain dans les locaux du cabinet de consultants des participants afin de réaliser l'entretien métaévaluatif. Pour les évaluateurs qui travaillent dans des bureaux en dehors de la ville, nous avons réalisé les entretiens en vidéoconférence. Avant de commencer chaque entretien, nous nous sommes assurés que les participants avaient bel et bien eu le temps de préparer l'activité de métaévaluation afin d'être en mesure de leur accorder un temps de préparation supplémentaire avant la passation de l'entrevue (si nécessaire). Comme on le mentionnait sur le formulaire de consentement que les évaluateurs ont signé avant de participer à la recherche, les entretiens ont été enregistrés avec un dictaphone. Nous nous sommes tout de même assurés que ces derniers consentaient toujours à l'enregistrement audio de la passation de cet entretien métaévaluatif. Nous avons également pris le temps de bien expliquer l'objectif général et le déroulement de l'entretien, et de vérifier le choix de la langue de

discussion (soit le français ou l'anglais). Dans l'ensemble, les participants se sont tous présentés préparés pour la passation de l'entretien et les entretiens ont duré en moyenne entre 60 et 90 minutes.

3.4.1.4 Procédures d'analyse des données de l'entretien métaévaluatif

Dans le cadre de l'analyse de contenu des données qualitatives obtenues dans l'entretien métaévaluatif, nous nous sommes inspirés de la méthodologie préconisée par Miles, Huberman et Saldaña (2014) et Saldaña (2013), et nous avons eu recours au logiciel d'analyse et de traitement de données qualitatives Atlas.ti. Nous avons suivi les recommandations des auteurs selon lesquelles le processus de codage et d'analyse des données débute par un premier cycle de codage du contenu des données textuelles en vue de préparer le corpus de texte pour l'activité analytique mise en œuvre durant le deuxième cycle de codage où il s'agit de classer, catégoriser, prioriser, synthétiser, conceptualiser et modéliser les résultats de la recherche (Miles, Huberman et Saldaña, 2014). Avant de débuter le 1er cycle de codage, nous avons exploré le contenu des données afin de déterminer la ou les méthodes de codages initiales (1er cycle) les plus appropriées (Saldaña, 2013). À partir du référentiel d'analyse de l'activité d'analyse critique et à l'aide d'une méthode de codage de type exploratoire, nous avons mis en exergue les éléments abordés par l'interviewé dans le cadre de l'analyse critique du devis d'évaluation.

À la suite de cette phase exploratoire, nous avons réalisé une première activité de validation du processus de codage et de l'arbre de codes à l'aide d'un expert, ce qui a montré que la structure des séries de codes proposées se situait à un niveau beaucoup trop « micro » pour être mesure d'analyser efficacement le corpus de données en vue de répondre aux questions de recherche. En outre, ceci nous a conduits à remettre en question la pertinence de la dichotomie établie entre le premier cycle et le deuxième cycle de codage compte tenu des considérations paradigmatiques, conceptuelles et méthodologiques de notre étude doctorale. En effet, dans notre cas, la combinaison et l'assortiment de différents types de codes de 1er cycle renseignent directement les questions de recherche (méthode de 2e cycle – codage théorique). Le codage du jugement de l'interviewé en ce qui a trait à l'objet méthodologique critiqué permet de mettre en lumière ses pratiques évaluatives modèles (mais aussi réels) et les facteurs contextuels qui peuvent avoir une influence plus ou moins importante sur ces dernières. De plus, le processus de codage permet d'établir des éléments de comparaison "intra" et "inter" cas de novices et cas d'experts. Nous

avons donc opté pour la méthode de codage éclectique, une méthode de codage hybride (1er-2ème cycles) qui met en œuvre un processus de combinaison et d'utilisation simultanée de plusieurs méthodes de codage de 1er cycle ciblées, qui propose une première tentative de catégorisation analytique des codes suivis d'un cycle itératif de révision et d'analyse des données codées en vue de répondre aux questions de recherche (Saldaña, 2013). Nous avons notamment eu recours aux méthodes de codage structurel (codes primaires à partir du contenu des questions d'entrevues), de codage secondaire (les éléments descriptifs du code primaire), de codage simultané (description et inférence au sein d'une même unité de texte), et de codage évaluatif (jugement sur la valeur, le mérite, la portée et la qualité des composantes de l'objet qui fait l'objet d'une évaluation ou d'une métaévaluation), etc. (Saldaña, 2013). L'utilisation combinée et assortie de ces méthodes s'est faite de manière juxtaposée avec le codage théorique (ou conceptuel), une méthode de deuxième cycle.

Enfin, la Figure 14 illustre le processus d'analyse des données collectées dans le cadre de la première activité de recherche. Et précisons qu'un rapport d'analyse et de présentation de résultats sous forme de méta-matrices partiellement ordonnées (Mile, Huberman et Saldaña, 2014) a été produit pour chaque participant.

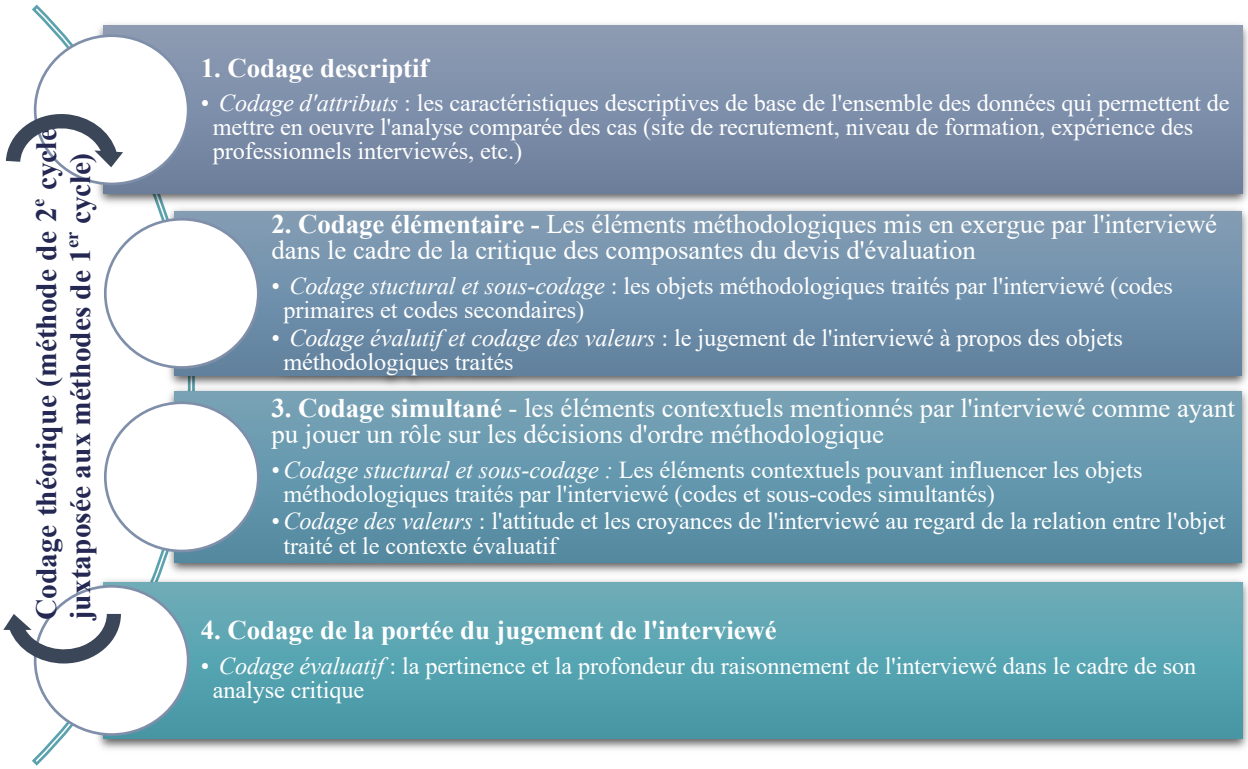


Figure 14. *Processus de codage éclectique des données qualitatives de l'activité métaévaluative*

3.4.2 Deuxième collecte de données : l'entretien d'explicitation

Dans le cadre de la première collecte de données, les évaluateurs participants ont été placés dans une posture critique, où ils étaient encouragés à nous faire part de leurs bonnes pratiques évaluatives. À travers cette deuxième collecte de données, on les a amenés à adopter une posture « méta » (autoréflexive, explicative) puisque nous désirions recueillir leurs pratiques évaluatives réelles ainsi que les significations et les interprétations qu'ils attribuent à ces dernières. En effet, l'objectif de cette collecte était de décrire et de comprendre l'opérationnalisation des choix méthodologiques (de la conception à la réalisation) des évaluateurs lors de leurs opérations d'évaluation et leurs perceptions quant à l'influence de facteurs contextuels sur ces choix. Pour répondre à ces objectifs, nous avons choisi de recourir à l'entretien d'explicitation (Vermersch, 1994, 1991). Cette technique d'entretien vise à obtenir la description la plus fine possible d'une activité passée, réalisée par un individu dans le cadre de sa pratique professionnelle. En outre, le but est de s'informer, auprès de l'acteur de la situation à l'étude (ou activité), sur ce qui s'est réellement passé et sur les connaissances implicites inscrites dans la réalisation des actions propres à la situation. Enfin, au regard des règles prescrites par Vermersch (1994, 1991) pour la conduite rigoureuse d'un entretien d'explicitation, cette deuxième collecte de données s'est déroulée en plusieurs étapes :

(1) la conception d'un référentiel d'analyse de l'activité évaluative qui fait l'objet de l'entretien d'explicitation, soit une grille d'analyse de rapports d'évaluation centrée sur le devis d'évaluation ;
(2) la validation de contenu de la grille d'analyse ;
(3) l'expérimentation de la grille à travers l'analyse du contenu de rapports d'évaluation produits par les participants ;
(4) la conception du guide d'entretien d'explicitation à partir de l'analyse des rapports d'évaluation ;
(5) la passation et l'analyse des entretiens d'explicitations.

3.4.2.1 Le référentiel d'analyse de l'activité évaluative : la grille d'analyse de rapports d'évaluation

Dans le cadre de cette deuxième collecte de donnée, il était question de recueillir les pratiques réelles des évaluateurs de programme, à partir d'un entretien d'explicitation dont l'objet

est une de leurs évaluations rapportées dans un rapport qu'ils nous ont communiqué. Aussi, l'objectif était de pouvoir aborder plus en profondeur les explications des évaluateurs quant aux choix méthodologiques opérés (ce qu'il s'est passé lors du déroulement de l'évaluation) et leur perception quant à l'influence potentielle des parties prenantes sur leur processus décisionnel opéré. Selon Vermersch (1994, 1991), afin de réaliser un bon entretien d'explicitation, le chercheur doit se familiariser avec la tâche produite par le participant, tâche qui fait l'objet de l'entretien. Pour ce faire, il recommande le recours à un référentiel d'analyse pour être en mesure de préparer l'entretien d'explicitation. Dans le cadre de cette deuxième collecte de données, l'objet en question est une évaluation de programme réalisée par le participant ou une évaluation de programme à laquelle il a contribué. À titre de référentiel d'analyse de cette activité évaluative, nous avons conçu une grille d'analyse de rapport d'évaluation centrée sur le devis d'évaluation. Nous allons voir par la suite les procédures de conception, de validation de contenu et d'expérimentation de cette grille.

3.4.2.1.1 Procédure de création de la grille d'analyse des rapports d'évaluation

La métaévaluation ou l'analyse critique d'un devis d'évaluation se base généralement sur les normes du *Joint Committee on Standards for Educational Evaluation* ou JCSEE (Scriven, 2005). Ainsi, la grille d'analyse de rapports d'évaluation a été conçue à partir de la traduction et de l'adaptation de la *liste de vérification de la précision des devis d'évaluation* de Sanders et Nafziger (2011), qui mobilise, entre autres, les critères de la norme « précision » approuvés par le JCSEE. Les auteurs ont également proposé l'ajout des critères « représentativité », « objectivité » et « fiabilité » en ce qui a trait à la précision concernant la collecte et l'analyse des données. Nous avons, par ailleurs, modifié la grille de vérification de Sanders et Nafziger (2011) au regard des normes SEVAL (standards de la Société suisse d'évaluation, 2000). En effet, ces derniers ont proposé de référencer les normes (et leurs critères respectifs) auxquelles se référer en fonction des activités principales d'évaluation (par exemple, la planification de l'évaluation, l'exploitation des informations ou encore la présentation du rapport d'évaluation). À la suite de l'analyse des différentes listes de normes de références SEVAL, nous avons sélectionné les critères relatifs aux différentes normes qui peuvent être mobilisées dans le cadre de l'analyse d'un rapport, centrée sur la méthodologie, l'analyse des résultats et les limites. Au regard des différents écrits consultés dans

le cadre de l'élaboration de la *grille d'analyse de rapports d'évaluation*, nous avons retenu les critères suivants (Tableau 9).

Tableau 9. *Normes et critères retenus pour la construction de la grille d'analyse de rapports d'évaluation*

La précision (JCSEE [2011] et SEVAL [2010])	
P3. Description des objectifs et de la démarche	« Les objectifs, les questions et la démarche de l'évaluation sont documentés avec suffisamment de précision pour qu'on puisse les identifier et les apprécier ».
P4. Sources d'informations	« Les sources d'information utilisées dans une évaluation sont décrites avec suffisamment de précision pour qu'on puisse apprécier la justesse des informations ».
P5. Informations valides et fiables	« La procédure de récolte d'informations est choisie ou développée puis appliquée de manière à garantir la validité et la fiabilité des interprétations dans la perspective de l'objectif donnée ».
P7. Analyse des informations qualitatives et quantitatives	« Les informations qualitatives et quantitatives d'une évaluation sont analysées de manière appropriée et systématique, afin de répondre effectivement aux questions de l'évaluation ».
La Précision concernant la collecte et l'analyse des données (traduit librement de Sanders et Nafziger, 2011)	
Fiabilité *(Replicability)*	Le devis d'évaluation (incluant la méthodologie) présente des procédures qui permettent de s'assurer que les informations collectées sont exactes. Ainsi, si l'évaluation est répliquée, les mêmes résultats seront obtenus.
Objectivité *(Objectivity)*	Le devis d'évaluation (incluant la méthodologie) contient des procédures qui permettent de contrôler les biais susceptibles d'affecter les informations collectées ainsi que les inférences émises à partir de l'analyse de ces dernières.
Représentativité *(Representativeness)*	Les informations collectées représentent avec précision-justesse le programme évalué. Le devis d'évaluation contient des procédures de contrôle de la validité des inférences faites à partir des données obtenues à l'aide des instruments de collecte de données.
L'utilité (JCSEE [2011] et SEVAL [2010])	
U5. Transparence de l'appréciation	« Les perspectives et raisonnements sur lesquels se fonde l'interprétation des résultats sont décrits de telle sorte que les bases des jugements de valeur apparaissent clairement ».
U6. Caractère complet et clair du rapport	« Les rapports d'évaluation décrivent clairement l'objet d'évaluation et son contexte, de même que les objectifs, questions, procédures et résultats d'évaluation, dans le souci de mettre à disposition les informations essentielles sous une forme facilement compréhensible ».

Les axes de la *grille d'analyse de rapports d'évaluation* que nous avons développée portent sur la précision concernant le devis d'évaluation :

- la description générale de la démarche d'évaluation (cahier des charges),
- la description des sources d'information et les procédures d'échantillonnage,
- la description des procédures de collectes de données
- la description des procédures d'analyse des données
- la description des limites méthodologiques

3.4.2.1.2 Procédure de validation du contenu de la grille d'analyse de rapports d'évaluation

Afin de nous assurer de la validité du contenu de la grille d'analyse de rapports d'évaluation, nous avons recruté des experts en mesure et évaluation, en évaluation de programme au fédéral et en évaluation de programme en milieu éducationnel et communautaire ; soit trois domaines d'expertise en évaluation de programme. Nous avons envoyé un courriel de recrutement à six experts (au total) et trois ont accepté de réaliser la validation du contenu de la grille d'analyse de rapports d'évaluation. Le premier expert est professeur agrégé spécialisé en mesure et évaluation, et le second expert travaille dans un département d'évaluation de programme au gouvernement fédéral et participe à la formation des nouveaux employés en partenariat avec la Société canadienne d'évaluation. Enfin, le troisième expert est professeur émérite spécialisé en recherche et en évaluation de programmes éducationnels et communautaires. Au regard de l'analyse de leurs commentaires concernant les items de la grille et de la complémentarité de ces derniers, nous n'avons pas invité davantage d'experts à participer à la validation du contenu de la grille.

Afin de faciliter le processus de validation de contenu ainsi que l'analyse des évaluations des experts, nous avons converti la grille d'analyse de rapports d'évaluation en formulaire Google. En guise d'introduction du formulaire, nous avons décrit l'instrument et les consignes relatives au processus de validation de la grille (Annexe 8). Nous avons demandé aux experts d'utiliser la case commentaire à la suite de chacun des items pour nous faire part de leurs critiques et de leurs suggestions pour améliorer les items qu'ils jugeaient nécessaire de modifier. Une fois le processus de validation du contenu de la grille terminé, nous avons analysé les évaluations (les critiques et les suggestions) du contenu des items de la grille des trois experts. La Figure 15 présente deux

extraits du tableau de validation du contenu de l'instrument (la version complète du tableau qui sur le processus décisionnel pour chaque item selon commentaires des experts est en Annexe 8).

	Commentaires concernant l'item 1.2.3 Le devis d'évaluation est fidèle au type d'évaluation choisi	**Décision concernant l'item**
	É.1 : le terme devis porte à confusion ici. Parle-t-on de la méthodologie?	Le terme devis d'évaluation fait référence au terme anglophone « *evaluation design* » qu'on définit comme l'empreinte méthodologique de l'évaluation. Aussi, au regard des commentaires il semble plus approprié de parler de plan d'évaluation celui-ci faisant explicitement référence à la méthodologie évaluative. Aussi, voici **notre nouvelle proposition concernant cet item**: **1.2.3 Le plan d'évaluation est fidèle au type d'évaluation choisi**
	É.2 : Cela dépend de ce que l'on entend par le terme "devis d'évaluation". Chez nous, le devis et un document appelé "workplan" élaboré spécifiquement pour l'évaluation d'un programme; chaque évaluation possède un workplan qui est le premier document à produire dans la conduite d'un processus d'évaluation. Comme nous avons très peu de flexibilité pour choisir le type d'évaluation, le devis devra nécessairement refléter cette réalité. Il serait dangereux de considérer la portion du rapport portant sur la méthodologie comme étant le devis d'évaluation, si cette hypothèse représente une option.	
	É.3 : Bien.	
1.2 Présentation des questions d'évaluation	**Commentaires concernant l'item 1.3.1 Les questions d'évaluation correspondent adéquatement aux objectifs d'évaluation**	**Décision concernant l'item**
	É.1 : N/A	Dans la continuité des changements apportés à l'item 1.2.1, voici **notre nouvelle proposition concernant cet item**: **1.3.1 Les questions d'évaluation sont pertinentes au regard des objectifs d'évaluation**
	É.2 : Il s'agit d'un critère qui fait appel au jugement de ceux qui utiliseront la grille, donc qui risque de varier pour plusieurs individus évaluant un même rapport.	
	É.3 : La précision et clarté des objectifs et du mandat risquent d'influencer la précision des questions d'évaluation. Donc, le sens exact du mot "'adéquatement" pourrait varier avec les circonstances de l'évaluation.	
	Commentaires concernant l'item 1.3.2 Les questions d'évaluation sont claires	**Décision concernant l'item**
	É.1 : N/A	Afin d'assurer la rigueur scientifique d'une évaluation, de la définition de ses objectifs à la rédaction de ses résultats, les questions d'évaluation doivent être claires, et ce, quelques soient les circonstances. Aussi, **nous conservons l'item tel quel**: **1.3.2 Les questions d'évaluation sont claires**

Figure 15. *Extrait du tableau de validation de contenu de la grille d'analyse de rapports d'évaluation*

3.4.2.1.3 Expérimentation de la grille d'analyse des rapports d'évaluation

Une fois les rapports d'évaluation collectés auprès des participants, nous avons entrepris leur analyse à partir de la grille d'analyse de rapports d'évaluation, en veillant à relever des informations clés quant au devis d'évaluation qui appuie la formulation des conclusions et des recommandations. La version finale de la grille d'analyse des rapports d'évaluation est disponible en Annexe 9.

Aussi, pour chacun des énoncés de la grille, nous avons procédé comme suit :

- Si un élément ne s'appliquait pas, nous avons coché « N/A ».
- Si un élément de la liste était absent du rapport, sans explication, alors qu'il aurait été adéquat et nécessaire de l'aborder, nous avons coché « non ».
- Si un élément de la liste était partiellement présent dans le rapport, alors qu'il aurait été adéquat et nécessaire de l'aborder plus en profondeur, nous avons coché « Part. ».
- Pour les autres éléments qui s'appliquaient et qui étaient dans le rapport d'évaluation, nous avons coché « oui ».

Pour faciliter la conception du guide d'entretien d'explicitation, nous avons systématiquement utilisé la case commentaire pour relever, analyser les éléments présents, ou non. Nous avons également fait des annotations directement sur le rapport d'évaluation, ce qui a facilité le déroulement de l'entretien et le processus d'explicitation de l'activité évaluative discutée.

Encadré 6. *Extrait de l'utilisation de la grille d'analyse du rapport du participant [COH7]*

CODE du participant [COH7]
Type d'évaluation : Évaluation sommative de la mise en œuvre d'un programme X – cohérence, mise en œuvre et résultats

3

Thèmes	Critères principaux	Critères détaillés	Respect des critères				Commentaires
			N/A	Non	Part.	Oui	
1. Précision concernant la description de la démarche d'évaluation : *[Les objectifs, les questions et la démarche de l'évaluation sont documentés avec suffisamment de précision pour qu'on puisse les identifier et les apprécier. Soit la description détaillée de :* *- l'organisation,* *- la collecte,* *- la mise au point, les réajustements,* *- la saisie et l'exploitation des données.]*	1.1 Description des objectifs de l'évaluation	1.1.1 Le mandat d'évaluation est clairement décrit	N/A	Non	Part.	Oui	
		1.1.2 Les objectifs d'évaluation sont clairs	N/A	Non	Part.	Oui	
		1.1.3 Les objectifs d'évaluation sont précis	N/A	Non	Part.	Oui	
	1.2. Justification du choix du devis d'évaluation	1.2.1 Le type d'évaluation choisi est pertinent pour répondre aux objectifs d'évaluation	N/A	Non	Part.	Oui	Le type d'évaluation choisi est pertinent pour répondre aux objectifs d'évaluation cependant nous nous interrogeons sur la décision d'inscrire une évaluation sommative dans une approche participative
		1.2.2 Le choix du type d'évaluation est justifié par l'auteur	N/A	Non	Part.	Oui	
		1.2.3 Le plan d'évaluation est fidèle au type d'évaluation choisi	N/A	Non	Part.	Oui	Comme nous l'avons mentionné pour l'item 1.2.1 nous remettons en question l'approche participative dans le cadre de la mise en œuvre d'une évaluation sommative
	1.3. Présentation des questions d'évaluation	1.3.1 Les questions d'évaluation sont pertinentes au regard des objectifs d'évaluation	N/A	Non	Part.	Oui	Critique : les questions d'évaluation sont seulement mentionnées en annexe dans le tableau des matrices. - Dans la matrice d'évaluation les sources d'information ne sont pas spécifiées = remise en question de la faisabilité/évaluabilité
		1.3.2 Les questions d'évaluation sont claires	N/A	Non	Part.	Oui	- Manque de catégorisation des questions - Plusieurs indicateurs dans une même question - Plusieurs questions dans une même question - L'auteur du rapport mentionne l'existence de modèle logique, mais ils ne sont pas dans le rapport d'évaluation
		1.3.3 Les questions d'évaluation sont précises	N/A	Non	Part.	Oui	

L'Encadré 6 fournit un extrait de l'utilisation de la grille d'analyse, et extrait est illustré plus tard dans le contenu du guide d'entretien présenté dans la section suivante (Encadré 7, p.133).

3.4.2.2 Le guide d'entretien d'explicitation

Comme nous l'avons mentionné à plusieurs reprises, l'objectif de l'entretien d'explicitation était d'amener les évaluateurs à nous décrire ce qui s'est passé durant le projet d'évaluation (soit la tâche analysée plus tôt), en discutant des éléments qui se trouvent (ou non) dans leur rapport d'évaluation. Ainsi s'agissait-il de faire ressortir l'historique des stratégies mises en place durant l'évaluation du programme, de même que : (1) le rôle des parties prenantes dans le processus décisionnel relativement à la méthodologie d'évaluation implantée ; et (2) l'influence des contraintes inhérentes au terrain.

3.4.2.2.1 Procédure de création des guides d'entretien d'explicitation à partir de l'expérimentation des grilles d'analyse des rapports d'évaluation

L'entretien d'explicitation est un essai de conversation, une technique de verbalisation qu'on utilise, a posteriori, une fois que l'action (ici l'activité évaluative) qui fait l'objet de l'explicitation a été réalisée (Vermersch, 1991). En effet, comme nous l'avons mentionné plus tôt, l'entretien d'explicitation s'intéresse au vécu de l'action du sujet interviewé et il s'attarde plus particulièrement sur le déroulement procédural de l'action, c'est-à-dire son exécution (Martinez, 1997). Aussi, durant l'entretien, le chercheur doit guider le participant dans la verbalisation du déroulement de l'activité passée, soit la suite de ses actions procédurales et l'émergence de liens qui existent entre l'action et la cognition. Durant l'entretien, l'intervieweur doit donc s'assurer de demeurer dans le questionnement afin de placer l'interviewé en situation d'évocation et de description. Il est essentiel d'éviter d'être dans l'interrogation, car cela stopperait l'évocation (et le recueil d'informations sur le déroulement de l'action) et conduirait l'interviewé à se justifier. Pour ce faire, contrairement à l'entretien métaévaluatif, cet entretien était directif dans la structure des questions et du schéma d'entretien, et dans la façon de canaliser l'interviewé pour s'assurer qu'il colle très finement à l'évocation de son expérience (Martinez, 1997). L'un des points sensibles dans la préparation et la conduite d'un tel entretien est donc de maintenir l'interviewé dans l'évocation du vécu de l'action, sans induire le contenu des réponses. Ainsi l'intervieweur doit-il, dans ses questions et ses relances, laisser suffisamment d'espace pour que l'interviewé puisse évoquer, décrire avec ses propres mots ce que fut réellement son expérience (Martinez, 1997 ; Vermersch, 1994, 1991).

Pour chaque participant, le guide d'entretien d'explicitation a été conçu à partir de l'analyse de leur rapport d'évaluation centrée sur le devis d'évaluation, incluant les procédures méthodologiques et les limites. Ce rapport constituait une trace privilégiée de l'activité évaluative à l'étude et son analyse fut une étape essentielle dans la préparation de la passation de l'entretien d'explicitation.

Le guide d'entretien reprenait ainsi les axes de la grille d'analyse de rapports d'évaluation (soit le référentiel d'analyse de l'activité évaluative qui fait l'objet de l'explicitation) :

- L'axe 1 portait sur la précision concernant la description de la démarche d'évaluation.
- L'axe 2 portait sur la précision concernant les sources d'information.
- L'axe 3 portait sur la précision concernant les procédures de collecte de données.
- L'axe 4 portait sur la précision concernant les procédures d'analyse des données qualitatives et quantitatives.

Les questions ou sous-questions au sein des différents axes ont été formulées à partir des éléments clés issus de l'analyse des rapports d'évaluation. En fonction des éléments trouvés, nous sommes revenus sur certains critères de la grille afin de comprendre ce qu'il s'est passé durant l'évaluation.

Compte tenu des dimensions importantes de notre étude, et les objectifs de cette dernière, les potentiels axes de relance furent les suivants :

- L'explicitation des choix méthodologiques opérés (de la conception à la réalisation).
- La perception des évaluateurs quant à l'influence potentielle des parties prenantes sur l'opérationnalisation des choix méthodologiques et la mise en œuvre du protocole d'évaluation.
- La perception des évaluateurs quant à l'influence potentielle des contraintes inhérentes au terrain sur l'opérationnalisation des choix méthodologiques et la mise en œuvre du protocole d'évaluation.

Au regard des caractéristiques spécifiques de la technique d'explicitation, nous avons construit les guides d'entretien d'explicitation et préparé les passations en tenant compte des consignes prodiguées par Vermersch (1994, 1991).

Dans le cadre de l'entretien d'explicitation, l'interviewé est considéré comme un observateur de ses propres actions procédurales et mentales qu'il se remémore en situation d'évocation (Vermersch, 1994, 1991). Faire entrer le sujet en situation d'évocation signifie le

conduire à faire exister mentalement la situation passée, afin de recueillir ses représentations de l'action telle qu'il l'a vécue. Aussi doit-on s'assurer que ce dernier se trouve bel et bien « dans la situation (ou cœur de l'activité évaluative) » au moment où il la verbalise. En outre, une telle caractéristique implique que le schéma de questionnement de l'entretien guide à la fois le recueil d'informations et le maintien de l'interviewé dans le domaine des faits (dans le rôle d'observateur de ses propres actions dans la cadre de l'activité en situation de rappel). Les questions posées devaient donc amorcer le processus de description des actions mises en œuvre dans le cadre de l'activité évaluative explicitée et la prise d'informations sur l'exécution de cette dernière. Pour s'assurer d'aller au-delà des généralités et des savoirs déclaratifs au moment de l'évocation, l'entretien devait demeurer très directif. Il était donc primordial de s'appuyer sur l'analyse de l'activité évaluative et de l'inviter à décrire et à discuter du « comment » et du « quoi » des éléments clés mis en exergue dans le cadre de la préparation de l'entretien. Il est important de noter qu'il fallait éviter le recours au « pourquoi » ainsi que toutes autres questions sur la causalité de la situation pour éviter de rompre l'évocation et d'amorcer un processus de justification (Martinez, 1997 ; Vermersch, 1994, 1991). Ainsi, dans le cas de nos entretiens, le fil conducteur du questionnement reposait sur la verbalisation descriptive de contenus qui portaient sur *le contexte et l'environnement* (les informations sur les circonstances de la réalisation de l'activité évaluative), et sur *les actions du sujet* (les informations relatives aux procédures, ainsi que les raisonnements mobilisés en amont des différentes prises de décision dans le cadre de l'exécution de l'activité évaluative, etc.). La formulation des questions et des relances était déterminée par la nécessité d'obtenir dans la verbalisation un niveau de détails optimal afin de rendre intelligible l'action décrite, au regard des objectifs de l'entretien (Vermersch, 2004). Dans le cadre de notre étude, il s'agissait d'accompagner l'évaluateur à nous faire part des détails sur les différents choix et procédures méthodologiques opérés au fil de la conception et de la mise en œuvre de l'évaluation qui fait l'objet d'explicitation. Nous désirions obtenir une description des processus décisionnels en amont de ces choix dits méthodologiques et de saisir l'influence que les facteurs contextuels ont pu avoir sur ces derniers.

Enfin, nous avons eu recours à des questions descriptives de la forme suivante (extrait d'un des guides d'entretien dans l'Encadré 7 ci-dessous) :

- *« Vous mentionnez que… pouvez-vous m'en dire davantage ? »*
- *« Vous avez recouru à telle méthode… comment avez-vous fait ? »*

- *« Vous faites référence au caractère complexe de la situation… quels sont les défis que vous avez rencontrés ? Quelles décisions avez-vous dû prendre ? etc. »*
- *« Quelles tâches avez-vous accomplies ? », « comment avez-vous procédé ? », « quelle méthode avez-vous utilisée pour… ? », etc.*
- Les relances étaient du type *« pouvez-vous m'en dire plus ? », « c'est-à-dire ? », « pouvez-vous décrire davantage ? », « comment ? », « quelles étaient les étapes ? », etc.*

Encadré 7. *Extrait du guide d'entretien d'explicitation du participant [COH7]*

Avant que l'on entre dans les détails de l'évaluation,
Question 1 : vous mentionnez dans le rapport le cabinet X ainsi qu'une équipe d'évaluation, quelle a été votre contribution dans cette évaluation et la rédaction du rapport ?
Question 2 : pouvez-vous me décrire brièvement le mandat d'évaluation du X ?

Axe 1 : Description de la démarche d'évaluation

1. Choix du type d'évaluation

Afin de répondre au mandat d'évaluation, vous avez proposé une évaluation sommative axée sur les résultats, la mise en œuvre et la cohérence. Pour ce faire, vous mentionnez que cette évaluation s'inscrit dans une approche participative,

1.1 Pouvez-vous m'en dire davantage quant à ce choix ?
1.2 Qui sont les personnes que vous avez impliquées ?
1.3 Comment s'est déroulée la mise en œuvre de cette approche participative ?
1.4 Comment s'est déroulée la coordination de l'évaluation ?

2. Complexité du programme

Vous mentionnez dans le rapport le caractère complexe du X, notamment du fait de sa mise en œuvre décentralisée,

2.1 Pouvez-vous m'en dire davantage ?
2.2 Quels sont les défis que vous avez rencontrés durant l'élaboration du devis d'évaluation ?
2.3 Vous mentionnez également un manque de ressource notamment au niveau de l'évaluation des résultats, pouvez-vous me décrire les choix que vous avez dû faire en ce qui a trait au développement de la méthodologie ?

3. Présentation des questions d'évaluation

Vous avez fait le choix de présenter les questions d'évaluation en annexe du rapport,

3.1 Pouvez-vous m'en dire plus quant à ce choix ?
3.2 Comment avez-vous développé les questions d'évaluation ?

Enfin, les procédures de passation de l'entretien d'explicitation ont été préexpérimentées auprès de deux doctorants en psychologie qui ont suivi l'ensemble des cours du programme de formation en évaluation de programme et ont réalisé leur internat dans un cabinet de consultants en recherche et en évaluation. Aussi, l'un des doctorants avait participé à l'activité de préexpérimentation de l'entretien métaévaluatif afin que l'on soit dans des conditions similaires à celles de la recherche. En effet, nous étions conscients qu'à la suite de l'entretien métaévaluatif, où les participants ont adopté une posture critique vis-à-vis du travail d'autres évaluateurs, ces

derniers pouvaient se percevoir comme étant l'objet de la critique à leur tour. Aussi, nous avons pu pratiquer le processus d'immersion de l'interviewé dans la situation d'évaluation qui fait l'objet de l'entretien, tout en le mettant en confiance. Nous nous sommes familiarisés avec le processus d'explicitation, une pratique d'entretien plutôt contre nature puisque le schéma d'entretien est très directif et que la formulation des questions et des relances doit avoir un format particulier afin de maintenir l'interviewé dans l'évocation et la description. Enfin, cela nous a permis d'identifier comment collecter les informations sur les procédures engagées durant l'exécution de l'activité évaluative qui fait l'objet de l'entretien (le déroulement de l'évaluation, les choix méthodologiques opérés, et les raisons évoquées dans le cadre de la description des dits choix, etc.), sans compromettre le processus d'explicitation et pousser l'interviewé à se justifier.

3.4.2.2.2 Procédures de passation et d'analyse de l'entretien d'explicitation

Les conditions de passation de l'entretien d'explicitation furent les mêmes que dans le cadre de l'entretien métaévaluatif. Les entretiens se sont déroulés pendant les heures de travail des évaluateurs participants, soit directement dans les locaux de leur cabinet, soit sur Skype. Les entretiens ont été enregistrés à l'aide d'un dictaphone, ils se sont déroulés dans la langue de préférence des participants (en anglais ou en français) et ils ont duré en moyenne entre 60 et 120 minutes. À chaque début d'entretien, nous avons pris le soin de réitérer les objectifs de la passation et de bien décrire l'activité évaluative qui faisait l'objet de l'explicitation. Aussi, nous leur résumions brièvement l'évaluation de programme rapportée dans le rapport qu'ils nous ont communiqué afin de déclencher les procédures de rappel. Nous leur expliquions que nous avions relevé différents éléments lors de notre lecture du rapport et que nous aimerions qu'ils nous décrivent le déroulement de l'évaluation qui fait l'objet de l'entretien, afin de les mettre progressivement en situation d'évocation. Tout au long de l'entretien, nous nous sommes assurés de les maintenir dans l'évocation en relançant autant de fois que nécessaire pour qu'il se représente les différentes actions discutées, etc. Nous avons également veillé à ne pas leur demander « pourquoi » ils avaient fait (ou n'avaient pas fait) telle ou telle chose. La plus grande difficulté fut de les amener à nous décrire des éléments qui faisaient partie intégrante de l'activité évaluative, mais qui n'étaient pas mentionnés dans le rapport. Aussi, nos connaissances pratiques en évaluation de programme et le recours à la grille d'analyse de rapports d'évaluation nous ont vraiment préparés à la conduite d'un tel entretien et à ce type de situation. Dans l'ensemble, les

entretiens ont été très fructueux et les évaluateurs ont grandement apprécié d'avoir eu l'occasion de discuter de leurs pratiques évaluatives et de revenir sur le déroulement d'une évaluation à laquelle ils ont pris part.

À la suite de la retranscription des verbatim, nous avons réalisé deux types d'analyse de contenu des données qualitatives collectées dans le cadre de l'entretien. D'abord, nous avons utilisé le référentiel d'analyse de l'activité évaluative dont chaque axe (avec ses thèmes, sous-thèmes et indicateurs) représente une étape de la démarche d'évaluation (de la conception du devis d'évaluation aux procédures d'analyses des données). L'objectif de cette analyse était de répertorier les informations relatives à la description des différentes actions, particulièrement celles d'ordre méthodologique, mises en œuvre durant l'exécution de l'activité évaluative explicitée (Gouju, Vermersch et Bouthier, 2003). Puis, nous avons réalisé une seconde analyse de contenu, plus fine, à l'aide du logiciel d'analyse de données qualitatives Atlas.ti. Cette analyse visait, dans un premier temps, à mettre en exergue les explications fournies par les évaluateurs quant aux choix méthodologiques opérés dans le cadre de l'évaluation à laquelle ils ont participé. Puis, dans un deuxième temps, à identifier (et décrire) les caractéristiques des pratiques réelles des évaluateurs ainsi que les facteurs contextuels qui exercent une influence sur ces dernières, particulièrement en ce qui a trait à l'opérationnalisation des choix dits méthodologiques.

Pour ce faire, nous avons eu recours au recours au même processus de modélisation et de validation du cycle de codage que celui mis en œuvre dans le cadre de l'analyse des données de l'entretien métaévaluatif. Nous avons eu recours à une méthode de codage itératif hybride (1er cycle-2e cycle). Aussi, la Figure 16 (p. 136) illustre le processus d'analyse des données collectées dans le cadre de la deuxième activité de recherche.



Figure 16. *Processus de codage éclectique des données qualitatives de l'activité d'explicitation*

Enfin, un rapport d'analyse et de présentation de résultats sous forme de méta-matrices partiellement ordonnées (Mile, Huberman et Saldaña, 2014) a été produit pour chaque participant.

3.5 Conclusion du chapitre

Si l'étude de cas est largement reconnue par la communauté scientifique pour sa contribution à la recherche exploratoire, la scientificité et la crédibilité de ce devis méthodologique, demeure souvent critiqué (Alexandre, 2013). Rappelons-le, « la crédibilité des recherches repose largement, sinon prioritairement sur la rigueur de leur méthodologie » (Dufays, 2001, p. 8). Ainsi à chaque étape de la conception du cadre méthodologique de cette recherche, nous nous sommes assurés de tenir compte des normes scientifiques de la recherche de type qualitatif et des stratégies proposées pour que cette étude de cas soit d'une rigueur optimale.

Afin de rendre compte de la rigueur de notre démarche méthodologique, nous proposons en guise de conclusion une représentation graphique du protocole de l'étude de cas multiples que

nous avons réalisée dans le cadre de cette étude doctorale (Figure 17, p. 138). Ce protocole illustre la méthodologie mise en œuvre afin de décrire en profondeur l'influence que les facteurs contextuels, soit le profil des évaluateurs, la relation avec les parties prenantes, les contraintes inhérentes au terrain, peuvent avoir sur l'opérationnalisation des choix méthodologiques des évaluateurs dans le cadre de leurs opérations d'évaluation.

Enfin, la réalisation de cette étude de cas multiples devait nous permettre de proposer une modélisation « cas de novices » et une modélisation « cas d'experts » qui illustrent les processus décisionnels en amont des pratiques méthodologiques des évaluateurs (selon qu'ils soient novices et experts) et leurs interactions avec les facteurs contextuels propres à leurs pratiques évaluatives. Nous abordons les procédures d'analyse de l'étude de cas multiples, réalisée à partir des rapports descriptifs de chaque cas individuel, dans le chapitre suivant qui porte sur la présentation des résultats.

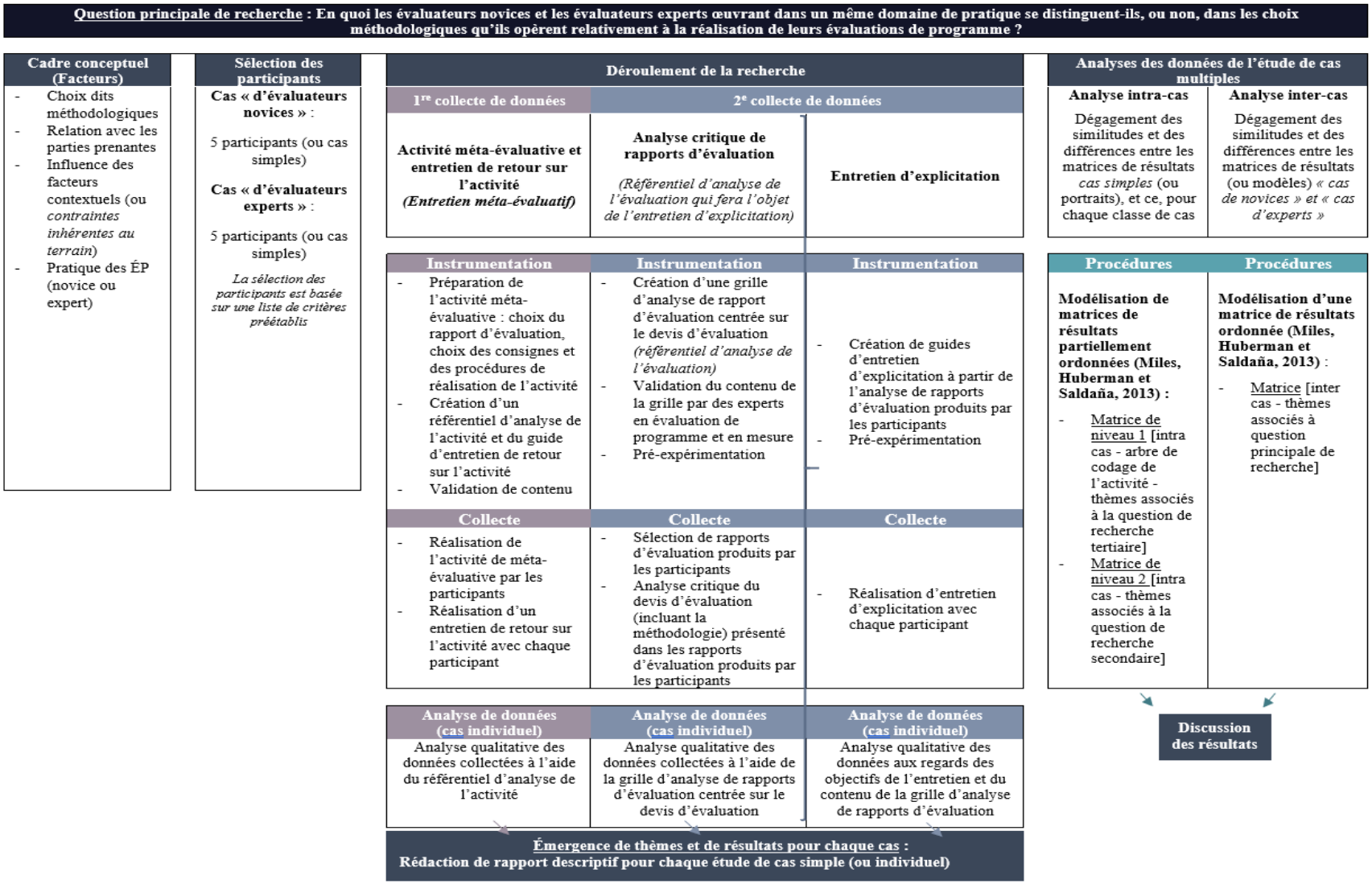


Figure 17. *Protocole de l'étude de cas multiples sur les processus décisionnels relatifs aux choix dits méthodologiques des évaluateurs dans le cadre de leurs opérations d'évaluation de programme (inspiré de Kinney-Walker, 2015 ; Yin, 1994)*

CHAPITRE 4. RÉSULTATS ET INTERPRÉTATION

À travers ce quatrième chapitre, nous présentons les résultats d'analyses des données collectées dans le cadre des entretiens métaévaluatifs et d'explicitation réalisés auprès des évaluateurs de programme participants, et leur interprétation au regard du cadre conceptuel de cette recherche doctorale. Aussi, à titre de rappel, dans le cadre de :

1. L'entretien métaévaluatif, les évaluateurs participants ont été invités à adopter une posture critique et analyser le contenu d'un rapport d'évaluation tiers. L'objectif principal de cette première collecte de données était d'identifier (et comprendre) ce qui constitue pour ces derniers de bonnes pratiques d'ordre méthodologique, et le rôle que peuvent jouer les facteurs contextuels (contraintes de terrain, statut de l'évaluateur, collaboration avec les parties prenantes, etc.) sur le plan de ces pratiques.
2. L'entretien d'explicitation, les évaluateurs participants ont été placés en situation d'évocation et amenés à nous décrire leur expérience, leur pratique, dans le cadre d'un projet d'évaluation auquel ils ont participé. L'objectif principal de cette deuxième collecte de données était de décrire les caractéristiques des pratiques d'ordre méthodologique réelles des évaluateurs de programme, et de comprendre dans quelle mesure les facteurs contextuels (contraintes de terrain, statut de l'évaluateur, collaboration avec les parties prenantes, etc.) influencent les choix méthodologiques opérés par ces derniers.

En outre, nous tenons à préciser que le contexte dans le cadre duquel les données ont été collectées et analysées présente des caractéristiques très spécifiques, caractéristiques qui jouent un rôle déterminant dans la couleur des résultats et de notre interprétation. Si nous prenons soin de les évoquer tout au long de ce chapitre, il nous paraît pertinent d'en introduire certaines dès à présent :

- Afin de garantir la possibilité de comparaison et la fiabilité des résultats, le rapport faisant l'objet de l'activité métaévaluative, et les rapports fournis par les participants dans le cadre de l'activité d'explicitation, concernent des évaluations externes commanditées par des départements ou agences du secteur public canadien (programmes fédéraux).
- Les évaluations en question s'inscrivent dans un cadre politique plus ou moins prescriptif, sous la gouvernance d'une institution que nous identifierons « IGF ». Elles ont été

commanditées par un comité d'évaluation interne constitué de parties prenantes clés du programme dont le principal interlocuteur est un évaluateur de programme de la division interne de l'évaluation de l'agence commanditaire qui est accrédité par la Société canadienne d'évaluation.

- Ces évaluations externes sont entreprises par des équipes évaluatives externes au sein desquelles les évaluateurs juniors (novices) détiennent davantage un rôle d'exécutant dans le cadre de(s) collecte(s) et, ou, analyse(s) des données évaluatives. De fait, les données collectées au cours de l'entretien d'explicitation se restreignent à l'exécution des choix méthodologiques opérés par l'évaluateur sénior responsable du projet (expert) et les clients. Par conséquent, les résultats portent, dans la mesure du possible, sur leurs explications des activités méthodologiques auxquelles ils ont pris part et des décisions prises par leurs supérieurs.

Enfin, ce quatrième chapitre est découpé en cinq parties. Sa structure (Tableau 10), correspond au processus de modélisation des résultats sous forme de *métamatrices partiellement ordonnées* (matrices de niveau 1 [intra cas - arbre de codage de l'activité - thèmes associés à la question de recherche tertiaire], puis matrices de niveau 2 [intracas - thèmes associés à la question de recherche secondaire), et de *matrice ordonnée* (matrice [intercas - thèmes associés à question principale de recherche]), qui est recommandé par Miles, Huberman et Saldaña (2013).

Tableau 10. *Structure de la présentation des résultats et de leur interprétation*

Titre		**Section du chapitre**
Les éléments relevés par les évaluateurs quant aux choix méthodologiques, et à leur influence des facteurs contextuels (activité métaévaluative)		**4.1 (p. 141)**
Question de recherche tertiaire 1	Le cas des évaluateurs juniors (novices)	4.1.1 (p. 142)
	Le cas des évaluateurs séniors (experts)	4.1.2 (p. 180)
Les informations fournies par les évaluateurs pour expliquer et décrire les choix méthodologiques entrepris durant l'évaluation (activité d'explicitation)		**4.2 (p. 223)**
Question de recherche tertiaire 2	Le cas des évaluateurs juniors (novices)	4.2.1 (p. 223)
	Le cas des évaluateurs séniors (experts)	4.2.2 (p. 271)

Interprétation des questions de recherche secondaire **1. Quelles sont les ressemblances et les différences entre les pratiques réelles et les pratiques recommandées des évaluateurs externes de programmes publics en ce qui a trait aux choix dits méthodologiques ?** **2. Quelles sont les perceptions des évaluateurs de programme externes quant aux possibles influences des facteurs contextuels sur les choix méthodologiques opérés durant leurs évaluations de programmes publics ?**		**4.3**
	Le cas des évaluateurs juniors (novices)	
	Le cas des évaluateurs séniors (experts)	4.3.1
Interprétation de la question principale de recherche **En quoi les évaluateurs novices et les évaluateurs experts œuvrant dans des cabinets de conseil en évaluation de programmes publics se distinguent-ils dans les choix méthodologiques qu'ils opèrent relativement à la réalisation de leurs évaluations de programme ?**		**Chapitre 5. Conclusion**

4.1 Les éléments relevés par les évaluateurs quant aux choix méthodologiques et à l'influence des facteurs contextuels sur ceux-ci (*activité métaévaluative*)

Ces données ont été collectées dans le cadre de l'entretien métaévaluatif et rendent compte du jugement critique des évaluateurs participants quant au contenu du devis d'évaluation présenté dans le rapport d'évaluation d'un programme fédéral tiers commandité en réponse à la politique évaluative fédérale. Durant cette activité, l'intervieweur n'a donné aucune indication quant à la marche à suivre pour réaliser ce type d'analyse critique. En effet, nous désirions mettre en exergue les éléments sur lesquels les participants ont choisi de porter leur attention, ainsi que leurs critères d'évaluation et leurs attentes, leurs perceptions en matière de « bonnes pratiques », et l'influence des facteurs contextuels sur ces dernières. D'ailleurs, la sélection d'un rapport d'évaluation d'un programme gouvernemental d'aide sociale desservant les communautés autochtones était ici intentionnelle, et ce, en vue des enjeux politiques et socioculturels qui sont associés à un tel contexte évaluatif.

L'ensemble des évaluateurs participants à cette étude, quel que soit leur niveau d'expertise, ont utilisé des critères tels que la précision, la pertinence et la justesse des informations fournies dans le rapport d'évaluation.

4.1.1 Le cas des évaluateurs juniors (novices)

À l'unanimité, les évaluateurs juniors ont pris le temps de décrire leur niveau d'expérience ou d'inexpérience en évaluation de programme, et la façon dont ils ont abordé cette activité d'analyse critique (Tableau 11).

Tableau 11. *Contexte de réalisation de l'activité d'analyse critique selon les évaluateurs juniors*

'J1' **Junior avancé** - Firme [B] - Formation [PhD SHS] [DES ÉP inclus] - Expérience pro. ÉP [18 mois/Activité1] [28 mois/Activité 2] - Passation en anglais	Il débute par mentionner qu'il commence tout juste à voir dans le contexte de sa pratique des projets d'évaluation du début à la fin. Nous allons voir qu'il s'appuie sur ses attentes en tant qu'évaluateur en ce qui a trait à la précision sur le plan de la description du devis d'évaluation ; mais aussi sur les critères de scientificité en recherche (son profil de chercheur), à savoir la fiabilité, la validité et la représentativité du devis d'évaluation et de l'utilisation/l'appréciation des résultats (réplicabilité, alignement « question - indicateurs - instrumentation - sources d'information », effort de triangulation, communication appropriée des résultats, etc.).
'J2' **Junior associé de recherche** - Firme [A] - Formation [PhD SHS] - Expérience pro. ÉP [6 mois/Activité 1] [16 mois/Activité 2] - Passation en anglais	Il déclare immédiatement qu'il n'y connait pas grand-chose en évaluation de programme et que cela fait à peine 6 mois qu'il travaille pour le cabinet de consultant, ayant juste terminé son doctorat. Aussi, il fit preuve d'une attitude plutôt nonchalante, et résistante au cours de la passation. Aussi, il part du principe qu'il est difficile de juger de la méthodologie évaluative sans connaitre les contraintes de terrain, plus particulièrement les intrants disponibles (remarque pertinente). Nous verrons qu'il s'appuie sur sa formation de chercheur, lorsqu'il se penche notamment sur la question de la transparence sur le plan du contenu du cadre méthodologique qu'il associe à la rigueur, ainsi que la réplicabilité et la gestion des biais sur le plan des procédures d'analyse de données qualitatives.
'J3' **Junior associé de recherche (avancé)** - Firme [A] - Formation [M. Écon.] - Expérience pro. ÉP - [4 ans - 4 ans1/2] - Passation en anglais	Il ancre son analyse critique sur ce qu'il perçoit comme les standards de l'évaluation au fédéral qui guide les attentes des clients, ainsi que les pratiques de sa firme. Nous verrons qu'il s'attarde sur l'effort de description et de documentation du devis d'évaluation, de la démonstration du raisonnement évaluatif, un effort qu'il associe au renforcement de la rigueur des pratiques méthodologiques des évaluateurs de programme, et à la garantie de la confiance des clients en la qualité de l'évaluation.
'J4' **Junior avancé** - Firme [B] - Formation [M.P.P.A] [DES ÉP] - Expérience pro. ÉP [2 ans] - Passation en français-anglais	Il mobilise ses attentes en tant qu'évaluateur, ce qu'il perçoit comme des standards de pratique en évaluation de programme. S'il se penche sur la précision sur le plan de la description du plan d'évaluation, l'explicitation des méthodes et le caractère systématique des procédures d'analyse de données qualitatives, nous verrons qu'il s'est particulièrement concentré sur la transparence dans l'appréciation des résultats, la représentation et l'utilisation appropriée de ces derniers.

Si le devis d'évaluation, tel que nous l'avons défini dans le chapitre 2, est constitué du cadre d'évaluation et du cadre méthodologique de l'évaluation, les évaluateurs juniors ont principalement concentré leur attention sur la section « méthodologie » du rapport d'évaluation faisant l'objet de leur analyse critique. Ils se sont plus particulièrement attardés sur le manque de précision au niveau de la description des méthodes de collectes et d'analyses de données et, ultimement, sur les limites méthodologiques que cela engendre. La Figure 18 illustre les principaux axes et leurs thématiques respectives abordés par les évaluateurs juniors (fréquences d'occurrence présentées en *unité de sens* [uds]), dans le cadre de leur analyse critique du contenu du cadre méthodologique du rapport d'évaluation faisant l'objet de l'activité métaévaluative. Nous abordons en détail le contenu de ces différents axes dans les sections qui suivent.

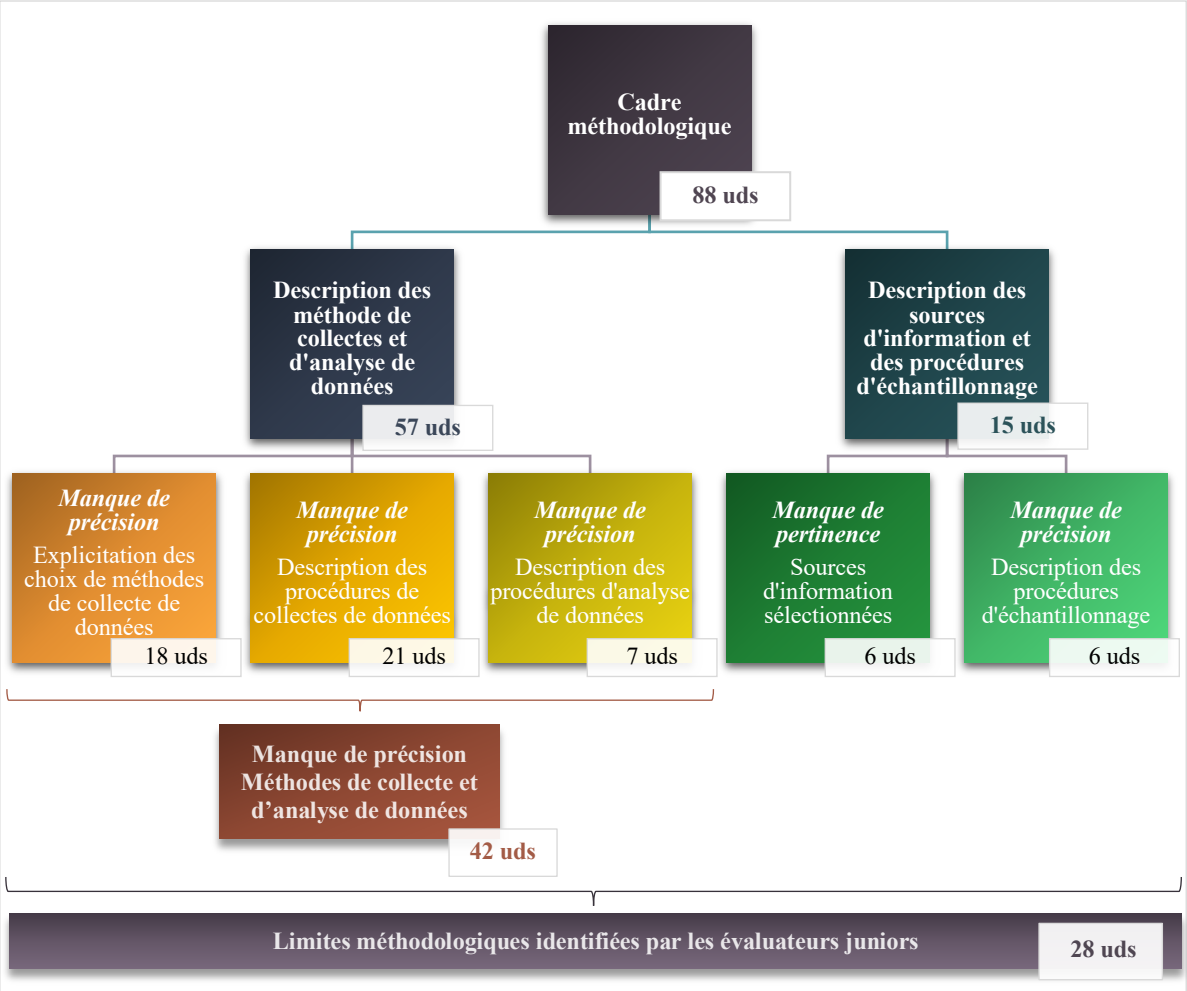


Figure 18. *Les principaux axes et thématiques d'analyse mobilisés par les évaluateurs juniors dans le cadre de l'activité métaévaluative.*

4.1.1.1 Le manque de précision au niveau de la description des méthodes de collectes et d'analyses de données au cœur du jugement critique des évaluateurs juniors

Comme la Figure 19 le montre, cet axe d'analyse (42 uds) rend compte des différents thèmes et éléments critiques mis en exergue par les évaluateurs juniors face au manque de précision (et par extension, de clarté) sur le plan de la description des méthodes de collectes et d'analyses de données.

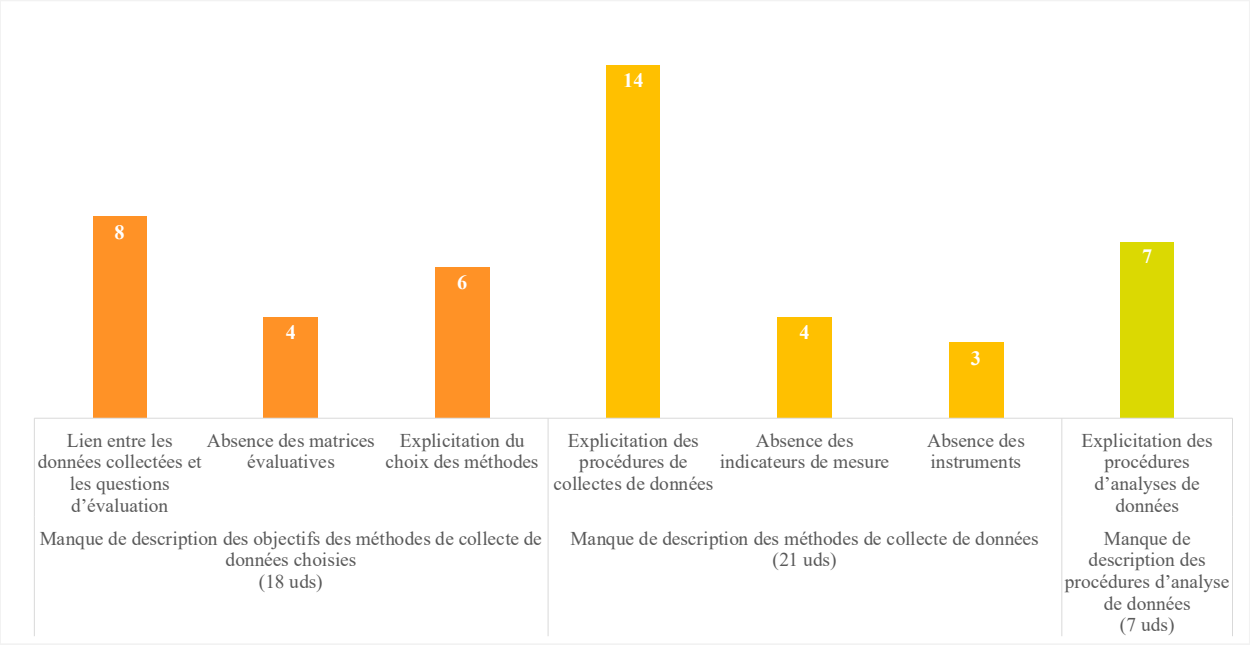


Figure 19. *Thématiques et éléments d'analyse relatifs aux méthodes de collectes et d'analyse de données (évaluateurs juniors)*

Remarquons que la conséquence du manque de description sur le plan des procédures de collectes et d'analyse de données, a conduit les évaluateurs novices à remettre en question la rigueur et la précision-justesse de la démarche méthodologique (validité, fiabilité et représentativité). Aussi, afin d'éviter la redondance, nous abordons les conséquences dudit manque de description dans la section portant sur les limites identifiées par les évaluateurs novices.

a. Le manque de description des objectifs des méthodes de collecte de données choisies

Si les évaluateurs novices reconnaissent à l'unanimité la pertinence du recours à de multiples méthodes de collectes de données, et ce, notamment à des fins de triangulation, ils trouvent d'autant plus important d'*expliciter le choix desdites méthodes* (18 uds).

En effet, pour eux, il est primordial de rendre compte du *lien entre les différentes méthodes de collectes de données et les questions d'évaluation* (8 uds). Et, au regard des informations données, le lecteur devrait être en mesure d'identifier pour chacune des sources de données[23] choisies, les questions d'évaluation auxquelles elles sont censées répondre. En outre, si dans les cas de 'J2' et 'J3', le lectorat en question fait référence aux utilisateurs ciblés de l'évaluation (clients - financeurs, hauts gestionnaires de programme, etc.), pour 'J1' et 'J4' il est question d'eux-mêmes, en tant qu'évaluateurs de programme, et de leurs attentes en ce qui a trait au caractère complet du contenu du devis d'évaluation rapporté. Par ailleurs, dans le cas de 'J1', il est également question de rendre compte des efforts de triangulation des sources de données dans le cadre de l'investigation de chacune des questions d'évaluation (préoccupations relatives à la validité, fiabilité et représentativité du devis d'évaluation).

Extraits de verbatim

"I think it would be important to people who are funding the evaluations to understand you know why you're using the lines of evidence that you're using. What questions it is that you're hoping to address." (J3).

"I am an evaluator, and it would be nice to see the specific indicators, and the data sources they use. For example, the literature review what questions where they answering to? Like you know they have 3 or four different lines of evidence, so then I could see if they had a good representation for each of them. You know they had a lot of lines of evidence but did each line of evidence actually informed the questions. But I still think that... I don't have any problems with the question, you just kind of wonder about the fidelity. The triangulation process wasn't always clear." (J1).

Ainsi, les évaluateurs juniors déplorent l'absence de *la matrice d'évaluation dans le rapport critiqué (4 uds),* cette dernière étant selon eux un outil indispensable qu'il s'agisse de permettre au lecteur (le client) de comprendre comment les évaluateurs vont répondre aux différentes questions d'évaluation, d'assurer la planification rigoureuse de la méthodologie évaluative (alignement « question d'évaluation - éléments à examiner/indicateurs - sources de données à utiliser »), ou encore de démontrer/examiner l'effort de triangulation des sources de données. En ce qui a trait à l'utilisation de cet outil, pour 'J2', l'évaluateur le plus junior (-6 mois

[23] Le terme source de données fait référence ici au terme anglais « *line of evidence* » utilisé dans le rapport d'évaluation ainsi que par nos évaluateurs participants. Aussi, ce terme évoque la combinaison méthode-source d'information - par exemple, entrevue d'informateurs clés, enquête de satisfaction du personnel, ou encore analyse des données.

d'expérience) sans formation initiale en évaluation de programme, il s'agit d'une pratique standard, une exigence de l'évaluation au gouvernement fédéral. 'J3', l'évaluateur junior formé sur le terrain avec le plus d'expérience professionnelle en évaluation de programmes fédéraux, nous explique quant à lui que sa firme utilise de manière standard un « *evaluation crosswalk* », qui est toujours annexé au rapport final. En outre, il souligne que si la matrice évaluative n'est pas incluse dans le rapport, cela ne signifie pas nécessairement que les évaluateurs ne l'ont pas conçu (elle peut avoir été communiquée dans un rapport technique). Enfin, les évaluateurs juniors ayant une formation initiale en évaluation de programme, expliquent qu'en tant qu'évaluateurs, ils s'attendent à ce que des matrices évaluatives aient été utilisées, et que celles-ci soient annexées dans le rapport d'évaluation final.

Extraits de verbatim

"I know that at my firm we prepare what we call an evaluation crosswalk. What that means is we take our evaluation crosswalk or evaluation matrix, and for each evaluation question we will list the lines of the evidence that we're going to use to answer that question. And they may have done it. I mean this is a final evaluation report. And the mention in this document that they did in fact prepare another evaluation report. So there may have been some more detail there in terms of you know what questions they were going to try and answer with each of the lines of evidence that they were planning to use. But then again you can't see that here." (J3).

"They didn't have the evaluation matrix. So I don't know if you're familiar with that term but it's pretty standard I think in a lot of federal evaluations it requires such a thing. It's basically just a chart that connects you know the questions for example that they've listed under section X for relevance. It's going to connect them all directly to the methodology they're going to use to address that. So that the reader and the person who orders the evaluation know right away why we used this methodology to address this." (J2).

"It would have been nice to see which evaluation questions you hoping to answer with which line of evidence. To see for each evaluation question, the indicators listed, with the method used. You have an evaluation matrix or an evaluation protocol that you use to organize this information. You should always have it included in the evaluation report, at least in the appendices. It's something I expect as an evaluator." (J4).

Nous avons vu dans le Chapitre 2 que la matrice d'évaluation est le produit direct de la conception du cadre d'évaluation et qu'elle sert de cahier des charges (synthèse du devis d'évaluation), de guide, dans l'opérationnalisation du cadre méthodologique de l'évaluation. Aussi, c'est pour cette raison que nous avons fait le choix de les inclure, ici, dans la discussion sur le cadre méthodologique.

Enfin, outre la nécessité de clarifier le lien entre les questions d'évaluation et les différentes méthodes de collectes (notamment en ayant recours à la matrice d'évaluation), les évaluateurs juniors soulignent qu'il est aussi essentiel de *justifier le choix des méthodes de collectes de données*

(6 uds). En effet, face à l'absence de telles informations, le lecteur s'interroge quant à la vraie raison en ce qui a trait aux décisions méthodologiques prises dans le cadre de l'évaluation. Par ailleurs, 'J4' fait allusion au caractère prescriptif de l'évaluation de programme au fédéral et le recours à des méthodes standards, ce qui selon lui ne justifie pas pour autant le manque d'explication du choix des méthodes.

Extraits de verbatim

"1st thing I noticed. There wasn't any rational for the methods that were chosen. For example, they did case studies. Why? As opposed to do a survey or other methods. [...]. So yes, it wasn't clear why these methods were chosen. Obviously, it's much bigger than what's presented in the report. And I guess federal evaluations are pretty prescriptive, there is this policy that requires to cover these issues and a standard way to do it in terms of methods. But still, we need more here." (J4).

"You start to wonder was there a theoretical reason for using these methodologies or did you just do them because you had funding or because you know you wanted to be as rigorous as possible." (J3).

b. Le manque de précision (et de clarté) dans le cadre de la description des différentes méthodes de collectes de données

Les évaluateurs juniors ont exprimé à plusieurs reprises leur frustration en ce qui a trait à la description, ou plutôt *le manque de description des différentes méthodes de collectes de données mises en œuvre dans le cadre de l'évaluation* (21 uds).

Ils expliquent avoir du mal à comprendre ce qu'impliquaient les différentes collectes de données ainsi que la manière dont les évaluateurs s'y sont pris pour les réaliser. Ils déplorent à l'unanimité *l'absence d'explicitation des procédures de collectes de données* (14 uds), insistant notamment sur le fait que donner le nom de la méthode de collecte de données ne suffit pas pour comprendre de quoi il s'agit et ce que cela implique en termes de procédures. Pour ces derniers, à lecture du contenu du cadre méthodologique, nous devrions être en mesure de connaitre les détails de la mise en œuvre des différentes méthodes de collectes de données (soit comment, par qui, etc., les données ont-elles été collectées ?).

Extraits de verbatim

"Again, in terms of methodology, it's just some vagueness in the term they used. They talk about under still the case studies which is the worst in terms of description. They talk about cultural appropriate methods. What are those? What does that mean? What are the methods you are referring to?" (J4).

"They don't tell me what they mean by literature review right. It can be anything from I googled this you know to a full Cochrane systematic review. […]. So, it was a little confusing you know because that term can mean so many things. I would have liked to have been more explicitly laid out right." (J2).

"They also didn't say who is conducting the interview. So, for example, here when we do evaluation with aboriginal communities, we make sure to have someone here has worked in these communities before and has a lot of experience." (J1).

Dans la lignée de leur questionnement sur le lien entre les questions d'évaluation et les différentes collectes de données, les évaluateurs juniors ont remarqué l'absence de *description des indicateurs de mesure* (4 uds) (mentionner en lien notamment avec l'*absence des matrices évaluatives*) ; et, ou l'impossibilité d'examiner en détail les questions qui ont été posées aux participants ou les différentes informations que les évaluateurs ont cherché à recueillir, notamment du fait de l'*absence des instruments de collectes de données* (3 uds). Selon 'J1' et 'J3', les instruments de collectes de données devraient toujours être inclus dans les rapports d'évaluation.

Extraits de verbatim

"I am an evaluator, and it would be nice to see the specific indicators they use. […]. They had a lot of lines of evidence but did each line of evidence actually informed the questions?" (J1).

"Again, here it would have been nice to see for each evaluation question, the indicators listed, with the method used. To see an evaluation matrix or an evaluation protocol. It should be included. Also, the logic model would have been useful too here." (J4).

"I think in the reports that we write we would attach the research instrument trade. So here we are doing I think we're doing a final report for a federal government entity. You know you will have the exact list of questions that we presented to key informants as well as you know we have a survey, for example." (J3).

"I don't even have the guides in here. Usually when you look at a report you look at the guides and then you critique. Like, did you ask the right questions? Is the content of the instrumentation aligned with the content of the matrix, the indicators?" (J1).

Le manque de description sur le plan des procédures de collecte de données, l'absence d'outils tels que le modèle logique, la matrice d'évaluation/protocole d'évaluation (incluant la définition des indicateurs) ou encore l'instrumentation, empêchent les évaluateurs juniors d'apprécier la pertinence/précision-justesse de la démarche méthodologique. Nous revenons davantage sur cet aspect dans la section portant sur les limites identifiées par les évaluateurs juniors.

c. Le manque de précision sur le plan de la description des procédures d'analyse de données

Tout comme ce fut le cas avec les procédures de collectes de données, les évaluateurs juniors s'interrogent sur *les rouages du processus d'analyses des données qualitatives* (7 uds) face au caractère inutile d'énoncés génériques tels que « j'ai analysé les données selon les thèmes de recherche ». Aussi, les évaluateurs juniors se sont notamment questionnés sur le type d'analyse, les procédures de codages, la prise en considération des biais (la triangulation ou non des

évaluateurs), ou encore le recours à un logiciel d'analyse de données qualitatives (*logiciel QDA*). S'ils ne font pas nécessairement référence à une méthode préférentielle, ils insistent sur l'importance d'être transparent et de permettre au lecteur d'être en mesure de discerner clairement la façon dont les résultats ont été obtenus. Ils insistent en revanche sur la nécessité de rendre compte du caractère systématique de la démarche d'analyse de données qualitatives, de l'effort de rigueur.

Extraits de verbatim

"Interview data were transcribed and analyzed according to research themes right. So again, what does that mean exactly? Was it one person that analyzed the themes and then coded those themes? Was there really like sort of consideration of how bias might have come into that? Did you have multiple people looking at the coding scheme? […] I'm not saying that you can't do a synthesis based on just reading what they did and doing a narrative synthesis, but they should say that's what they did right so that I can interpret that in light of that. Also, […] what did they mean by analyzing? There are a variety of ways you could do it like coding at the sentence level and concept level. Using different types of software, different themes." (J2). *"You don't want to use terms like we did an analysis, and this was the result. Everyone would say well what analysis did you do? You can't just say I analyzed the data and then give me your conclusion. Like there would be very specific requirements depending on the type of analysis you are doing, and how you would report the results. Also, there should be some kind of systematic approach.*

"I think more and more the readers want to know exactly what you did to arrive at the patterns that you actually got. And I think that that leads into what I alluded before about the process of analysis. […] particularly in the context of qualitative information […] because the word analysis gets thrown around a lot and there's not necessarily very much thought about exactly what that means. What is it actually? What are the nuts and bolts of analyzing something?" (J3)

"Which also leads me to question how the analysis was conducted? Which they also didn't explain. Like they did all these key informant interviews. Did they just look, and they decide oh this is how…like was it systematic? We don't know." (J4)

Le manque de description, et l'absence de démonstration du recours à une démarche rigoureuse et systématique dans le traitement des données qualitatives poussent les évaluateurs juniors à remettre en question la précision-justesse des procédures d'analyses de données. Nous revenons davantage sur ce point dans la section sur les limites identifiées par les évaluateurs juniors où nous discutons des enjeux de rigueur, de validité, de fiabilité et de validité du devis d'évaluation face au manque de transparence sur le plan des procédures de collectes et d'analyses de données.

4.1.1.2 Le manque de précision et de pertinence en ce qui a trait aux sources d'information et caractéristiques des échantillons

Nous l'avons mentionné plus tôt, l'analyse critique du devis d'évaluation réalisée par les évaluateurs juniors fut concentrée sur le cadre méthodologique (88/108 uds), plus particulièrement

sur **le manque de précision au niveau des méthodes de collectes et d'analyses des données (42 uds)**. En outre, comme la Figure 20 le montre, les participants se sont tout de même penchés sur le *manque de précision et de pertinence en ce qui a trait aux sources d'information et caractéristiques des échantillons* (12 uds).

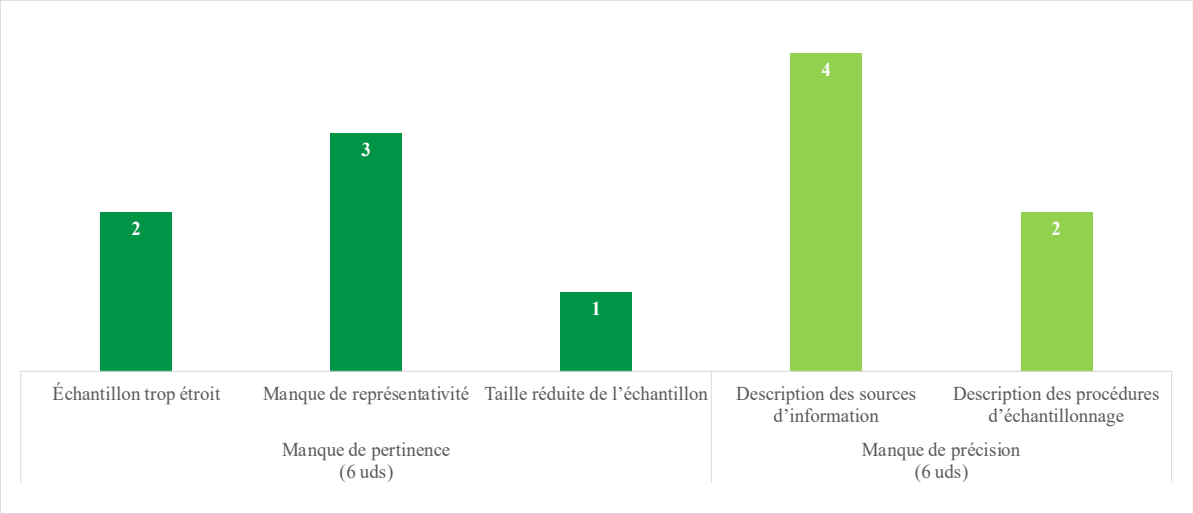


Figure 20. *Thématiques et éléments d'analyse relatifs aux sources d'informations et échantillonnages (évaluateurs juniors)*

a. Le manque de pertinence des sources d'information et échantillonnages

Au regard des caractéristiques du programme, et des objectifs de l'évaluation, les évaluateurs juniors ayant une formation initiale en recherche, et (ou) en évaluation de programme ('J1'et 'J2') expliquent qu'ils s'attendaient à l'inclusion d'une perspective comparative sur le plan de critères de sélection des sources d'information dans le cadre de l'analyse documentaire et les entrevues d'informateurs clés. En effet, 'J1' et 'J2' souligne que l'évaluation de la conception et de l'implantation d'un programme gouvernemental d'aide sociale, récent, ciblant une population sensible (communautés autochtones) nécessitait un effort de comparaison interjudictionnelle, ont critiqué l'absence de l'inclusion d'une perspective comparative dans le cadre du choix des sources d'information, particulièrement en ce qui a trait à l'analyse de documents et de données secondaires provenant de l'évaluation de programmes plus matures, et ce, au travers des différentes provinces canadiennes (comparaison interjuridictionnelle). Aussi, pour ces derniers l'*étendue de certains échantillons est trop étroite* (2 uds). Pour 'J1', il aurait été pertinent de recueillir la perspective des membres du personnel, des clients du programme, mais aussi de recueillir l'expérience de gestionnaires de programmes similaires dans d'autres provinces du pays. De plus,

en ce qui a trait aux informateurs clés dans le cadre des entrevues, 'J4' mentionne qu'il a eu l'impression que la consultation des membres de la population ciblée par le programme semblait être un ajout ou une pensée « après coup ». L'ensemble des constats énumérés *remet donc en question la représentativité de certains échantillons* (3 uds).

Extraits de verbatim

"The key informant interviews, I don't think they were very representative here. For example, they could have spoken more to the other provinces doing the same type of work that's also financed by X because that would have been important to inform their work since they are a newer program. Also, other provinces have done evaluations in the past and they could do a comparative analysis or even a cross-jurisdictional review, look at it internationally and other programs with the same type of work to inform their work because they are so new. At a high level this would be my main critic in terms of weakness regarding the information sources" (J1).

"Their data sources are a little narrow as far as I can tell. Again, it's difficult to know the constraints or timeframe they may be under, but it seems like they really just looked at stuff inside province X to evaluate a program in province X. […]. I thought it was a bit weird to do a document review that seemed to focus so much just on what's happening in province X. But I would think that programs of this nature for the indigenous communities probably occur in other provinces or even other countries. Australia is a really good comparison for Canada on many X-related and social issues because our government structures are quite similar. I thought that there was a bit of a weakness there." (J2)

"My biggest issue is it like it almost seems like the beneficiaries were almost like an afterthought. They are not really involved in the data collection. In the case studies they said they interviewed chief, counselled, community elders and parents. I guess they did some, but we don't know how many of them. Obviously, they did not talk to any children and it's a child prevention, protection program. I understand there are some sensitivities involving children for this type of evaluation. But at least, they could have explained why they did not » (J4).

b. Le manque de précision sur le plan de la description des sources d'information et caractéristiques de certains échantillons

Les évaluateurs juniors regrettent le *manque d'identification et de description des sources d'information ainsi que des caractéristiques de certains échantillons et des procédures d'échantillonnage* (6 uds). Par exemple, s'ils ont la liste des sites où ont été réalisées les entrevues avec les informateurs clés, ils ne sont pas en mesure d'identifier les caractéristiques descriptives des individus interviewés et les raisons quant à la consultation de ces derniers. De même, en ce qui a trait à la recension des écrits, aucune indication n'a été donnée quant au type (littérature grise, articles scientifiques, etc.) et à l'objet des écrits consultés ainsi que le processus d'identification et de sélection des écrits ayant conduit à l'échantillon final. Aussi, un tel manque de description, empêche les évaluateurs juniors d'apprécier la pertinence, la représentativité des sources d'information sélectionnées. D'ailleurs, face à l'absence de description (et justification) de la procédure de sélection des écrits recensés, 'J3' se demande dans quelle mesure les évaluateurs savaient vraiment ce qu'ils cherchaient, ce dont ils avaient besoin en termes de références, et si en

fin de compte ils sont en mesure de dire que l'échantillon d'écrits qu'ils ont recensés est bel et bien pertinent.

Aussi, une fois de plus le manque de transparence pousse les évaluateurs juniors à remettre en question la pertinence/précision-justesse des procédures méthodologiques, ici la sélection des sources d'information. Nous revenons sur les enjeux associés à ce constat dans le cadre de la discussion des limites identifiées par les évaluateurs juniors. Nous verrons également que le manque de transparence quant aux sources d'information s'étend jusque dans la section résultats du rapport d'évaluation, où les évaluateurs juniors, plus particulièrement 'J1' ont eu de la difficulté à identifier la source de certains résultats utilisés pour répondre aux différentes questions d'évaluation.

Extraits de verbatim

"For the key informants, how many interviews did you do? Who did you speak to at these places? We only have a broad list here, did they speak to the director or the manager, did they speak to multiple people? Different categories of people? Did they speak to the program users? I can't see if I have a good representation here. […]. Also, for the literature review they only used 32 documents. But what type of document did they use? Was it gray literature, academic articles?" (J1)

"The citations for that sort of literature review section are pretty light, right like they say I don't know in the end that they've got 32 documents or something but I'm like Okay well that's a bit of a weirdly specific facts to report without telling me how you arrived at those being the 32 documents that they should be based on. Also, I would have liked to see a lot more citations to the sources that they spoke to or at least the identification of. You know this is the methodology." (J2)

"There is sort of a sparseness about the detail that was used. Regarding the literature review, there's a line at the end of it where they say a total of 32 documents were found on the subject and analyzed. And for me it's problematic because first of all, where are the documents? They use footnotes throughout the document but there's no reference list. For an interested reader, it's nice to know what these documents were. Second, they don't discuss how they identified the literature. So how do you know which documents you wanted to use, how do you know what was important to look at?" (J3)

En somme, pour les évaluateurs juniors, les évaluateurs ont pris beaucoup de raccourcis sur le plan de la description du cadre méthodologique de l'évaluation, raccourcis qui remettent en question la confiance qu'ils peuvent avoir en la force des résultats et la qualité de l'évaluation.

Nous allons voir à présent les limites identifiées par les évaluateurs juniors en lien avec les carences qui ont émergé au fil de leur analyse critique.

4.1.1.3 Les limites identifiées par les évaluateurs juniors

L'analyse du discours des évaluateurs juniors dans le cadre de leur analyse critique du cadre méthodologique de l'évaluation faisant l'objet de l'activité métaévaluative a facilité

l'émergence de différentes croyances venant appuyer l'identification de leurs pratiques recommandées. En outre, comme le montre la Figure 21, deux axes de préoccupations émergent dans le discours des évaluateurs juniors :

- Axe 1 : les préoccupations liées au manque de transparence, et de précision-justesse sur le plan du contenu du devis d'évaluation (18 uds)
- Axe 2 : les problèmes liés au manque de transparence, et de précision-justesse sur le plan de la synthèse et de l'interprétation des résultats (14 uds).

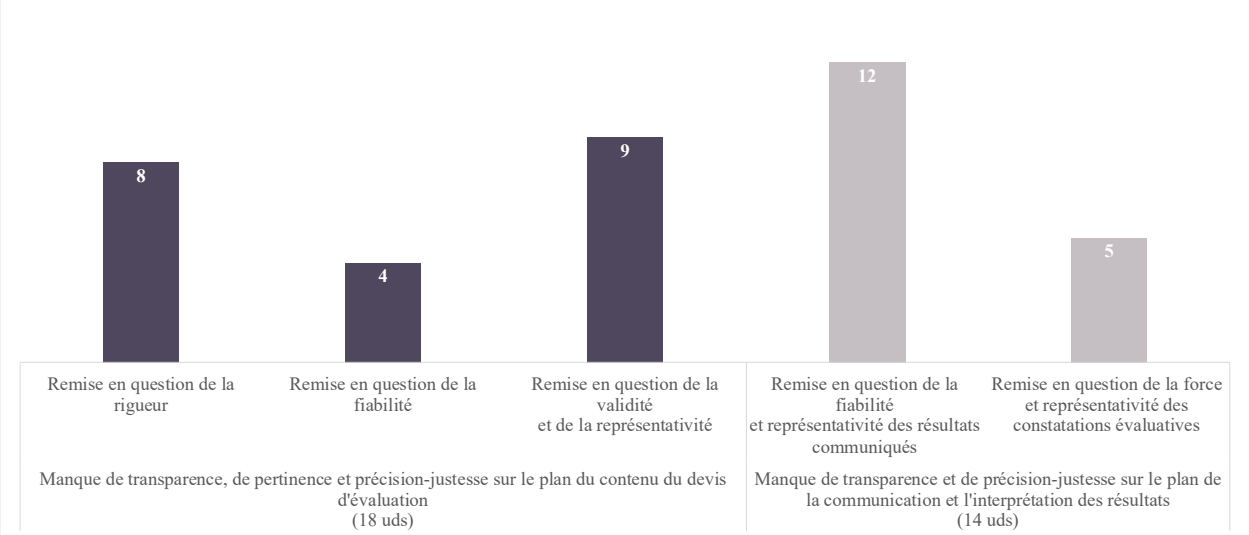


Figure 21. *Thématiques et limites identifiées par les évaluateurs juniors face au manque de description et d'explicitation du cadre méthodologique (évaluateurs juniors)*

Remarquons que si l'ensemble des évaluateurs juniors ont évoqué des limites en lien avec l'axe 1, ces derniers ne se sont pas nécessairement concentrés sur les mêmes éléments, ou encore font preuve du même degré de remise en question. L'axe 2 s'est avéré la plus grosse préoccupation de l'évaluateur 'J4', et ce, principalement du fait de la mérepresentation des données/résultats dans le cadre de la formulation des constatations (réponses aux questions d'évaluation).

a. Les préoccupations liées au manque de transparence, de pertinence et de précision-justesse sur le plan du contenu du devis d'évaluation

Ce premier axe de limites identifiées par les évaluateurs juniors émerge des points critiques centraux en ce qui a trait au manque de transparence, de pertinence et de précision-justesse des éléments clés du devis d'évaluation. Aussi, cet axe traite de la remise en question de : 1) la rigueur ; 2) la fiabilité ; et 3) la validité et la représentativité, de la démarche méthodologique.

La remise en question de la rigueur face au manque de transparence (8 uds)

S'ils reconnaissent être mal à l'aise de prononcer un tel jugement, pour les évaluateurs juniors, le manque d'information concernant les éléments clés du cadre méthodologique (les sources d'information et procédures d'échantillonnage, les objectifs des méthodes choisies, les indicateurs de mesure, les procédures de collecte et d'analyse de données, et de synthèse des résultats) signifie un manque de rigueur de la part des évaluateurs du programme. Par ailleurs, nos participants, si les évaluateurs ont en effet un espace limité pour décrire la méthodologie dans le rapport d'évaluation, il n'en demeure pas moins qu'ils ont la possibilité de donner suffisamment d'information pour clairement rendre compte des procédures méthodologiques mises en œuvre dans le cadre de leur évaluation.

À un niveau global, pour 'J1', le caractère incomplet du cadre méthodologique rapporté le pousse à « assumer », par défaut, que les évaluateurs n'ont pas fait un travail de qualité. Un jugement de valeur qui est selon lui renforcé par l'absence de la matrice d'évaluation et de l'instrumentation. 'J1' explique qu'il devrait être en mesure, par exemple, de critiquer le contenu de l'instrumentation, de vérifier que les instruments permettent bien de recueillir les informations nécessaires pour répondre aux questions d'évaluation auxquelles ils sont censés répondre. Il devrait être en mesure d'apprécier le raisonnement évaluatif (de remonté du résultat à la question d'évaluation) ainsi que l'effort de triangulation. Nous poursuivons sur ce point un peu plus loin dans les différentes discussions autour des préoccupations quant à la fiabilité, la validité et la représentativité du devis d'évaluation.

Extraits de verbatim

"They did have a lot of different methodology which is good and it's just that the details they put in the reports aren't quite clear. It seems like what you would put in an executive summary of the report and then you will elaborate or include more things in the appendix like the matrix, the instruments. So, I don't know why they didn't. It just seems to me that when you are questioning like which sources did you use? Where are your references? You know they had a lot of lines of evidence but did each line of evidence actually informed the questions. [...]. The triangulation process wasn't always clear. [...]. How many interviews did you do? Who did you speak to at these places? We only have a broad list here, did they speak to the director or the manager, did they speak to multiple people? Did they speak to the program users? I can't see if I have a good representation here. Also, I don't even have the interview guides. Usually when you look at a report you look at the guides and then you critique. Like did you ask the right questions? [...]. The case studies, we don't know the method used. So that wasn't clear and if they actually included more details that would strengthen the report and the evaluation. Who knows though they might have done an excellent job but we can see it in here. [...]. They took a lot of short cuts and flushed a lot of things out in terms of the methodology." (J1)

La relation intriquée entre la rigueur et l'effort de description des procédures de collecte et d'analyse de données se trouve au cœur de la réflexion de 'J3' qui nous explique qu'au fil des dernières années, il a pu observer un renforcement des exigences des clients en termes de description, et de documentation/communication des procédures de collecte et d'analyse de données, et ce, selon lui dans une optique de renforcer la rigueur des pratiques des évaluateurs, le recours à des processus systématique. 'J3' remarque également que la confiance des clients en la qualité de la démarche d'évaluation est tributaire de l'effort de transparence de la part des évaluateurs.

Extraits de verbatim

"Over the years I find they are demanding more rigours in terms of how we document these things, which in turn pushes the evaluators to be more rigorous. You must describe the data collection and data analysis processes, provide an evaluation crosswalk to show how it all lines up like the questions and the lines of evidence. [...] Protocols like what I mentioned earlier about using a search and reference management protocol for the literature review to show that there is a systemic process to how you search and identify, and to gradually determine which articles are important, which items you'll eventually use to answer the different evaluation questions. [...]. Also, make sure the client, or interested readers, can look at the content of the data collection instruments. It provides more confidence of about how things were done. The quality of the evaluation process. [...]. I do hope there is a more complete report somewhere with everything included in it. Probably a technical report not available to the public. But still, in my opinion, it would not have been a terrible thing for them to have three more sentences talking about how they did that. Any amount of detail would have been better to what they provide which is very, very thin." (J3)

Dans la continuité de la démonstration de l'effort de rigueur, 'J3' discute de la nécessité de concevoir et de communiquer un plan d'analyse des informations qualitatives et de synthèse des résultats provenant des différentes sources de données (processus de triangulation), un élément du cadre méthodologique qui semblerait trop souvent négligé. En effet, 'J3' nous explique que le terme « analyse » est utilisé comme si à lui seul il rendait compte de la démarche, mais que les clients désirent comprendre les bases du jugement évaluatif, les rouages en amont de la formulation des réponses aux questions d'évaluation. 'J2' fait part d'une remarque similaire face au caractère inutile de l'énoncé unique et générique « les données ont été analysées selon les thèmes de recherche et voici les conclusions », stipulant qu'en l'absence d'explicitation de la démarche d'analyse de données, il n'a pas d'autre choix que d'assumer un manque de rigueur de la part des évaluateurs.

Enfin, face au manque de description des procédures méthodologiques, l'ensemble des évaluateurs juniors font particulièrement preuve d'inquiétude quant au manque de rigueur dans le traitement des données d'entrevues, plus particulièrement le possible non-recours à une démarche

d'analyse qualitative systématique permettant de répondre avec justesse aux questions d'évaluation. 'J2' fait d'ailleurs part de préoccupations importantes en ce qui a trait à la réplicabilité et la gestion des biais. Quant à 'J4', la remise en question du caractère systématique de la démarche d'analyse des données d'entrevue émerge lorsqu'il soupçonne une interprétation ou une utilisation/représentation biaisée des résultats. Nous explorons davantage ce point dans le cadre de l'axe 2 des limites, qui porte sur le manque de transparence dans l'appréciation et l'utilisation des résultats (manque de précision-justesse dans la formulation des réponses aux questions d'évaluation).

Extraits de verbatim

"I think the readers, more and more they want to know exactly what you did to arrive those patterns […] particularly in the context of using multiple sources of qualitative information like interview notes, documents, literature. And it goes back to what I alluded before about being more rigours in terms of documenting our procedures. More and more when we're writing a request for proposals, we're being asked to outline what our data analysis plan is, because often the word analysis gets thrown around a lot and there's not necessarily very much thought about what that means exactly. And I think you know when people use the word analyze there is much more explicit thought about what that process actually refers to, like what are the nuts and bolts of analyzing something? What about the process in which you procure greater insight to your evaluation question by using data, results from multiple line of evidence? How do you go about it?" (J3).

"You don't want to use terms like we did an analysis, and this was the result. Everyone would say well what analysis did you do? You can't just say I analyzed the data and then give me your conclusion. Like there would be very specific requirements depending on the type of analysis you are doing, and how you would report the results. Also, there should be a systematic approach. […]. I think the natural assumption would be if you don't specify, I am going to assume you did the least rigorous thing right. […]." (J2).

"Barely any criticism it's pretty rare for a program like that. Also, when it's presented like that, I don't know, I am not sure all the evidence is here. Which then, you just want to ask: what about the rest of the information? Which also leads me to question how the analysis was conducted? Which they also didn't explain. Like they did all these key informant interviews. Did they just look, and they decide oh this is how…like was it systematic? We don't know." (J4).

La remise en question de la fiabilité face au manque de transparence du devis d'évaluation (4 uds)

Pour 'J1', le contenu du devis d'une évaluation, plus particulièrement de son cadre méthodologique, doit être décrit et documenter de manières suffisamment précises pour que le lecteur du rapport d'évaluation soit en mesure de critiquer la méthodologie évaluative, de la répliquer ou de l'utiliser pour soutenir la réalisation d'une autre évaluation de programme. Pour 'J2', il devrait être en mesure d'identifier et de comprendre clairement les procédures de collectes, et plus particulièrement des procédures d'analyses de données qualitatives, tout en ayant l'assurance du contrôle des biais. Or, pour ces évaluateurs novices, ici ce n'est pas le cas, et le manque de transparence au niveau du cadre méthodologique soulève donc des problèmes de

fiabilité de l'évaluation. En outre, veuillez noter que si le concept de « fiabilité » n'est pas explicitement mentionné par les évaluateurs juniors, c'est ce dont il est question à la lumière de leurs propos. 'J2' remet en question la possibilité de répliquer les procédures d'analyse qualitative des données d'entrevues, et nous l'avons vu plus haut, fait preuve de fortes préoccupations en ce qui a trait à la rigueur des procédures d'analyse de contenu, et de la gestion des biais. En ce qui a trait à 'J1', outre l'impossibilité de répliquer le devis d'évaluation à la lumière du peu de détails fournit au sein du cadre méthodologique rapporté, l'évaluateur junior fait preuve de préoccupations quant à l'impossibilité de connecter les différents éléments du devis d'évaluation et d'apprécier de quelle manière les évaluateurs en sont arrivés aux résultats d'évaluation communiqués dans le rapport.

Extraits de verbatim

"In an evaluation you need to be able to speak to both audiences, someone who just wants just to read the executive summary and get some information and someone who wants more, who can critique it or using to inform another evaluation, or future evaluation and with this one they would have a lot of questions about the methodology. […]. You should be able to take the guides, to take whatever in the report that gives an understanding of how they approach it and how they came to these findings. Because that's what is important here. You should be able to replicate the evaluation using the report." (J1).

"So yeah, in terms of procedures, was it one person that analyzed the themes and then coded those themes? Was it, was there really like sort of consideration of how bias might have come into that? Did you have multiple people looking at the coding scheme? [...] And when they say analyses, you know it makes me think about what they mean by analyzing. There are a variety of ways you could do that you know down to like coding at the sentence level and concept level. Using different types of software, different themes. So again, just telling you that you analyze looking at the research theme doesn't tell me what you did. Right. So that was a little confusing. [...] There's transparency in replicability and what I'm doing. I think that's what I'm sort of driving at with the interview reports saying well we just analyzed themes… and well your process is not transparent, and I don't know that I can replicate it. We have no guarantee that there is no bias or that if we tried to do it again, we will get the same result." (J2).

Enfin, remarquons que l'ensemble des préoccupations des évaluateurs juniors que nous avons pu observer face au manque d'explicitation composantes du cadre méthodologique, et la remise en question de la rigueur méthodologique face à un tel manque de transparence, soulignent un potentiel manque de pertinence et (ou) de précision-justesse susceptible d'engendrer des problèmes de fiabilité, de validité et de représentativité.

La remise en question de la validité et de la représentativité face au manque de transparence et de pertinence/précision-justesse du contenu du devis d'évaluation (9 uds)

Face au manque de description des objectifs des différentes méthodes de collectes de données, à l'absence de la matrice d'évaluation contenant les indicateurs et l'absence de l'instrumentation, 'J1' explique avoir du mal à cerner la contribution de chacune des sources de

données (méthodes-sources d'information) et à identifier l'effort de triangulation, et se demande dans quelle mesure le devis d'évaluation a permis le recueil d'informations (ou de preuves) qui représentent avec précision et justesse le programme évalué. Ici, 'J1' parle explicitement de l'enjeu de la représentativité, une préoccupation qui, nous verrons plus bas, se manifeste à nouveau dans la discussion sur les enjeux de validité et de représentativité des conclusions évaluatives face au manque de transparence sur le plan de l'interprétation des résultats. Par ailleurs, son regret de ne pas être en mesure d'examiner la pertinence-justesse du contenu des guides d'entrevue, touche à la vérification du lien de l'alignement « questions d'évaluation - indicateurs - instrumentation », une préoccupation liée également à la validité et par extension la fiabilité.

Dans le cas de 'J3', s'il ne fait aucunement mention de préoccupations quant à la démonstration de l'effort de triangulation, ou la représentativité, il s'interroge tout de même quant à l'alignement « question d'évaluation - sources de données ». Un élément qu'un évaluateur s'attend à pouvoir observer, et qu'un client devrait vouloir connaitre. On revient également sur la perception du manque d'effort de rigueur dans la démonstration, et la planification de la méthodologie évaluative, renforcer par l'absence de la matrice évaluative.

Extraits de verbatim

"I am an evaluator, and it would be nice to see the specific indicators they use. For example, the literature review what questions where they answer? Like you know they have three or four lines of evidence, so then I could see if they had a good representation for each of them. They had a lot of lines of evidence but did each line of evidence actually informed the questions. But I don't have any problems with the question, you just kind of wonder about the fidelity. The triangulation process wasn't always clear. [...]. Also for the key informants, how many interviews did you do? Who did you speak to at these places? We only have a broad list here, did they speak to the director or the manager, did they speak to multiple people? Did they speak to the program users? I can't see if I have a good representation here. Also, I don't even have the interview guides. Usually when you look at a report you look at the guides and then you critique. Like did you ask the right questions?" (J1)

"I think it's important for me as someone who prepares evaluations, and I should think it would be important to people who are finding the evaluations to understand why you're using the lines of evidence that you're using, what questions you're hoping to address it wasn't clear. You know they don't talk about why they chose these lines of evidence and what specific questions they had in mind. And I think there is a relationship here. [...]. I find that having evaluation matrix is important because it's more rigorous. There is an organization to it you can sort of see how everything lines up. And so, in this case it's a little more difficult and it's possible that if I read through the entire report word for word it could become clearer what line of evidence apply to each question. I can speak to the parts I read, the executive summary and the methodology, and I didn't see that discuss. They may have used like us, like an evaluation crosswalk to plan for each evaluation question the lines of evidence they were going to use to answer that question. But you can't see that here." (J3)

Les évaluateurs juniors 'J1', 'J3' et 'J4' se sont également demandé dans quelle mesure l'évaluation pouvait rendre compte ou démontrait l'efficacité de l'implantation du programme.

Ainsi, si 'J3' ne l'a pas mentionné dans son analyse critique, lorsqu'interrogé sur les questions d'évaluation, ce dernier semble avoir remarqué un potentiel problème d'alignement « fonction - objets/axes - questions d'évaluation et données disponibles ». En effet, pour lui l'intention des évaluateurs de se concentrer sur la conception, l'implantation, et les résultats primaires de la mise en œuvre du programme sous-entendait une évaluation formative. Entendu qu'il nous avait mentionné un plus tôt que l'évaluation du rendement (efficacité, efficience et économie) se prêtait davantage à l'évaluation sommative, il ne désire pas pour autant remettre ici en question la présence de questions de rendement. Il remarque en revanche que cela nécessite l'accès à un échantillon important de données robustes, une chose que le programme ne semblait pas prêt à fournir. Dans la lignée de cette observation, 'J1' explique que l'échantillon restreint de la documentation sur le programme le pousse à s'interroger quant à la force des conclusions, ces derniers ne semblant pas avoir eu accès à suffisamment de données de monitorage du programme. Enfin, 'J4' explique ne pas être en mesure de pleinement saisir s'il est question d'une démonstration théorique ou empirique de l'efficacité.

Extraits de verbatim

"I would say it is a formative evaluation because it says will largely focus on the status of implementation of the program, particularly its design and delivery, and the initial results of the implantation. [R - So what does that mean for the performance questions?] - I am not going to say. Well, I am just going to say this, it's not necessarily... I think it's not inappropriate for them to have questions about outcomes or performance, effectiveness and so on. But you do need to have robust evidence, and a lot of them to answer these questions and they may not be there yet." (J3).

"*I like the monitoring effectiveness questions. You know as an evaluator. But... ya... They did say that they didn't get a lot of documentation too from the sites. So, I don't know how strong the finding would be because they did not have enough documentation, like actual monitoring data*." (J1)

"Sometimes it was hard to discriminate what was coming from the literature and what was coming from the program. It wasn't clear for me if it was theory or if it was proven to be working in other contexts, etc. That wasn't clear. They answer the question but I'm not sure the findings were sound. They presented the approach, but I didn't know if it was just theory or if it's actually sound based, like if they actually went to see if it works."

Enfin, les différents écueils observés par les évaluateurs juniors en ce qui a trait aux sources d'information et procédures d'échantillonnage ont également soulevé des problèmes de validité et de représentativité. Aussi, nous faisons référence ici :

- *Au manque de pertinence des sources d'information sélectionnées*
    - Non-inclusion d'une perspective comparative sur le plan de critères de sélection des sources d'information dans le cadre de l'analyse documentaire ('J2') et des entrevues d'informateurs clés ('J1').

- Absence du recueil de la perspective du personnel de première ligne, de la clientèle (‘J1’) ou de la population ciblée du programme (‘J4’).

- *Au manque de description des sources d’information sélectionnées*
  - Absence de la liste de références de la documentation analysée et des écrits recensés (‘J2’ et ‘J3’).
  - Absence de description du type de document sélectionné (‘J1’).
  - Absence de description des caractéristiques des individus ou groupes d’individus interviewés et de la répartition des entrevues (‘J1).
- *Et à la remise en question de la pertinence et de la précision-justesse dans la sélection des sources d’information face au manque d’explicitation des procédures d’échantillonnage*
  - Absence de démonstration d’une procédure systématique dans la sélection des écrits pour la recension (‘J2’ et ‘J3’).

b. Les préoccupations liées au manque de transparence sur le plan de la synthèse et de l’interprétation des résultats

Ici, les évaluateurs juniors ‘J1’ et ‘J4’ s’attardent sur le manque de transparence, de précision-justesse sur le plan de l’appréciation des résultats, un constat qui les a conduits à s’interroger sur certains éléments relatifs à la fiabilité et la représentativité de l’agrégation des preuves et de l’interprétation des résultats. Par exemple, dans le cadre de l’évaluation de l’efficacité du programme :

- Dans quelle mesure les résultats mobilisés sont-ils bel et bien représentatifs du programme ?
- Dans quelle mesure répond-on vraiment aux questions d’évaluation ? Est-ce que les résultats mobilisés sont fondés sur des données probantes ? D’où proviennent les résultats ?
- Au regard du caractère, peut-être trop positif des résultats évaluatifs, est-ce que l’ensemble des preuves a bel et bien été utilisé ? Dans quelle mesure le poids accordé à certains résultats (et la manière de les présenter) représente-t-il avec justesse le programme évalué ? Quelles sont les motivations derrière ces choix ?
- Etc.

Extraits de verbatim

"Okay one place did that. But how many other places completely change things or tailored their program? The key informant said this but how many people said it? What's the quantity? I think that's really important because again when you're writing reports it's pretty easy to change things to make it sound like that has more impact or more weight than it does that's why. I like to know how many people you did interview and what percentage of them actually said this. I think it's very important to inform this. That was something I definitely notice that was lacking in the findings." (J1)

"They didn't source some things. […]. Like for example page 21 - this evidence suggests this agency X implemented more preventive measure to reduce the number of X. Where did they get that from? What evidence? Was that from key informant interviews? From case a study? From document reviews? Which lines of evidence did they used? You need to be more aware adding more details. They did source some things. I always notice you know when there is only one line of evidence, I'm like oh oh." (J1)

"It's the evaluator's job to synthesize all this information and make a judgment of the program but I felt like I was left to make the judgment myself. I wasn't clear for me if it was theory or if it was proven to be working in other contexts, etc. That wasn't clear. They answer the question but I'm not sure the findings were sound. They presented the approach, but I didn't know if it was just theory or if it's actually sound based, like if they actually went to see if it works." (J4).

"But yeah, barely any criticism it's pretty rare for program like that. Also, when it's presented like 42% of the agencies did… as the main result to answer the question, I don't know, I am not sure all the evidence is here. Which then you just want to ask: what about the rest of the information? Also, when it's presented like that, I don't know, I am not sure all the evidence is here. Which then you just want to ask: what about the rest of the information?" (J4)

Enfin, si le manque de transparence sur le plan de l'appréciation des résultats remet en question, pour 'J1' et 'J4', la force des conclusions évaluatives, il soulève également un potentiel problème de mérepréssentation et de mésusage des résultats mobilisés. En effet, pour 'J1' et 'J4', lorsque des résultats tels que « *40 % - 50 % des agences ont mis en place avec succès l'activité X du programme* » constituent la preuve principale pour démontrer l'efficacité d'une des dimensions du programme évalué, et que les informations relatives aux 60 - 50 % restants ne sont pas communiquées, cela limite la confiance des lecteurs envers les résultats-conclusions présentés.

Extraits de verbatim

"Maybe it's because they don't have enough information. Like the key informant said that you don't associate it with a percentage or quantity it. It seems like the evidence wasn't strong enough or... that's what I'm going to assume, I wouldn't take that as a strong finding myself." (J1).

"*For the monitoring questions, implementation, and effectiveness, I wonder... It seems like they didn't get a lot of administrative documents from the sites about program implementation. So, I don't know how strong the findings are... or if there is actual evidence."* (J1)

"I thought the evidence was not really convincing throughout the report. For example, page X, this aspect of the relevance question regarding the quality of the relationships between agencies, like the analysis here I am not sure. Like 42% is not even the majority of the agencies so how is that being used to support your findings that there is a strong relationship, engaged directors and all of that. And you see things like that throughout the report." (J4).

"The use of stats like 40% or 50% of agencies are doing this to support the ultimate finding, answer the question but you don't even talk about what the rest of the evidence shows. So, to me the findings are not necessarily strong. I don't think I have a clear idea of what the evidence actually shows. I have no idea. What the evaluator or the person who finalized the report wants us to see. Like why you don't tell me what the other 50% was about, you don't inspire me with confidence in the findings." (J4).

4.1.1.4 Les facteurs contextuels exerçant une influence potentielle sur les éléments clés du devis d'évaluation

Au fil de l'activité d'analyse critique, différents éléments contextuels ont émergé comme étant susceptibles d'influencer les décisions prises par les évaluateurs dont l'évaluation faisait l'objet de l'activité métaévaluative et, ou, sur le jugement critique même des évaluateurs novices participants à la recherche. On dénombre au total *50 cooccurrences (cos)* entre les éléments clés du devis d'évaluation et divers éléments contextuels constitutifs de trois groupes de facteurs contextuels. Aussi, les fréquences desdites cooccurrences sont distribuées de la manière suivante :

- Dyade [contraintes politiques - devis d'évaluation] : 66 % (33 cos).
- Dyade [intrants disponibles - devis d'évaluation] : 22 % (11 cos).
- Dyade [sensibilité culturelle du milieu - devis d'évaluation] : 12 % (6 cos).

La Figure 22 (p. 163) montre, de manière générale, la répartition des cooccurrences dans le cadre des interactions entre chaque groupe de facteurs contextuels et les éléments clés du devis d'évaluation. Ainsi, on peut voir que les **contraintes politiques** ont été les plus mises en exergues dans le cadre de la discussion du contenu de l'ensemble des éléments clés du devis d'évaluation. Aussi, ces dernières semblent tenir une part plus importante en ce qui a trait aux *objectifs, types et questions d'évaluation* (respectivement 6 cos, 5 cos et 5 cos), aux *méthodes de collectes de données et instrumentation* (7 cos) ainsi qu'aux *limites méthodologiques* identifiées par les évaluateurs juniors (5 cos). En ce qui a trait aux **intrants disponibles pour l'évaluation**, ceux-ci ont été principalement pris en considération lorsqu'il était question du cadre méthodologique, et plus particulièrement, des *méthodes de collectes de données et instrumentation* (6 cos). Quant à la **sensibilité culturelle du milieu**, celle-ci touche majoritairement les *méthodes de collectes de données et instrumentation* (4 cos). Enfin, le choix des méthodes de collectes de données et les procédures de collectes de données sont les éléments du devis d'évaluation qui interagissent le plus avec les facteurs contextuels.

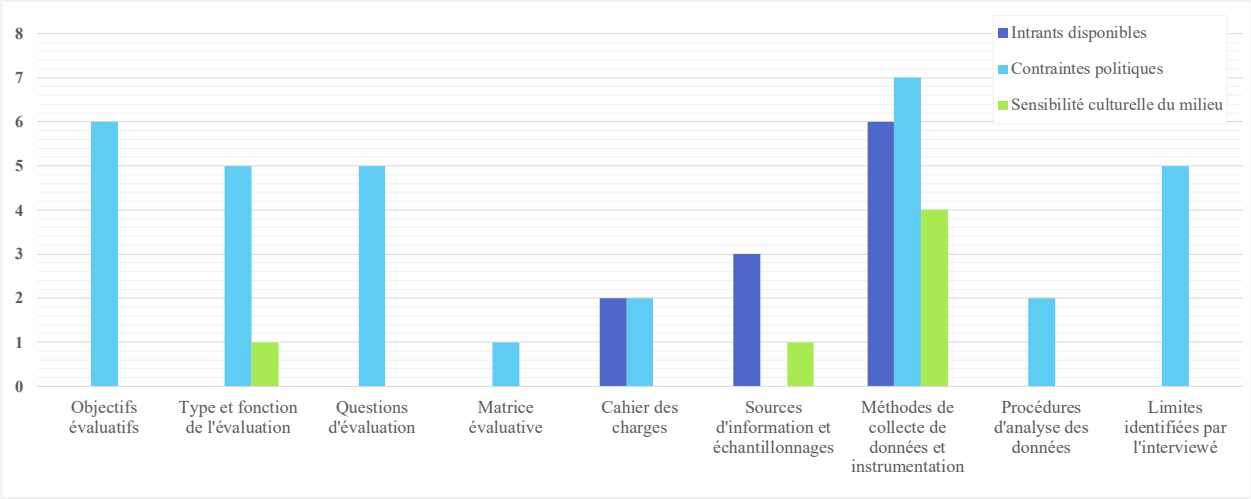


Figure 22. *Distribution des fréquences de cooccurrences dans le cadre des interactions entre chaque groupe de facteurs contextuels et les éléments clés du devis d'évaluation*

Afin d'analyser plus en profondeur les cooccurrences entre les différents éléments constitutifs de chacun des groupes de facteurs contextuels et les éléments clés du devis d'évaluation, nous avons conçu la Figure 23 (p. 164). Cette figure rend compte des interactions au sein de chacune des dyades introduites précédemment. Et, nous abordons en détail ces interactions à la suite du graphique, à savoir le contenu de :

a. la dyade [contraintes politiques - devis d'évaluation] (33 cos) - principalement l'influence des *standards de l'évaluation et de la politique évaluative du gouvernement fédéral* sur le mandat d'évaluation (objectifs évaluatifs [6 cos], fonctions de l'évaluation [5 cos], et questions d'évaluation [5 cos]) ; et du *comité d'évaluation interne* sur les méthodes de collectes de données (4 cos) et les limites méthodologiques (3 cos).
b. La dyade [contraintes politiques - devis d'évaluation] (11 cos) - essentiellement l'influence des *contraintes budgétaires et temporelles* sur le cahier des charges (4 cos), les sources d'information (4 cos) et les sources de données (3 cos) ;
c. La dyade [sensibilité culturelle du milieu - devis d'évaluation] (6 cos) - surtout l'influence du *caractère culturel sensible du milieu, soit le contexte autochtone*, sur les méthodes et procédures de collectes de données (5 cos) et les limites méthodologiques identifiées par les évaluateurs juniors participants (5 cos).

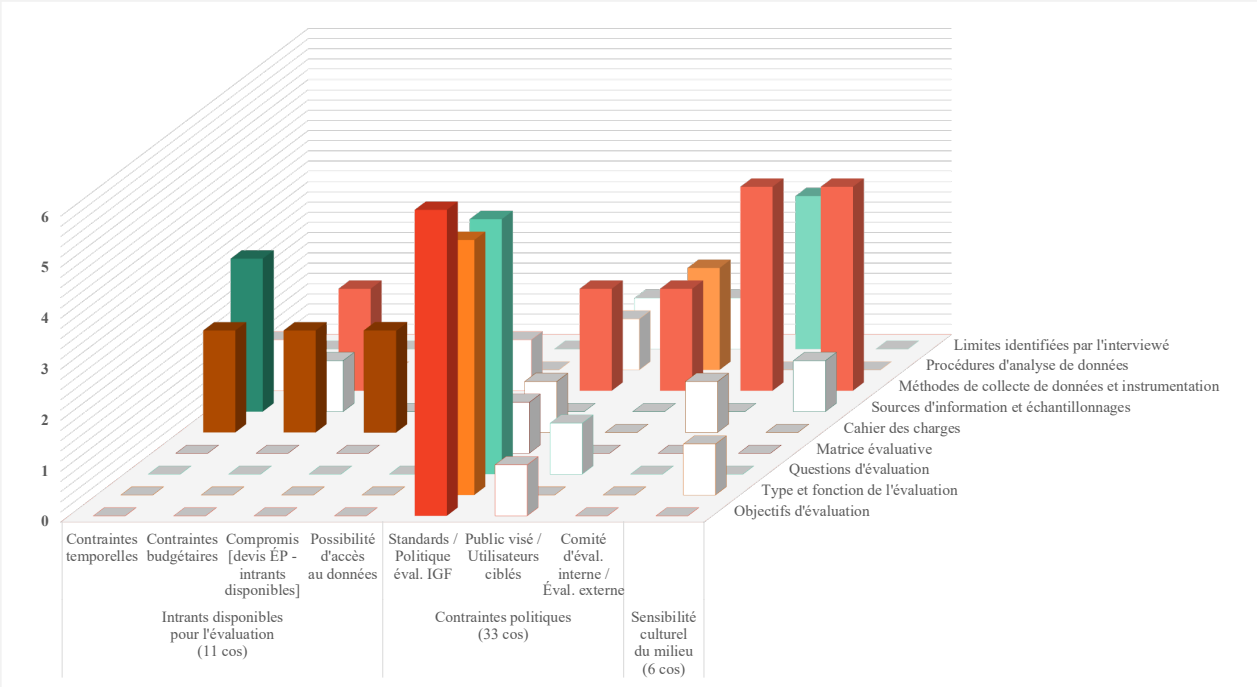


Figure 23. *Fréquences de cooccurrence entre les éléments constitutifs de chacune des dyades [devis d'évaluation - groupe de facteurs contextuels]*

a. La dyade [contraintes politiques - devis d'évaluation] (33 cos)

Parmi les contraintes politiques identifiées dans le discours des évaluateurs juniors, ce qu'ils qualifient de *standard de l'évaluation au fédéral* constitue le facteur le plus influent, plus particulièrement au niveau du contenu du cadre d'évaluation, qu'ils prescrivent. Il est question ici de la *Politique d'évaluation de l'IGF*[24] *(18 cos)*, politique que seul 'J3' a explicitement évoquée. Aussi, selon nous, la valeur normative des standards de l'évaluation au fédéral perçue par les évaluateurs juniors expliquerait notamment pourquoi ces derniers n'ont pas spontanément critiqué le contenu du cadre d'évaluation (hormis lors de la mention de l'absence des matrices, perçue d'ailleurs comme une pratique standard au fédéral par 'J2' et 'J3'). En outre, c'est lorsque nous les avons interrogés sur les questions d'évaluation que ces derniers ont utilisé comme critère d'analyse leur caractère standard au regard des attentes du gouvernement fédéral (5 cos).

[24] L'acronyme *IGF* fait ici référence à l'institution du gouvernement fédéral responsable, entre autres, de la politique d'évaluation du gouvernement fédéral et de l'assurance qualité des projets d'évaluation de programmes fédéraux.

Pour ‘J2’, s’il les trouve plutôt *jargonneuses*, ces dernières illustrent typiquement ce que les agences gouvernementales désirent savoir. ‘J4’ souligne qu’elles sont *centrées sur les attentes du gouvernement fédéral* et typiquement les mêmes pour toutes évaluations de programmes de ce type, et qu’en dépit de leur forte densité, elles donnent une bonne idée du contenu du programme. Il remet cependant en question le poids des standards de l’évaluation au fédéral, déclarant que le gouvernement fédéral est trop strict en ce qui a trait aux critères qu’il recherche, et les questions qu’il veut que les évaluateurs examinent. Nous verrons un peu plus bas que pour ‘J4’, cette rigidité provoque des occasions manquées lors de l’évaluation d’un programme au sein de communautés autochtones, évaluation qui mériterait le recours à une approche plus participative, et plus holistique. Pour ‘J1’, les questions évaluatives sont adéquates puisqu’elles couvrent la pertinence, la performance, l’efficience, etc., soit ce qui est attendu au gouvernement fédéral. En outre, s’il semble apprécier le respect du cadre d’évaluation du gouvernement fédéral, il demeure cependant sceptique face à certaines questions qui semblent avoir été incluses de manière automatique, avec une formulation générique qui ne fonctionne pas nécessairement au regard des caractéristiques du programme évalué, et pour lesquelles il a du mal à envisager des indicateurs de mesure. Enfin, en ce qui concerne ‘J3’, ses propos rapportés ci-dessous illustrent assez bien le rôle prescriptif de la Politique d’évaluation de l’IGF.

Extraits de verbatim

“The questions were fairly standard. It is pretty common to look at relevance and design delivery. These are the broad categories: relevance, design, delivery, performance and effectiveness, and economy and efficiency. In Federal evaluations there is a policy on evaluation, there is a policy that the IGF has, relating to evaluation and it generally fashions the types of question that are being asked, it guides what's being asked. [..]. And so, a formative evaluation would be more inclined to focus on design and delivery whereas in summative you would be more likely to focus on performance, economy and efficiency.” (J3).

Au sein des agences gouvernementales canadiennes, les évaluations de programme sont commanditées par un *comité d’évaluation interne*, comité qui définit initialement le cahier des charges de l’évaluation (incluant la méthodologie) dans l’appel de proposition, et ce, à partir des lignes directrices de la Politique d’évaluation de l’IGF. Ce comité, à titre de partie prenante primaire, est le principal collaborateur de l’équipe d’évaluation externe, et il semble détenir un pouvoir décisionnel important tout au long de l’évaluation. Aussi, il constitue le second facteur politique ayant une incidence importante sur les choix méthodologiques de l’équipe d’évaluation externe et la communication de l’évaluation (8 cos). Toutefois, ‘J3’ et ‘J4’ perçoivent ce comité

d'évaluation interne de manière bien distincte. Si pour 'J3', ce comité joue le rôle d'auditeur de qualité, d'agent de vérification de la démarche d'évaluation ; il représente en revanche, selon 'J4' une potentielle source de biais et, ou, de manipulation des résultats-conclusions de l'évaluation.

Extraits de verbatim

"There is this multi-stakeholder group (research evaluation and assessment group) that is watching what's going on and making sure that the approach is reasonable and in fact you would see that even in some non-federal evaluations. […]. It provides an extra layer of quality assurance. […]. It's quite common to have working groups with various individuals and various roles. And we will present our findings to them, and they may say well you know where this finding comes from? Or maybe we should be asking this. And in so doing, it's just an added layer to ensure that due diligence is being done." (J3).

"*For example, page X, this aspect of the relevance question regarding the quality of the relationships between agencies, like the analysis here I am not sure. Like 42% is not even the majority of the agencies so how is that being used to support your findings that there is a strong relationship, engaged directors and all of that. And you see things like that throughout the report. Also, I have questions for some of the other findings… There are very few areas of criticism.* [R - Why do you think…] - *I don't know why or maybe yes but… maybe there is some hidden objectives, or maybe it was internally? You know when they hire us as consultants, they sometime scrap our names from the report so… But yeah barely any criticism is pretty rare for a program like that. Also, when it's presented like that, I don't know, I am not sure all the evidence is here. Which then you just want to ask: what about the rest of the information?"* (J4)

En outre, si 'J3' et 'J4' ne sont pas en mesure de se prononcer sur l'étendue du degré de négociation entre les commanditaires et l'évaluateur sénior en charge de l'évaluation externe, leurs expériences passées les amènent à reconnaitre le potentiel manque de flexibilité du comité d'évaluation interne en ce qui a trait aux éléments clés du devis d'évaluation et la possibilité de mobiliser des approches évaluatives plus novatrices. En effet, selon 'J3', si les appels de proposition invitent parfois les évaluateurs de programme à sortir des sentiers battus, très souvent (voir même plus souvent), que ce soit par habitude ou par assurance (ils préfèrent s'en tenir à ce qu'ils connaissent), par idéologie, ou par respect des standards de l'évaluation au fédéral ou du contenu du devis d'évaluation prescrit par l'IGF, le comité d'évaluation interne y prescrit les éléments clés du cahier des charges de l'évaluation commanditée. Enfin, pour 'J4', une telle rigidité au niveau des procédures va à l'encontre des préoccupations relatives à la sensibilité culturelle d'un programme en milieu autochtone. Ici, il fait particulièrement référence à son observation du manque d'implication des parties prenantes clés des différentes communautés autochtones dans le comité d'évaluation interne, dans la conception de l'appel de proposition, dans la conception du cahier des charges avec les évaluateurs externes, etc.

Extraits de verbatim

"Sometimes the client they want to stick with what they had in mind. And they're not open to other ideas will sometimes happen but not always. You know sometimes they explicitly ask the people who are applying, the various contractors, to put forward their own ideas about what might be done. […]. But then again, the level of negotiation or decision, I judge from the proposals for the projects I've worked on in the past where they will often say you know at minimum evaluations should involve the various lines of evidence. You know there will be many that many interviews, etc. or whatever. However, the submitter may also consider other lines of evidence." (J3).

"I don't know, it seems like they could have had the opportunity to use a more participative approach, especially for this type of program where they talk about cultural sensitivity and all that. Opportunity to involve beneficiaries or people from the community in the process. I know they talk a bit about the governance, and they mention the evaluation communities where most of the people are from the department. I understand that maybe it's the federal government, it's not that flexible so they couldn't be more innovative." (J4).

Le public visé par la communication de l'évaluation et les utilisateurs ciblés de l'évaluation constituent un autre élément factoriel de la dyade [contraintes politiques - devis d'évaluation] exerçant une influence sur les décisions prises en ce a trait aux informations méthodologiques fournies, et la pertinence du mandat d'évaluation et des collaborateurs clés de l'équipe d'évaluative externe.

Nous avons vu plus tôt que l'un des points critiques centraux de l'analyse des évaluateurs juniors concernait le manque de précision vis-à-vis des informations fournies en ce qui a trait aux éléments clés du cadre méthodologique (sources d'information, échantillonnage, choix des méthodes, procédures de collectes et d'analyse de données, etc.). Aussi, lorsque nous leur avons demandé pourquoi, selon eux, il manquait autant d'informations dans la section méthodologique du rapport d'évaluation, 'J1' et 'J3' ont mentionné que cela dépendait du *public visé (5 cos)*. En effet, ces derniers semblent émettre l'hypothèse que le niveau de détails est jaugé selon le public visé par le rapport (c.-à-d. du contenu le plus basic pour le public général au contenu le plus riche et précis pour les commanditaires de l'évaluation). En outre, si le rapport final publié en ligne est le même pour tout le monde, J1 et J3 espèrent qu'il existe un rapport beaucoup plus complet, ou un rapport technique détaillé de la méthodologie.

Extraits de verbatim

"Again, it could be about the audience they are talking to and they want to keep it basic. But you usually I hope you would have another report or do a presentation or something." (J1)

"It depends on what you choose to report to the public and what you choose to report to the person who is funding and administering the research, it might be something different. So maybe they thought in the report that is published online, that level of detail wasn't necessary, that it is possible. I certainly do hope they prepare a separate methodology report. It sounds like they did that they actually you know went just sort of this detail." (J3)

En ce qui a trait aux commanditaires de l'évaluation (comité d'évaluation interne), 'J3' explique que de plus en plus ces derniers désirent une communication claire et précise des procédures méthodologiques et des processus ayant conduit aux résultats sur lesquels reposent les conclusions évaluatives. Un tel souci du détail exigé sur le plan de l'explicitation des procédures méthodologiques force ici les évaluateurs à être plus rigoureux dans leurs pratiques. La communication de l'évaluation aux utilisateurs ciblés semble constituer ici une forme d'assurance qualité des résultats obtenus.

Extrait de verbatim

"I think the reader, more and more they want to know exactly what you did to arrive at the patterns that you actually got. And this is something that he's been struggling with. It's something that is newer. Particularly in the context of qualitative information, more and more when we're writing a request for proposals, we're being asked to outline what our data analysis plan is because the word analysis gets thrown around a lot and there's not necessarily very much thought about exactly what that means. They want to know, to understand what the nuts and bolts are of analyzing something, summarizing it and compiling it with other lines of evidence in order to produce answers and greater insights." (J3)

Enfin, pour 'J4', il ne s'agit pas tant du public visé par la communication de l'évaluation, mais plutôt l'identification des *utilisateurs ciblés de l'évaluation*. En effet, pour lui, l'appréciation de la pertinence du mandat d'évaluation et de la composition du groupe de parties prenantes clés impliqué dans l'évaluation se fait à travers le prisme de la raison d'être d'une évaluation et des utilisateurs ciblés de cette dernière. Pour lui, il s'avère donc primordial d'être en mesure d'identifier les parties prenantes clés membres du comité d'évaluation interne, et ce, en vue de s'assurer que chacun des utilisateurs ciblés est représenté, ou du moins consulté.

Extraits de verbatim

"I've participated in federal evaluation, and I know the users. Here they didn't talk about who the users of the evaluation were. I did see management was there. It's just usually you mention the people who are going to use the findings, etc. It's not really touched upon." (J4)

"What I am more interested in finding out who the users of the evaluation are and are these questions relevant to them. For what purpose? Is it mostly for internal purposes? Then maybe these are great. If they are looking to learn from the evaluation and improve themselves then they have an opportunity to give their input in designing these questions. Probably not. I just find that they are opportunities for them to be more inclusive not in the design but. They did provide their working group. Well, it includes most of the internal people." (J4)

a. La dyade [intrants disponibles - devis d'évaluation] (9 cos)

Pour 'J2', les différents choix ou limites méthodologiques qu'il a pu souligner, notamment en ce qui a trait au spectre restreint de l'échantillon dans le cadre de l'analyse documentaire, sont *tributaires des exigences (ou besoins) des clients en termes de contraintes budgétaires et*

temporelles. Il mentionne d'ailleurs à plusieurs reprises qu'il est difficile de juger les choix méthodologiques des évaluateurs sans connaitre les contraintes relatives aux intrants disponibles pour l'évaluation.

Extrait de verbatim

"You basically agreed in your contract to say like we're going to tell you by this date what's happening right, so certain information sources you may not even be able to consider because you just don't have time to reach everyone or go deeper in the document review. (…). There are maybe methodological choices that they made that were predicated on the fact that they didn't have time to do anything better or their client didn't want to spend the money to do anything better. […]. An evaluation is designed to meet a set of clients' needs in terms of the timeframe that it needs to be completed in and the amount of money that they have available to spend on it. To evaluate the quality of an evaluation (…). I'm just saying that if we don't know how much time or money, they had to spend then you know it's very easy to say things that might have been better but maybe this is still the best choice given that they didn't have the money to do that." (J2)

En d'autres termes, il est question d'un *compromis entre le cahier des charges et la répartition efficace des ressources pour l'évaluation.*

Extraits de verbatim

"It's everything's a cost-benefit trade-off and I think that's a really important decision when you're doing research here. Like I could get better answers than we get for certain evaluations. But could we do it in a way that's timely and efficient and practical in terms of the needs of the organization? If it takes me six years to produce a report that you know fully understands the issue. Well, that doesn't help them plan for the next five years right." (J2)

"The other theory is probably about funds. Probably some of that as well in the sense that you can do maybe something that's more exotic, but it will cost more money. It's better to stick with what's tried and true. So that's what I would say. I do want to emphasize that although it is common to rely on the same method over and over again, we do have cases where some innovative approaches are used. It does happen from time to time." (J3)

b. La dyade [sensibilité culturelle du milieu - devis d'évaluation] (6 cos)

Le lien entre la sensibilité culturelle du milieu, ici un programme qui touche la communauté autochtone, et la démarche d'évaluation est abordé de manière différente.

Pour 'J1', il est question de la composition de l'équipe évaluative et la nécessité d'inclure un évaluateur qui comprend les problématiques en jeu, un évaluateur autochtone. Il aborde également l'importance de la prise en compte de la sensibilité culturelle dans l'approche des entrevues et le contenu ainsi que le langage des énoncés des guides de discussion. D'où, pour lui, l'importance d'inclure les instruments dans le rapport. 'J2' quant à lui effleure le sujet lorsqu'il souligne l'importance de recourir aux entrevues comme méthode de collecte de données, permettant ainsi la prise en compte de la communauté, d'inclure la voix des parties prenantes communautaires au sein de populations isolées. Enfin, pour 'J4' la prise en compte de la communauté autochtone devrait aller au-delà de leur statut de sources d'information dans le cadre

de collectes de données. Selon lui, les évaluateurs ont manqué une occasion de recourir à une approche plus participative, où des parties prenantes clés de la communauté autochtone auraient pu être impliquées dans le processus d'évaluation comme membre du comité d'évaluation, par exemple. En outre, il reconnait le manque de flexibilité du gouvernement fédéral sur le plan de la gouvernance des programmes et de leur évaluation, ainsi que la méthodologie évaluative.

Extraits de verbatim

"Well, that's the question I was talking about before about how your team is structured when you do an evaluation. For example, here we have someone who has a lot of experience working with aboriginal communities, who grew up in an aboriginal community, understand the issues. […]. *For an evaluation we have done in the past…In view it looked very much like here, we had questions similar to the evaluation questions, very straight to the point, and we didn't get any participation from the communities, so we completely changed the guide, went with a more holistic approach. The wording was softer. That's why I like to see the guides, to see how they actually did get to this question. […]. It needs to kind of reflect the cultural sensitivity, your language needs to resonate with them."* (J1)

"*Also interview methodology was… I mean I think interviews are absolutely essential in this case and there were some nice things here you know that they sort that they seemed to integrate the community into that and I think that can be really important in getting stakeholders to get on board and open up and particularly with the most isolated populations like this"* (J2)

"*I don't know, it seems like they could have had the opportunity to use a more participative approach, especially for this type of program where they talk about cultural sensitivity and all that. Opportunity to involve beneficiaries or people from the community in the process. I know they talk a bit about the governance, and they mention the evaluation committee where most of the people are from the department. I understand that maybe it's the federal government, it's not that flexible so they couldn't be more innovative."* (J4)

4.1.1.5 *Interprétation de la question de recherche tertiaire 1* – les pratiques dites recommandées des évaluateurs novices de programmes publics

Question de recherche tertiaire (1) : Lors de l'analyse critique d'une évaluation de programme tierce, quels sont les éléments mis en exergue par les évaluateurs de programmes publics relativement aux choix méthodologiques, et à l'influence des facteurs contextuels sur ceux-ci ?

Rappelons que dans le cadre de la consigne donnée aux participants pour cette activité d'analyse critique d'un devis d'évaluation, aucune indication n'a été donnée quant à la marche à suivre, les éléments à examiner, etc.

Les évaluateurs novices se sont principalement penchés sur la question de la « précision » du cadre méthodologique, plus particulièrement sur l'importance de la transparence et de la précision dans la description des choix et des procédures méthodologiques (sélection des méthodes et des sources d'information, procédures de collectes et d'analyses de données, plan de synthèse

des résultats, et limites méthodologiques). Aussi, s'ils débutent leur analyse critique au niveau des méthodes utilisées, ils font assez rapidement des liens avec le cadre d'évaluation, plus particulièrement la question de la démonstration du lien entre les questions d'évaluation et les différentes sources de données. Dès lors ils déplorent l'absence de la matrice d'évaluation, un outil qu'ils jugent indispensable en évaluation de programme, qu'il s'agisse de permettre au lecteur (le client) de comprendre comment les évaluateurs vont répondre aux différentes questions d'évaluation (lien questions-méthodes), d'assurer la planification rigoureuse de la méthodologie évaluative (rendre compte du lien « question d'évaluation - éléments à examiner - sources de données à utiliser »), ou encore démontrer/examiner l'effort de triangulation des méthodes et sources d'informations. Ils abordent ici un élément essentiel de la planification de la démarche d'évaluation (McDavid, Huse et Hawthorn, 2019 ; Secrétariat du Trésor du Québec, 2018 ; *Better Evaluation*, 2014)

Pour les évaluateurs novices, le devis d'évaluation rapporté doit rendre compte avec transparence, clarté, et précision de l'ensemble de la démarche évaluative. S'ils reconnaissent que dans le contexte de l'évaluation de programmes fédéraux, l'espace accordé à la description du cadre méthodologique dans le rapport d'évaluation est généralement restreint, cela ne devrait pas empêcher, selon eux, de fournir suffisamment d'information pour clairement rendre compte de la démarche méthodologique entreprise. De fait, de tels manquements semblent entraîner chez ces derniers une remise en question de la pertinence et (ou) de la précision-justesse de la méthodologie évaluative, et par conséquent, la perception d'un manque de rigueur, de procédures méthodologiques systématiques, voire de potentiels problèmes de fiabilité, validité et représentativité. Remarquons que les préoccupations méthodologiques observées chez les évaluateurs novices participants semblent émerger de différentes combinaisons de savoirs théoriques/procéduraux et de savoirs pratiques/expérientiels acquis dans le cadre de formations et (ou) sur le terrain, dans le domaine de l'évaluation de programme et (ou) des domaines connexes, plus précisément la recherche en sciences sociales. Il est question ici de ce que LeBoterf (2002) appel *l'équipement incorporé dans lequel le praticien puise pour agir avec compétence*, et selon les caractéristiques du profil de chacun de nos évaluateurs novices, les pratiques recommandées émergentes dans leurs discours en ce qui a trait aux éléments clés du devis d'évaluation sont plus ou moins riches. Ainsi, de façon générale, on observe que :

- **‘J2’** semble pallier un certain manque de savoirs dans le domaine de l’évaluation de programme en s’appuyant sur son profil de chercheur, mobilisant un cadre de référence centré sur les critères de scientificité de la recherche (reproductibilité du devis, considération des biais dans les analyses, etc.). S’il fit initialement preuve de résistance dans le cadre de l’activité d’analyse critique, évoquant à de multiples reprises son inexpérience en évaluation, il fut tout de même capable de faire part d’éléments pertinents, fondés sur la perception des pratiques de son cabinet de conseil, des standards de l’évaluation du gouvernement fédéral. Dans le Tableau 12 (p. 178-179), nous décrivons de manière détaillée les caractéristiques des pratiques recommandées de chacun évaluateurs novices participants, mais en résumé, **les préoccupations de ‘J2’ se concentrent particulièrement sur l’importance de :**

Profil de ‘J2’
Junior associé de recherche
- Firme [A]
- Formation [PhD SHS]
- Expérience pro. ÉP [-1 an/passation 1] [16 mois/passation 2]

  - Rendre clairement compte de la démarche méthodologique entreprise pour répondre à chaque question d’évaluation – recourir à la matrice d’évaluation un outil standard du gouvernement fédéral.
  - Examiner la pertinence du devis méthodologique à la lumière des ressources disponibles (échéancier, budget, accès aux données) et des exigences contractuelles.
  - Décrire les sources d’information et procédures d’échantillonnage, et rendre compte des contraintes existantes.
  - S’assurer de la pertinence des sources d’information sélectionnées (taille, étendue, critères d’échantillonnage) – inclure une perspective comparative lors l’évaluation d’un programme gouvernemental d’aide sociale avec des enjeux sociopolitiques et culturels complexes.
  - Expliciter clairement les procédures méthodologiques, plus particulièrement le traitement des données qualitatives – assurer la reproductibilité, démontrer l’effort de rigueur et la fiabilité des résultats.
  - Recourir à de multiples stratégies de contrôle des biais dans le cadre de l’analyse de données qualitatives – démarche rigoureuse et systématique, triangulation des évaluateurs.
  - Énoncer et décrire les limites méthodologiques.

- Dans le cadre de son analyse, **'J4'** fait part de ses attentes en tant qu'évaluateur de programme et des pratiques méthodologiques qui s'y rattache. Il s'est notamment attardé sur la section résultat du rapport d'évaluation, ce qu'il l'a conduit non seulement à questionner la justesse des conclusions évaluatives, mais aussi la pertinence du devis d'évaluation. Ce dernier est, par ailleurs, assez critique de l'environnement évaluatif dans le contexte du gouvernement fédéral et de son influence sur la pertinence et la précision-justesse de la démarche évaluative et sa communication (nous revenons sur ce point un peu plus loin). En somme, **les préoccupations de 'J4' se concentrent particulièrement sur l'importance de :**

Profil de 'J4'
Junior avancé
- Firme [B]
- Formation [M.P.P.A][DES ÉP]
- Expérience pro. ÉP [2 ans – 2 ans ½]

  - Rendre clairement compte de la démarche méthodologique entreprise pour répondre à chaque question d'évaluation – communiquer le plan d'évaluation/protocole méthodologique, les indicateurs (le rapport final devrait inclure la matrice évaluative et le modèle logique).
  - Promouvoir le recours à des approches innovantes lors de l'évaluation de programmes en contexte culturellement sensible, et ce, malgré le cadre rigide associé au contexte du gouvernement fédéral – p. ex. recourir à une approche évaluative de type participative en milieu autochtone.
  - Bien identifier les utilisateurs ciblés et veiller à ce que l'évaluation réponde à leurs besoins – dans une perspective d'utilisation des résultats à des fins d'amélioration, les impliquer dans la définition du cadre d'évaluation.
  - Décrire et justifier les procédures de collectes et d'analyse de données entreprises (en dépit de certains choix plus ou moins prescrits/standards de l'évaluation de programme au gouvernement fédéral).
  - Recourir à des procédures d'analyse de données qualitatives rigoureuses – rendre compte du caractère systématique de la démarche (justesse des interprétations).
  - Faire preuve de transparence dans l'appréciation des résultats – utiliser et présenter les résultats de manière appropriée (les présenter dans leur entièreté, s'assurer que les conclusions sont fondées sur des preuves tangibles), prononcer clairement le jugement évaluatif (représentativité des constats évaluatifs/justesse des interprétations).

- Formée en évaluation de programme sur le terrain, l'expérience professionnelle de **'J3'** semble exclusivement inscrite dans l'évaluation externe de programmes du gouvernement fédéral. Il est l'évaluateur junior avec le plus d'expérience dans le présent contexte, et ses expériences de collaborateur transparaisse dans la richesse de son analyse. Son cadre de référence s'inscrit dans les pratiques de sa firme-conseil et les standards de l'évaluation du gouvernement fédéral (de l'IGF). Il s'attarde sur la question du caractère complet du devis rapporté, ce dernier associant l'effort de transparence au renforcement de la rigueur méthodologique, et l'assurance de la confiance des clients en la qualité de l'évaluation. Ainsi, **les préoccupations de 'J3' se concentrent particulièrement sur l'importance de :**

Profil de 'J3'
Junior associé de recherche avancé
- Firme [A]
- Formation [M. Écon.]
- Expérience pro. ÉP
- [4 ans - 4 ans ½]

  - Décrire et documenter rigoureusement la démarche évaluative, les procédures de collecte et d'analyse de données (standard de l'IGF, assurance qualité) – communiquer le plan d'évaluation, démontrer le caractère systématique des procédures méthodologiques, s'assurer que l'on comprenne les bases du jugement évaluatif (cheminement entre les séries de données collectées et les constatations évaluatives émises, particulièrement le traitement et la triangulation des données qualitatives). Il est question du partage des outils méthodologiques (matrice éval., protocoles de collecte/analyse de données, instrumentation, procédure de synthèse/interprétation des résultats).
  - Veiller à la pertinence du mandat d'évaluation – définir les questions d'évaluation à partir des directives de l'IGF, s'assurer de l'alignement « fonction - axe/objet éval. - questions - ampleur des données de prog. disponibles/maturité du prog. ».
  - Recourir à l'« *evaluation crosswalk* » (matrice éval. de son cabinet conseil) pour planifier rigoureusement la méthodologie – démontrer le lien « question traitée - éléments à examiner - sources de données à utiliser ».
  - Décrire les procédures d'échantillonnage et référencer les sources d'information consulter (rigueur, transparence) – s'assurer la démarche systématique de recensement des sources documentaire (utiliser un protocole de recherche/identification/sélection des écrits, un logiciel de gestion de référence).

- Enfin, les pratiques recommandées de ‘J1’ semblent être ancrées dans l’alliance de ses attentes en tant que chercheur et en tant qu’évaluateur de programme. Ceci se traduit notamment par l’attention qu’il porte au caractère complet du devis d’évaluation rapporté et sa reproductibilité, à la démonstration du raisonnement évaluatif, la vérification de la qualité de l’instrumentation, l’appréciation de la fiabilité et de la représentativité des résultats. Ainsi, **les préoccupations de ‘J1’ se concentrent particulièrement sur l’importance de :**

Profil de ‘J1’
Junior avancé
- Firme [B]
- Formation
  [PhD SHS]
  [DES ÉP inclus]
- Expérience pro. ÉP
  [18 mois/Act.1]
  [28 mois/Act.2]

  - Décrire et documenter rigoureusement la démarche évaluative ainsi que l’ensemble des procédures de collecte et d’analyse de données (fiabilité et représentativité) – assurer la reproductibilité du devis, démontrer la rigueur méthodologique, et l’alignement « question - indicateurs - sources de données/instrumentation - résultats » ainsi que l’effort de triangulation (partager la matrice éval., la liste d’indicateurs, l’instrumentation, référencer les sources d’information consultées, etc.).
  - S’assurer de la pertinence des questions évaluatives – vérifier l’évaluabilité des questions telles que formulées, et à la lumière de l’ampleur des données existantes, éviter l’utilisation mécanique des questions standards du cadre d’évaluation du gouvernement fédéral (fiabilité et représentativité).
  - Veiller au caractère culturellement approprié de la démarche évaluative – composition de l’équipe éval. (inclure des évaluateurs qui comprennent les enjeux socioculturels associés au contexte), instrumentation (utiliser un langage et des modalités de passation appropriés et accessibles).
  - Décrire les sources d’information utilisées, et s’assurer de leur pertinence et représentativité à la lumière des questions d’évaluation traitées (taille, étendue, et critères d’échantillonnage).
  - Faire preuve de rigueur et de transparence dans l’appréciation des résultats (fiabilité et représentativité) – présenter les résultats de manière juste et équilibrée : trianguler les preuves, citer la provenance des résultats utilisés, rendre compte du poids des résultats (idée de quantité/proportion, p. ex. 3 pers./4, la majorité), présenter les résultats dans leur entièreté, etc.

Il importe de souligner que, si les pratiques méthodologiques recommandées des évaluateurs novices (ici, dans le contexte de l'évaluation externe de programmes du gouvernement fédéral) résultent de différents cadres de référence mobilisés selon le profil des participants (critères de scientificité de la recherche, expérience professionnelle dans le présent contexte, standards de qualité du gouvernement fédéral, pratiques standards en évaluation de programme, etc.), celles-ci sont alignées avec un certain nombre de normes du standard de précision-justesse (« *accuracy* ») de l'évaluation (JCSEE, 2011). Par ailleurs, outre l'attention portée au contenu du devis méthodologique, ces deux évaluateurs se sont penchés sur la question de la précision et de la justesse sur le plan de l'appréciation (et de l'utilisation) des résultats. 'J1' évoque clairement l'influence de son profil de chercheur, alors que 'J4' mobilise ces attentes en tant qu'évaluateur de programme, lorsqu'ils se penchent sur le manque de transparence quant à la provenance des preuves sur lesquelles s'appuient certains constats évaluatifs, et sur la mérepréssentation des résultats. Ces derniers abordent l'importance de communiquer les données/résultats de manière juste et équilibrée, et ce, afin de veiller juste à la fiabilité, la représentativité et l'utilité des conclusions évaluatives. 'J4' insiste également sur le rôle essentiel de l'évaluateur sur le plan de l'interprétation des résultats où son jugement évaluatif doit être clair (en somme ce n'est pas au lecteur de formuler son propre jugement).

'J1' et 'J4' couvrent ici un certain nombre de normes relatives au standard de la précision-justesse et de l'utilité de l'évaluation (JCSEE, 2011). Précisons que les préoccupations relatives à la mérepréssentation des résultats face à l'insuffisance des preuves mobilisées pour appuyer les constats évaluatifs, la difficulté de cerner la provenance des preuves, la présentation non équilibrée des résultats, etc., évoquent des risques de mésévaluation, de mésusage intentionnel, ou de non-utilisation/mésusage non intentionnelle (Alkin et King, 2017 ; Patton, 2015a). En effet, pour les évaluateurs 'J1' et 'J4' les problèmes observés peuvent provenir d'un manque de rigueur de la part de l'équipe évaluative sur le plan :

(1) de la démarche méthodologique, conduisant à une *mésévaluation*, soit l'utilisation par l'évaluateur et les utilisateurs ciblés de résultats qui ne sont pas valides et fiables (absence de triangulation des preuves, absence de données probantes, manque de représentativité des échantillons, etc.) ;
(2) de la communication des résultats conduisant à *une non-utilisation* (mise à l'écart de résultats valides et fiables de manière non intentionnelle), ou encore une *mésévaluation*

(mauvaise présentation des résultats qui poussent à remettre en question la justesse des conclusions évaluatives).

Cependant, les évaluateurs novices, plus particulièrement ‘J4’, évoquent une autre possibilité, celle du *mésusage intentionnel*. Soit la manipulation des résultats par les utilisateurs ciblés, le comité d’évaluation interne (*data cherry-picking*), ou encore la restructuration voire la réécriture des conclusions initialement rédigées par l’équipe évaluative externe.

Enfin, si les évaluateurs novices n’ont pas nécessairement fait part d’une analyse profonde du lien entre les facteurs contextuels et le devis d’évaluation, ces derniers ont tout de même mis de l’avant certains éléments ayant une influence sur la pertinence de la démarche évaluative, tels que : les intrants disponibles pour l’évaluation (compromis entre le cahier des charges et la répartition efficace des ressources) ; le cahier des charges prescrit dans la demande de propositions émise par le comité d’évaluation interne commanditaire (méthodes, sources d’information, nombre d’entrevues, ressources allouées, etc.) ; les attentes politiques (politique évaluative fédérale, directives de l’IGF, pratiques standards perçues, etc.), les besoins des utilisateurs ciblés, la rigidité du contexte organisationnel et la possibilité de recourir à des approches méthodologiques innovantes ; ou encore la sensibilité culturelle du milieu. On recouvre ici des préoccupations ayant trait à des éléments de contraintes de l’évaluation en situation réelle (Bamberger et Marbry, 2020) ainsi qu’à l’utilité et à la faisabilité (*viabilité contextuelle, utilisation des ressources*) de l’évaluation (JCEE, 2011). Nous revenons plus en détail sur ces éléments dans la section sur l’interprétation des questions de recherche secondaires.

Tableau 12. *Pratiques méthodologiques recommandées des évaluateurs novices dans le contexte de l'évaluation de programmes du gouvernement fédéral*

		'J1'	**'J2'**	**'J3'**	**'J4'**
		• Firme [B] • Formation [PhD SHS] [DES ÉP] • Expérience pro. ÉP [18 mois/Act.1] [28 mois/Act.2]	• Firme [A] • Formation [PhD SHS] • Expérience pro. ÉP [-1 an/Act.1] [16 mois/Act.2]	• Firme [A] • Formation [M.Écon] • Expérience pro. ÉP [4 ans - 4 ans1/2]	• Firme [B] • Formation [M.P.P.A] [DES É.P] • Expérience pro. ÉP [2 ans - 2 ans 1/2]
Cadre de référence		*Critères scientificités recherche expérimentale, pratiques standards en évaluation de programme, directives politique évaluative*	*Critères de scientificités en recherche expérimentale, pratiques de sa firme/standards du gouvernement fédéral*	*Lignes directrices IGF, politique évaluative, exigence des clients, modèle de sa firme, de sa pratique*	*Pratiques standards en évaluation de programme, politique évaluative*
Cadre d'évaluation	Fonction/Type/Approche éval.			• Suivre le cadre de référence de la politique évaluative. • S'assurer de l'alignement « fonction - axe(s) éval. ». ◦ F. formative - éval. de la conception et l'implantation (statut d'implantation, résultats initiaux). ◦ F. sommative - éval. de la pertinence et du rendement (efficacité, efficience et économie). • Formuler des questions standards selon les axes d'éval./de questionnement fondamentaux de la politique éval.	• Promouvoir l'utilisation d'une approche participative lors de l'évaluation d'un programme avec une composante culturelle sensible (p.ex. en milieu autochtone). • Tenir compte des besoins des utilisateurs ciblés/objectif éval. - utilisation du processus en interne (reddition de compte), ou des résultats à des fins d'amélioration. ◦ Si à des fins d'amélioration, impliquer les utilisateurs ciblés dans la conception des questions. • S'assurer que les questions reflètent le contenu du prog., particulièrement dans le contenu des questions prescrites par la politique éval. qui sont assez large.
	Questions d'évaluation	• Suivre le cadre de référence du gouvernement fédéral (axes de questionnement - pertinence, rendement, efficience). • S'assurer de la pertinence des questions d'évaluation. ◦ Évaluabilité des questions telles que formulés. ◦ Ampleur des données de programme existantes.	• S'assurer de formuler des questions claires et précises.		
	Plan d'évaluation/Matrice éval.	*Standard évaluation de prog.* • Préoccupations centrées sur la fiabilité, la validité et la représentativité du devis d'évaluation • Développer une matrice éval. ◦ Expliciter clairement la contribution de chacune des sources de données aux différentes questions d'évaluation. ◦ Démontrer l'effort de triangulation des méthodes et sources d'information. • Démontrer/examiner l'alignement « question - indicateurs - sources de données - contenu instrumentation ».	*Standard gouvernement fédéral* • Utiliser matrice éval. pour ◦ Expliciter clairement aux clients la manière dont la méthodologie évaluative va répondre à chaque question d'éval. ◦ Assurer le découpage rigoureux des questions d'éval. de l'efficacité souvent très génériques, en sous-questions/indicateurs précis (p.ex. lister chaque activité, leurs extrants et les résultats attendus). ◦ Vérifier que la méthodologie permet bien de répondre à chacune des questions éval.	*Pratique standard de sa firme « Evaluation crosswalk »* • Référencer pour chaque question d'évaluation, les sources de données utilisées pour y répondre. • Assurer la planification rigoureuse de la méthodologie évaluative - rendre compte du lien « question d'éval. - éléments à examiner - sources de données à utiliser ».	*Standard évaluation de prog.* • Rendre compte du lien « questions éval. - indicateurs - sources de données ». ◦ Communiquer le plan d'évaluation/protocole de la méthodologie évaluative. ◦ S'assurer que la démonstration de l'efficacité du programme et de son implantation repose sur des données probantes.

		'J1' • Firme [B] • Formation [PhD SHS] [DES ÉP] • Expérience pro. ÉP [18 mois/Act.1] [28 mois/Act.2]	'J2' • Firme [A] • Formation [PhD SHS] • Expérience pro. ÉP [-1 an/Act.1] [16 mois/Act.2]	'J3' • Firme [A] • Formation [M.Écon] • Expérience pro. ÉP [4 ans - 4 ans1/2]	'J4' • Firme [B] • Formation [M.P.P.A] [DES É.P] • Expérience pro. ÉP [2 ans - 2 ans 1/2]
Cadre de référence		*Critères scientificités recherche expérimentale, pratiques standards en évaluation de programme, directives politique évaluative*	*Critères de scientificités en recherche expérimentale, pratiques de sa firme/standards du gouvernement fédéral*	*Lignes directrices IGF, politique évaluative, exigence des clients, modèle de sa firme, de sa pratique*	*Pratiques standards en évaluation de programme, politique évaluative*
Cadre Méthodologique	Sources d'information et échantillonnages	• Décrire et (justifier) les sources d'information et les procédures d'échantillonnage. • S'assurer de la pertinence, représentativité des échantillons (taille, étendu, caractéristiques). • Inclure une perspective comparative lors de l'éval. d'un jeune prog. gouv. d'aide sociale avec enjeux sociopolitiques complexes ◦ Comparaison interjuridictionnelle - sélection documentation, informateurs clés de prog. d'autres provinces, ou pays du Commonwealth). • S'assurer de recueillir la perspective des employés, clients, bénéficiaires du prog.	• Décrire (et justifier) les sources d'information et les procédures d'échantillonnage. ◦ Mentionner les contraintes existantes. • S'assurer que les critères de sélection des sources d'information sont alignés avec les questions d'évaluation. • Inclure une perspective comparative dans l'éval. d'un prog. gouv. ciblant une problématique sociale avec des enjeux sociopolitiques complexes. ◦ Comparaison interjuridictionnelle - sélection documentation de prog. d'autres provinces du pays, ou pays du Commonwealth) • Importance d'inclure les perceptions de la pop. ciblée par le prog.	• Décrire (et justifier) les sources d'information et les procédures d'échantillonnage. ◦ Référencer la documentation, les écrits scientifiques - utiliser un logiciel de gestion de référence (renforce la rigueur et la transparence). ◦ S'assurer du processus systématique de la démarche de recherche et de sélection de la documentation pour la recension des écrits (p.ex. utiliser un protocole de recherche, d'identification et de gestion/d'organisation de la documentation).	• Importance d'inclure les perceptions de la population ciblée par le programme (si certains groupes ne sont pas inclus, justifier pourquoi).
	Procédures de collectes de données	• Décrire de manière précise les procédures de collectes de données afin d'assurer la réplicabilité. • Expliciter clairement la contribution de chacune des sources de données aux différentes questions d'évaluation. • S'assurer de l'alignement entre les questions, les indicateurs et le contenu de l'instrumentation. • Communiquer la matrice éval. et l'instrumentation.	• Décrire de manière précise les procédures de collectes de données. • S'assurer de la pertinence/faisabilité des procédures de collectes de données à la lumière des besoins des clients et des contraintes de temps et d'argent.	• Justifier le choix des méthodes (i.e., raisons théoriques, intrants disponibles, rigueur). • Expliciter clairement et documenter les procédures de collectes de données (exigence des clients, standard de qualité éval. fédéral). ◦ Démontrer/soutenir l'effort de rigueur des évaluateurs. ◦ Assurer la confiance des clients dans la qualité de la démarche méthodologique. • Communiquer l'instrumentation.	• Justifier le choix des méthodes même si plus ou moins prescrit/standard fédéral (justifier exclusion de méthodes comme l'enquête). • Décrire avec clarté et précision les procédures, particulièrement dans le cadre des études de cas qui comprennent le recours à plusieurs méthodes de collectes de données.
	Procédures d'analyse de données	• Décrire de manière précise les procédures d'analyse de données qualitatives afin d'assurer la réplicabilité.	• Décrire de manière précise les procédures d'analyse de données qualitatives afin d'assurer la réplicabilité. • Considérer les biais ◦ Recours à la triangulation des évaluateurs, illustration de la rigueur de la démarche d'analyse qualitative des données d'entrevues afin qu'on puisse apprécier la justesse des interprétations (caractère systématique de la démarche).	• Expliciter clairement et documenter le processus d'analyse des informations qualitatives et de triangulation des résultats des différentes sources de données (exigence des clients, standard de qualité éval. fédéral). ◦ Recours à une démarche systématique. ◦ Démontrer/soutenir l'effort de rigueur des évaluateurs. • Communiquer le plan d'analyse des données et de synthèse des résultats.	• Décrire avec précision les procédures d'analyse. • Faire preuve de rigueur dans l'analyse qualitatives des données d'entrevues ◦ Rendre compte du caractère systématique des procédures d'analyse qualitatives afin qu'on puisse apprécier la justesse des interprétations.
	Limites	• Citer les limites rencontrées	• Citer les limites rencontrées	• Citer les limites rencontrées et décrire les tentatives de gestion de ces dernières.	

4.1.2 Le cas des évaluateurs séniors (experts)

À l'unanimité, les évaluateurs experts ont annoncé en amont de leur analyse critique que le contexte évaluatif, programme fédéral en autochtonie, était très spécifique, sensible et complexe (Tableau 13, ci-dessous). Aussi, ils ont immédiatement adopté une posture ancrée dans un contexte socioculturel et politico-économique dont les particularités sont plus ou moins déterminantes dans leur manière d'aborder cet exercice. Les évaluateurs 'S1', 'S2' et 'S4' ont explicitement déclaré avoir de l'expérience en évaluation de programmes fédéraux dans le présent contexte. 'S2' et 'S4' semblent avoir une expérience plus avancée, particulièrement 'S2', qui a réalisé des évaluations de services et interventions similaires.

Tableau 13. *Caractéristiques de l'analyse critique selon l'expérience des évaluateurs séniors*

'S1' **Sénior consultant** • Firme [A] • Formation [M.A. SHS] [Juriste droit civ.] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [16 ans] • Passation en anglais	Il annonce dès le début que pour lui la démarche évaluative en réponse à la politique évaluative fédérale fera généralement preuve d'un manque de considération pour la population autochtone dont les communautés ne sont pas (ou très peu) considérées comme des utilisateurs ciblés de l'évaluation - non consultées dans la conception du devis d'évaluation, méconnaissance (ou manque de prise en considération) des méthodes et procédures de collectes de données culturellement appropriées. Nous verrons que pour lui le devis d'évaluation devrait être suffisamment décrit (et documenter) pour qu'il puisse apprécier la qualité de la démarche évaluative, et la représentativité des résultats.
'S2' **Sénior partenaire** • Firme [B] • Formation [PhD SHS] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [16 ans] • Enseigne ÉP au niveau universitaire • Passation en anglais	Il déclare que son jugement critique sera ancré dans ses expériences évaluatives passées de services fédéraux en milieu autochtone. Il explique avoir mobilisé le même cadre de référence que celui qu'il utilise dans sa pratique à savoir les lignes directrices du guide de revue des évaluations fédérales. Nous verrons qu'il s'est particulièrement concentré sur la pertinence et la précision-justesse du devis d'évaluation, sur la fiabilité, la validité, la représentativité des résultats et l'interprétation, et l'utilité/crédibilité de l'évaluation.
'S3' **Sénior associé** • Firme [B] • Formation [M.A. SHS] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [20 ans] • Membre du conseil d'administration association professionnelle en ÉP • Passation en anglais	Il mentionne avoir tenu compte de la complexité du contexte évaluative dans son analyse. Nous verrons qu'il s'est particulièrement concentré sur la pertinence et la précision du devis d'évaluation et la qualité de la gestion des contraintes de terrain (politique évaluative fédérale, paramètres du cahier des charges prescrits dans la demande de propositions (*RFP*), maturité du programme, problèmes d'accès aux données, etc.), sur la fiabilité et la représentativité des résultats et interprétations, et ce, sans que son jugement se fasse sous le spectre de la sensibilité culturelle du milieu.

‘S4’ Sénior partenaire (directeur)	
• Firme [A] • Formation [M.A. SHS] [Juriste droit civ.] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [17 ans] • Ancien président d’une association professionnelle en ÉP • Passation en français	Il explique que le problème fondamental qui va se poser aura trait au recours à un devis d’évaluation gouvernemental standard, et très traditionnel, qui ne sera pas adapté selon lui à la réalité de l’exercice évaluatif auprès des communautés autochtones. Malgré tout, ses critères d’analyse seront définis par la politique évaluative fédérale et les lignes directrices de l’IGF - ancrée dans un « *paradigme non autochtone* ». Nous verrons qu’il évalue la pertinence du devis d’évaluation à la lumière des besoins des utilisateurs ciblés, soit la reddition de compte, et de la réalité des contraintes inhérentes terrain (politique évaluative fédérale, paramètres du cahier des charges prescrits dans la demande de propositions, maturité du programme, problèmes d’accès aux données, etc.).

Enfin, dans le cadre de l’activité d’analyse critique des évaluateurs juniors, le point critique centrale portait sur *le manque de précision dans la description des procédures de collectes et d’analyses des données et les limitations que cela engendre* (42 uds). L’influence des facteurs contextuels se trouvait davantage au second plan, faisant souvent l’objet de relances de la part de la chercheuse. Aussi, avec une fréquence de cooccurrence de 50 cos entre les éléments clés du devis d’évaluation et les facteurs contextuels, il était pertinent de présenter les deux axes de résultats successivement sans crainte de redondance. En revanche, dans le cadre de leur analyse critique, si les évaluateurs séniors ont abordé le manque de description du contenu rapporté, ils ont également concentré leur attention sur la pertinence ainsi que sur la précision au sens de la rigueur et de la justesse des choix méthodologiques et de leurs mises en œuvre. Aussi, avec une **fréquence de cooccurrence de 159 cos au cœur de la dyade [devis d’évaluation-facteurs contextuels]**, les éléments critiques mis en exergue par ces derniers s’avéraient étroitement intriqués avec leur perception du rôle joué par les facteurs contextuels. Afin de renforcer la pertinence et la clarté de la présentation des résultats, nous les avons donc présentés de manière conjointe.

4.1.2.1 Le devis d’évaluation : une analyse de contenu ancrée dans la perception de la pertinence et de la précision-justesse de la démarche évaluative, et de la gestion des facteurs contextuels

4.1.2.1.1 L’analyse du contenu du cadre d’évaluation

Comme l’illustre le diagramme circulaire à la page suivante (Figure 24, p. 182), la majeure partie de l’analyse critique des évaluateurs en ce qui a trait au contenu du cadre d’évaluation porte sur le *cahier des charges* (43 uds), et plus particulièrement sur son *manque de pertinence* (13 uds) et son *manque de justesse* (17 uds). Ensuite, dans de moindres mesures, ils se sont penchés sur *la*

pertinence et la précision-justesse des questions d'évaluation (16 uds) ainsi que le *manque d'information sur le plan du contenu du cadre d'évaluation rapporté* (12 uds). Concernant le dernier point, il s'agit de constats particuliers émis par 'S1' et 'S2', qui déplorent l'absence du modèle logique (5 uds) et l'absence des matrices (4 uds). Pour 'S1', l'absence de ces deux outils essentiels freine grandement sa capacité de juger de la pertinence et de la précision des questions d'évaluation, du choix des indicateurs et de leur alignement avec ces dernières. Pour 'S2', le modèle logique devrait toujours être réalisé et inclus dans le rapport d'évaluation finale. Il est à la fois un outil de la conception et de la mise en œuvre du devis d'évaluation, et il est également un extrant (ou produit) de l'évaluation. Selon lui, sans modèle logique, il est difficile de fournir des preuves de performance et d'efficience du programme évalué. 'S2' aborde également le manque de précision sur le plan de la description du profil du programme et de ses ressources ainsi que des problèmes de traduction (problème de fiabilité et de représentativité des informations).

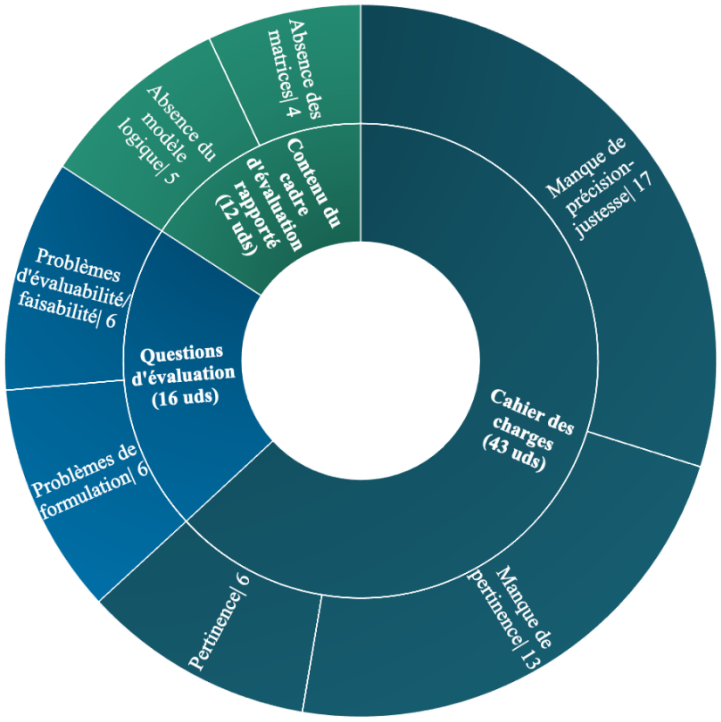


Figure 24. *Thématiques et éléments d'analyse relatifs au cadre d'évaluation (évaluateurs séniors)*

En ce qui a trait à **l'appréciation du contenu des questions d'évaluation**, les évaluateurs séniors se sont penchés sur leur alignement avec les lignes directrices de l'IGF, ce qui les a conduits en majorité à souligner leur pertinence, *celles-ci demeurant standards* (4 uds). En revanche, comme le montre la Figure 24, les évaluateurs séniors ont souligné des *problèmes*

d'évaluabilité/faisabilité au niveau de certaines questions d'évaluation (6 uds), en raison notamment de *problèmes de formulation* (6 uds). De son côté, 'S4' mentionne les attentes irréalistes de la politique d'évaluation et par extension, du comité d'évaluation interne ainsi que les contraintes importantes d'accès aux données dans la production de preuves empiriques sur l'efficience et le rendement du programme. Ces problèmes d'évaluabilité auraient pour lui conduit à la proposition d'une version détournée du contenu de l'axe attendu de questionnement, rendant la question intéressante à ses yeux, mais non pertinente au regard de la politique évaluative fédéral. Nous y reviendrons par la suite dans le cadre de la discussion du contenu du cahier des charges, mais il semblerait que les évaluateurs externes soient dans l'obligation de répondre à un axe de questionnement pour lequel ils sont dans l'impossibilité de répondre, ces derniers ayant un pouvoir décisionnel limité quant à la définition du cadre d'évaluation. Pour 'S3', ceci en revient à remettre en question la pertinence de proposer une question et tenter d'y répondre alors que les évaluateurs ne sont pas en mesure d'accéder aux données probantes nécessaires. 'S1' souligne d'ailleurs que cette obligation d'aligner les questions d'évaluation avec les priorités de la politique gouvernementale diminue la pertinence de certaines d'entre elles, réduisant ainsi les chances d'utilité de l'évaluation. En ce qui concerne 'S2', celui-ci note que certaines questions ne sont pas des questions d'évaluation, mais plutôt des questions de recherche, et il ne voit pas comment il serait censé y répondre puisqu'il ne comprend pas l'intention derrière ces dernières. Enfin, selon 'S1' si les questions d'évaluation paraissent pertinentes au regard des standards de l'IGF, celles-ci s'avèrent, comme toujours, très larges. Aussi, pour lui, ce manque de précision aurait pu être contrecarré par la présentation du modèle logique et des matrices dans le rapport d'évaluation, permettant ainsi au lecteur de bien saisir les intentions des évaluateurs et ce qui sera considéré comme preuves permettant de répondre à ces questions.

L'appréciation du contenu du cahier des charges est étroitement liée aux facteurs contextuels et à la gestion de ces derniers par les évaluateurs de programme (Figure 25, p. 184). En effet, nous remarquons que le cahier des charges constitue l'élément du devis d'évaluation qui a la plus haute fréquence de cooccurrence avec les *facteurs contextuels* (84 cos), plus particulièrement avec les *contraintes politiques* (51 cos). Aussi, ceci n'est pas surprenant puisqu'il constitue la pierre angulaire du devis d'évaluation (assurance de l'alignement [questions d'évaluation - indicateurs de mesures - méthodes - sources d'information - répartition des intrants]), et de l'entente partenariale entre le comité d'évaluation interne commanditaire de

l'évaluation et l'équipe évaluative externe. En effet, initialement conçu par la *division interne de l'évaluation* (6 cos), à partir des *lignes directrices de la politique de l'IGF* (12 cos), le cahier des charges est ensuite soumis au *comité d'évaluation interne* (9 cos), puis intégré dans *la demande de propositions* (3 cos). Une fois le contrat obtenu, *l'équipe évaluative consultante collabore avec l'évaluateur interne afin de finaliser le cahier des charges* (3 cos). Cependant, les évaluateurs mentionnent à plusieurs reprises que les évaluateurs externes sont ici des *fournisseurs de services* (5 cos), leur *pouvoir décisionnel demeurant assez restreint* (7 cos). L'évaluateur interne fait d'ailleurs la *liaison avec le comité d'évaluation interne qui semble valider tous choix, tous changements proposés, etc.* (4 cos).

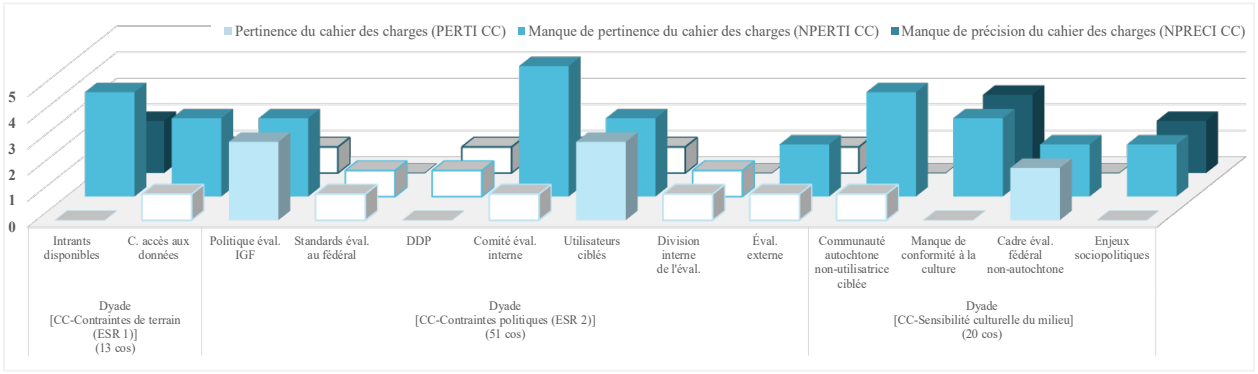


Figure 25. *Fréquences de cooccurrence entre les éléments constitutifs de chacune des dyades [cahier des charges - groupe de facteurs contextuels]*

L'analyse des cooccurrences (Figure 25) entre les différents contenus clés du cahier des charges et groupes de facteurs contextuels (contraintes de terrain [ESR 1], contraintes politiques [ESR 2][25], et sensibilité culturelle du milieu) a mis en exergue différentes problématiques reliées à la pertinence (ou au manque de pertinence) du cahier des charges (PERTI CC - NPERTI CC) ainsi que son manque de description et de précision-justesse (NPRÉCI CC). <u>En voici les thématiques clés</u> :

[25] Il est question ici des contraintes de l'évaluation en situation réelle – C. politiques, temporelles, budgétaires, et d'accès aux données (Bamberger et Marbry, 2020). Compte tenu de l'ampleur des contraintes politiques, nous avons fait le choix de les présenter à part, et d'inclure les trois autres contraintes ESR sous le facteur 'contraintes de terrain' où les contraintes budgétaires et temporelles sont intégrées dans l'item 'intrants disponibles'.

a. La “quasi-impossibilité” d’innover ou de mobiliser un devis plus sophistiqué

[*Facteurs contextuels*]

- Caractère prescriptif de la politique évaluative de l’IGF et de la demande de propositions (6 cos)
  - Cadre d’évaluation - cahier des charges
  - Paramètres stricts en ce qui a trait aux intrants disponibles
- Le fonctionnement institutionnel et la culture évaluative du gouvernement fédéral (2 cos)
- Le statut de fournisseur de services de l’évaluateur externe (2 cos)
  - Pouvoir décisionnel restreint

‘S4’ aborde l’impossibilité d’innover ou de proposer des devis plus sophistiqués au regard des problèmes de faisabilité engendrés notamment par le caractère prescriptif de la politique évaluative de l’IGF, les paramètres stricts en termes d’attentes et d’intrants ainsi que la question de l’accès aux données. Aussi, pour lui, comme pour les autres évaluateurs séniors, il n’est pas possible de recourir au devis expérimental ou même quasi expérimentale, de même qu’à une réelle analyse coût-bénéfice dans le cadre de l’évaluation de l’économie et de l’efficience. On revient d’ailleurs un peu plus loin sur l’impossibilité de fournir des données probantes, un problème reconnu, et une attente persistante de la politique évaluative et par extension des clients qui désirent satisfaire aux exigences de cette dernière et du gouvernement fédéral.

Extrait de verbatim

> *« Au fédéral, ça ne permet pas vraiment d’innover en termes de devis. La politique, si elle a un niveau de flexibilité, demeure prescriptive avec des paramètres très stricts en termes d’intrants. Ce sont des programmes évalués aux cinq, ça nous limite énormément en termes de ce qu’on est capable de mettre en œuvre. Ici, ils ont eu 9 mois. Qu’est-ce qui reste comme métho ? Impossible de faire du quasi expé, de l’expé, de vraies études coûts-bénéfices, pas de temps, pas d’argent, etc. En 9 mois, développement de la métho, des instruments et de leur validation, ça prend deux mois et puis à la fin deux à trois mois pour l'analyse, rédiger le rapport, cycle de révision, etc. Ça laisse 4 mois pour - 30 entrevues, 4 études de cas, analyse documentaire, recension, etc. Déjà ça prend une équipe. C’est comme ça au fédéral, ils n’ont pas le temps, il y a de nombreux niveaux d’approbation, etc. Ce n’est pas comme aux États-Unis où quand ils veulent piloter un programme, il demande un devis quasi expé ils donnent plusieurs années pour faire l’évaluation, 800 000 $ pour faire l’évaluation. Tout est fait pour que ce soit faisable. Nous, ici, dès qu’on obtient le contrat, on est en mode panique. Alors là en milieu autochtone pour planifier la collecte il y a une dimension géographique et socioculturelle très spécifique, on est dépendant de la saison, du climat, s’il y a des élections au conseil de bande, etc. Il y a très très peu de marche de manœuvre pour l’innovation à cause de tout ça. »* (S4).

‘S3’ nous rappelle, quant à lui, qu’à titre de consultants, les évaluateurs externes sont des fournisseurs de services qui répondent à une demande de propositions où les commanditaires décrivent clairement, voir prescrivent, le contenu du mandat d’évaluation et la façon dont ils doivent y répondre. En somme, dans le présent contexte, les évaluateurs externes sont davantage

des exécutants qui ont un pouvoir décisionnel assez limité à ce niveau-ci du devis d'évaluation. En outre, ils font également face à des contraintes de terrain qui limitent considérablement la possibilité de proposer d'autres alternatives sur le plan méthodologique. En effet, l'ajout de méthodes ou encore l'enrichissement de la portée de ces dernières s'avère toujours trop coûteux en termes d'intrants, l'application de la politique évaluative fournissant par ailleurs une définition des paramètres plutôt stricts en ce qui a trait au cahier des charges ainsi qu'aux ressources disponibles pour sa mise en œuvre.

Extrait de verbatim

"[R - The design brief is already set by the clients?] - *Definitely. Luckily for us, we are suppliers. We respond to request for proposals sent out to consultants that says we want somebody to do an evaluation of this program, and this is what we want to find out and the methods that we see as appropriate/the way we want you to find out. […]. Sometimes if we have suggestions to improve the design, the clients would say to go right ahead and provide a rationale to do it. But because always, there's a dollar value attached they would add we want you guys to do this amount for this amount of work for fifty thousand dollars. So you do that and then it's very difficult to say we'd also like to do a survey within their budget. I mean, sometimes you can do that and sometimes you can't. It's kind of hard to innovate given the parameters they've already given*." (S3).

b. La question du caractère culturellement approprié, ou non, du cahier des charges

[*Facteurs contextuels*]

- La perspective non autochtone de la politique évaluative de l'IGF qui se reflète sur le contenu prescrit dans la demande de propositions (10 cos)
  - Cadre d'évaluation - cahier des charges standard pour l'évaluation de tous programmes fédéraux
  - Intrants disponibles non adaptés à la réalité culturelle du milieu
  - Choix méthodologiques culturellement inappropriés
- Parties prenantes clés de la communauté autochtone non considérées comme un utilisateur clé de l'évaluation (7 cos) et enjeux sociopolitiques associés (5 cos)
  - Absent du comité d'assurance qualité de l'évaluation - comité d'évaluation interne responsable de la demande de propositions, de la validation du contenu du cahier des charges ainsi que de tous choix méthodologiques et la communication de l'évaluation.
  - Manque de transparence quant au respect, ou non, des principes de l'éthique de la recherche « OCAP » des Premières Nations.

Nous l'avons vu plus tôt, 'S1' annonce dès le départ que le gros point noir de l'évaluation de programmes fédéraux autochtones repose sur le fait que la démarche évaluative ne tient pas compte des particularités politico-économiques et socioculturelles du milieu. Pour lui cela s'observe dès la réception de la demande de propositions dont les termes relatifs aux intrants

s'avèrent toujours inappropriés, mettant ainsi en exergue les attentes irréalistes du cahier des charges des clients, cela conduirait les évaluateurs externes à procéder de manière culturellement inappropriée pour accéder aux données. Par ailleurs, il déplore l'absence de parties prenantes clés des communautés autochtones au sein du comité d'assurance qualité, soit le comité d'évaluation interne qui, entre autres, valide les choix méthodologiques (contenu du cahier des charges, sources d'information, contenu des instruments, procédures de passation, etc.).

Extrait de verbatim

"This goes to one of my rants. It always surprises me when an RFP comes out and it involves indigenous communities, and they will want the work done in like six months and they'll want the data collection done in two or three months and the research techniques selected are almost never appropriate with indigenous people. I don't know if that's the case for this one, but evaluators are often, under these tight timelines and design constraints so what's asked is already culturally inappropriate. I have been in this position where you are pressing people on timelines and committing to you, and often it doesn't work that way. You need to be more respectful, and you're made to be kind of demanding which I don't like. And you know it also doesn't consider things that won't happen the way you planned, you probably won't be going directly into the communities, you must talk to elders. […]. Also, if something happens in a First Nations community and it's a tragedy, everything shuts down. And so, your timelines really must consider that unfortunately some tragedy happens, and you need to have the time built in. You can't cancel a site and set it up again later and it can be quite a bit later. So, I'm guessing that part of the reason is that the timelines were tight, and a more in depth and meaningful involvement of Indigenous people would require longer timelines. But you need to be culturally appropriate, and I believe the inclusion of the key community stakeholders will add more value to the evaluation" (S1).

'S4' quant à lui reconnait que la perspective non autochtone du cadre d'évaluation rend l'évaluation à la fois pertinente et non pertinente. Pertinente, dans la mesure où elle respecte les lignes directrices de la politique évaluative de l'IGF et les standards de l'évaluation de programmes fédéraux. Non pertinente, cependant, en ce qui a trait au non-recours à une approche évaluative et des méthodes dites autochtones. Aussi, pour lui cela renvoie à analyser si l'évaluation répond aux attentes des utilisateurs ciblés de l'évaluation, soit, ici, les parties prenantes clés internes du gouvernement fédéral - l'organisme financeur, le département et les gestionnaires responsable du programme et la division de l'évaluation interne de l'IGF. Dans le cas présent la démarche évaluative est donc, pour 'S4', pertinente.

Extrait de verbatim

« *On a un problème fondamental, la métho va être gouvernementale, car c'est entrepris par le ministère et donc on ne sera pas en lien avec ce que les communautés autochtones voudraient comme approche évaluative, comme métho.* [montre un livre sur la décolonisation des méthodologies évaluatives]. *C'était une critique justement du gouvernement, des approches mainstream, du cadre d'éval traditionnel établi par le gouvernement fédéral, une métho hyper traditionnelle. [...]. C'est délicat comme question parce la politique sur l'évaluation, on part complètement dans un paradigme non autochtone, et ce, malgré le département concerné. Est-ce qu'ils auraient pu faire autrement ? Je ne sais pas. S'ils faisaient autrement, ils ne respecteraient plus la politique en évaluation, le rapport ne serait pas accepté par le Conseil du trésor. Là ça ramène la question fondamentale, j'imagine, je m'en vais sur un terrain glissant, mais j'y vais quand même. Pour qui fait-on l'évaluation ? On le fait pour le ministère, on ne le fait pas pour les autochtones. Le ministère investit là-dedans et veut savoir si son argent est bien investi donc j'imagine de le faire d'une faire perspective non autochtone comme on regarde n'importe quel autre programme, en ce sens-là je trouve que la démarche et les choix sont bons.* » (S4).

À l'unanimité, les évaluateurs déclarent que la sous-traitance des études de cas à une firme autochtone était une décision pertinente au moment de la définition du cadre d'évaluation/cahier des charges. En effet, ils expliquent que cela facilite grandement l'accès aux données ainsi que la richesse des informations collectées. En revanche pour faire écho à ce qui a été mentionné plus haut, 'S4' rappelle que l'évaluation demeure malgré tout inscrite dans un paradigme non autochtone. Le devis est développé par la firme externe mandatée, en collaboration avec la division de l'évaluation et le comité d'évaluation interne. La firme autochtone a ici pour seule responsabilité la collecte des données et leur transmission à la firme consultante pour analyse.

Extrait de verbatim

« Embaucher une firme autochtone pour les études de cas, ça ne change rien au paradigme. Ils s'en sont servi juste pour ça, pour qu'ils aillent collecter les données. Le Framework demeure le même. C'est délicat, parce que ça a donné quoi, ça a fait que les gens sont plus à l'aise de parler, ça a facilité, on l'espère, la collecte de données, l'accès aux informations. L'équipe autochtone a probablement préparé un rapport avec les informations collectées qu'ils ont envoyé à la firme. Ce ne sont pas eux qui font l'analyse et l'agrégation de toutes les données évaluatives. Mais là dans le rapport on n'y fait pas beaucoup référence. Ça n'a pas été la source de données la plus riche au premier abord. Et puis, on ne peut les mettre trop en avant parce que ce n'est pas les communautés qu'on évalue. Le but c'est de dire que d'après les quatre cas, pour X ça marche, il faudrait que... C'est une consultation parmi d'autres, au même titre que les entrevues, pour observer la mécanique du programme c'est tout. Donc les études de cas ne deviennent pas proéminentes. » (S4).

Si on aborde la limite évaluative soulevée par les évaluateurs séniors en ce qui concerne l'utilité de l'évaluation, il nous paraît important de rebondir ici sur les propos de 'S4'. Des propos qui vont à l'encontre des préoccupations de 'S1' vis-à-vis des principes éthiques *OCAP des Premières Nations sur le droit de propriété, d'accès, de contrôle et de possession des données et informations*. En effet, 'S1' s'attendait tout particulièrement à ce que dans le cahier des charges les clients rapportent la façon dont ils ont respecté les principes OCAP dans le cadre de l'évaluation.

Extrait de verbatim

"My continued sort of surprise in terms of federal evaluations that deal with Indigenous people is that the evaluation reports never really tried to address upfront how they are complying with OCAP principles. For doing research with Indigenous people. I kind of feel like this should be a part of any evaluation report that deals with Indigenous people. […]. And that they didn't specify other things that they might have clearly done which would've been you know for the case studies. In making a case study report and giving it back to the individuals that participated in the case studies as part of you know ownership and control of possession and access. You know I was a little surprised by that." (S1)

c. Le manque de pertinence et de précision-justesse du contenu du cahier des charges au regard des divers axes évaluatifs

[*Facteurs contextuels*]

- Les attentes de la politique évaluative de l'IGF et de l'évaluation de programmes fédéraux en général (10 cos)
  - Cadre d'évaluation - cahier des charges prescrit
  - Alignement du programme avec les priorités du gouvernement
  - Conformité du programme avec les lois et responsabilités fédérales
- Les contraintes ESR (7 cos)
  - Intrants disponibles pour l'évaluation
  - Contraintes d'accès aux données (existence et (ou) disponibilité des données, résistances des utilisateurs ciblées, etc.)
- L'évaluateur externe mandaté (3 cos)
  - Statut de fournisseur de service
  - Pouvoir décisionnel restreint
  - Rôle de vérificateur pour lequel il n'est pas nécessairement compétent

Axe 1 - Évaluation de la conception et de la mise en œuvre (8 uds)

Les évaluateurs 'S2' et 'S3' expliquent qu'au regard des caractéristiques politiques et socioculturelles particulières du milieu, ils s'attendaient à ce que le devis évaluatif s'inscrive dans une perspective comparative, et ce, pour plusieurs raisons :

i. Observer les services offerts par des programmes similaires d'autres provinces canadiennes, voir d'autres pays du Commonwealth avec des caractéristiques proches (S3).
ii. Répondre à l'une des questions fondamentales de la politique évaluative fédérale : « *[d]ans quelle mesure le programme est-il aligné avec les priorités gouvernementales* ? », l'une desdites priorités étant l'assurance d'offres de services comparables à la population autochtone dans les réserves, à la population autochtone en dehors des réserves et le reste de la population de la province (S2).

Extrait de verbatim	*« La question fondamentale, c'est que c'est une des responsabilités du gouvernement depuis quelques années... Il y a des ententes... On veut offrir des services comparables aux autochtones au sein des communautés, et en dehors, des services qui sont comparables avec ce que la province livre dans le reste de la province. [...]. Le point de comparaison devrait être - est-ce que oui, ou non, les autochtones reçoivent des services équivalents, excuse-moi, mais je vais dire aux blancs, dans les réserves et en dehors des réserves sur l'ensemble de la province? Le devis ne répond pas à cette question qui est pourtant fondamentale. C'est très regrettable. »* (S2)

À la lecture des résultats sur la conception et la mise en œuvre du programme, 'S2' a remarqué que l'une des questions évaluatives avait été modifiée. Si à première vue ce n'est pas trop préoccupant, il reconnait que cela cause tout de même un problème d'alignement au niveau du cahier des charges, et que la question aurait dû être modifiée dans le cadre d'évaluation. Ce qui en revanche semble problématique, c'est l'ajout de l'élément « de manière efficace » à la fin de la question sur la mise en œuvre du programme. En effet, pour 'S2', cela ferait de cette dernière une question se rapportant au rendement, ce qui signifie une tout autre manière de répondre au mandat comparativement à la version initiale.

Extrait de verbatim	*« L'approche est-elle mise en œuvre comme prévu de façon efficace. Sinon, pourquoi ? Ce n'est pas la même question présentée dans le cadre d'évaluation au début. C'est semblable, mais ce n'est pas la même. L'approche a-t-elle été mise en œuvre comme prévu ? Ils ont ajouté « de manière efficace ». Ce n'est pas un gros problème. Faudrait juste qu'ils aient ajusté cette question-là dès le début. Moi, je suis d'accord avec le changement, mais ici, ce n'est pas ce qu'on avait annoncé. Maintenant, le problème, c'est que de « façon efficace » signifie « efficacité réelle » - un concept qui fait référence au RAM (Risk Assessment Matrix), et ici il faut faire très attention au choix des mots parce que pour les évaluateurs comme moi, une approche qui est efficace c'est une approche qui a des résultats, donc là ça veut dire rendre compte du rendement, de la performance, et ça fait appel à une approche différente, avec des paramètres précis »* (S2).

Axe 2 - Évaluation du rendement - efficacité (8 uds)

Dans la lignée des constatations émises précédemment, et face à l'absence de preuves tangibles en ce qui a trait au rendement des activités du programme, 'S2' et 'S3' émettent l'hypothèse d'un problème au moment de la définition et du choix des indicateurs de mesure, et (ou) de la mise en œuvre du cahier des charges.

Selon 'S3', ceci ne remet pas en cause le manque de rigueur des évaluateurs, mais indique plutôt l'inexistence de données probantes due à l'absence probable de procédures systématiques de monitorage interne de l'implantation du programme. En revanche, pour lui, les évaluateurs ont, ou auraient dû, identifier ce problème lors de la définition de la portée de l'évaluation. Et que, si donné la possibilité (ou plutôt si les intrants le permettaient), ils auraient pu tenter de concevoir eux-mêmes les bases de données afin de pouvoir utiliser un minimum de preuves empiriques pour

soutenir leurs constatations. Sinon, ils auraient dû au moins expliquer dans le rapport pourquoi ils n'ont pas été en mesure de contourner cette contrainte.

Extrait de verbatim

"*There was not a lot of performance data regarding the program activities. What the program was doing in a tangible way. […]. An implementation evaluation it's an accounting exercise - the milestones attained, like the number of people reached since the program's implantation especially because you cannot yet get the mid/long-term outcomes. [...]. I can only surmise that they didn't have that kind of performance data. I don't think they forgot to look for or at it, they are professional evaluators. But they probably knew that early on and could have revised the evaluation scope or explain what alternative they thought about but couldn't do, or how they manage this lack of performance data. There is an evaluation that we are doing where we have very little data on some of the main activities that the evaluation needs to assess. We decided to go in and create a database. In this case, it could have been that some of their outputs or the records of what they were doing were in paper files. Evaluators would have had to redirect evaluation resources to create and populate a database. Ultimately it probably wouldn't be cost-effective, and the evaluators did not approach their client. It becomes a limitation of your evaluation, which I mean, they've been upfront about it and made a recommendation to kind of wait for the next time. However, in the end still tried to demonstrate the performance without empirical proof."* (S3).

Pour 'S2', il s'agit ultimement d'un travail qu'il qualifie de « bâclé ». Si à première vue, le devis paraissait standard pour ce type d'évaluation, il remarque très vite des incohérences au niveau des informations rapportées ainsi qu'une absence de preuves empiriques signalant un problème sur le plan de la définition des indicateurs et (ou) au moment de la mise en œuvre du cahier des charges et de la synthèse des résultats - *amalgame entre extrants et activités, absence de modèle logique, mention des « coûts » dans l'axe évaluatif portant sur l'efficacité alors qu'il s'agit de données relatives à l'efficience, etc.* Face à une question évaluative aussi importante que celle touchant à l'atteinte des objectifs attendus du programme, 'S2' déplore de manière beaucoup plus « caustique » (que 'S3') l'absence de données statistiques clés, de preuves tangibles. Nous l'aborderons davantage plus tard, mais pour 'S2' comme pour 'S3' des déclarations générales du type « *des progrès ont été effectués dans l'atteinte de tous les objectifs visés* » soulignent un manque de précision-justesse au niveau du plan d'investigation de cet axe évaluatif, mais aussi un potentiel manque de transparence au niveau des procédures de collectes de données, et un problème de représentation erronée des informations recueillies, et par extension du programme.

Par ailleurs, pour 'S2' et 'S3' le cahier des charges ne répond pas à une question fondamentale, soit dans quelle mesure le programme vient bel et bien en aide à la population autochtone, et si on observe des changements positifs au sein des communautés. Aussi, nous verrons sur le plan des limites identifiées par les évaluateurs que pour eux cela remet en question l'utilité de l'évaluation pour les bénéficiaires du programme.

En revanche, à la lecture du cadre d'évaluation 'S4' a déclaré ne pas s'attendre à voir de réponses concernant la performance du programme, puisqu'il a été mis en œuvre assez récemment. Il insiste d'ailleurs à plusieurs reprises sur le fait que l'évaluation n'a pas pour objectif d'observer les effets du programme sur la population. Pour lui, il ne s'agit pas d'une question de manque de pertinence, mais de la nécessité de répondre aux besoins des utilisateurs ciblés de l'évaluation, à savoir de répondre aux attentes de la politique évaluative fédérale qui s'inscrit dans une perspective de reddition de compte.

Extrait de verbatim

« *Ici on ne cherche pas vraiment à voir si le programme amélioré le sort des x et des y autochtones. Donc en fait la logique, ici, on regarde la mise en œuvre, donc on regarde est-ce que c'est pertinent, la conception, l'exécution est-ce que ça va. Rendement et efficacité on parle réellement d'opérationnalisation, les facteurs qui aident à la mise en œuvre - donc en fait on ne parle pas vraiment de rendement, on regarde est-ce que ça a bien été conceptualisé. Puis là on a ici une petite question « dans quelle mesure l'approche a-t-elle contribué à l'atteinte des résultats visés ? » - ça, on ne va pas vraiment trouver des réponses dans le rapport. En tout cas rien de convaincant, c'est normal, ça fait un an ou deux de mises en œuvre donc on ne peut pas voir comment on a transformé la vie de la population autochtone. Pas en deux ans. Et puis une fois de plus ce n'est pas l'objectif de l'évaluation.* » (S4)

Axe 3 - Évaluation du rendement - efficience et économie (9 uds)

À différents moments de leur analyse critique, les évaluateurs 'S2' et 'S4' ont remarqué que les constatations relatives à l'efficience et l'économie du programme ne répondaient pas aux attentes de la politique évaluative de l'IGF, et ils ont remis en question de manière plus ou moins directe le cahier des charges.

'S4' a constaté que les évaluateurs ont établi leur cahier des charges à partir d'une conceptualisation de l'efficience et de l'économie différente de celle du gouvernement fédéral. Il émet l'hypothèse que ces derniers ont tenté de contourner une problématique selon lui bien connue, à savoir le fait que dans la très grande majorité des cas les données empiriques n'existent pas. Ici, 'S4' remet donc en question les exigences irréalistes de l'IGF et par extension les attentes des clients qui persistent à vouloir les satisfaire. Cependant, s'il reconnait le caractère ingénieux des évaluateurs, il n'en demeure pas moins que le cahier des charges ne respecte pas les attentes évaluatives. Il explique qu'il ne dénigre pas le manque d'effort de conformité des évaluateurs, puisqu'il aurait fait exactement la même chose. Aussi, il revendique que « *les conditions dans lesquels on exécute les évaluations, c'est complètement irréaliste de penser qu'on va être capable de faire l'analyse de l'efficience et de l'économie* ». En effet, pour lui « *ça prendrait beaucoup de données, beaucoup de temps, et d'argent* », et ce, « *dans un milieu comme ici hyper décentralisé et rigide* ».

Extrait de verbatim

« La façon dont le fédéral conceptualise l'économie et l'efficience, c'est de voir si l'utilisation des ressources a été aussi efficiente que possible et s'il existe diverses manières pour y arriver. Ici, le rapport a un peu tout retourné en disant que la X coûte moins cher que l'Y. C'est sûr que le principe fondamental c'est que la X c'est toujours plus payant. Mais, ce n'est pas ce que l'IGF veut dire - comment ces ressources ont-elles été utilisées ? Est-ce qu'elles ont été utilisées à leur maximum. Malheureusement, 90% du temps, les données n'existent pas. On n'est pas capable de répondre aux attentes. Ce rapport n'est pas différent des autres. Je pense plus ici, à quel point la politique n'a pas été réaliste. C'est plutôt une critique de la politique sur l'efficience et l'économie, plutôt que le rapport lui-même, mais je dirais qu'il n'est pas allé dans le sens de l'IGF, idem pour la méthodologie. » (S4).

'S2' quant à lui a remis en question la précision-justesse du cahier des charges au regard de différents constats :

i. *L'amalgame entre les notions de budget et de dépense au moment de la description des ressources du programme* - 'S2' a observé un manque de clarté et d'exactitude sur le plan des données financières. Aussi, il explique que l'accent mis sur l'économie par l'IGF nécessite que l'évaluateur mobilise des compétences de vérificateurs qu'il ne possède pas.
ii. *L'absence de description sur la manière dont les évaluateurs allaient s'y prendre pour évaluer l'efficience et l'économie* - une chose que 'S2' explique faire de plus en plus dans le cadre de ses propres évaluations.
iii. *L'absence de données probantes au niveau des résultats-constatations relatives à l'efficience et l'économie du programme* - 'S2' fut fortement interpellé par le fait qu'ils abordaient l'efficience du programme sans recourir à des données budgétaires (et autres preuves empiriques). Pour lui il est impossible d'attester de l'efficience d'un programme en basant le jugement évaluatif uniquement sur les écrits recensés et le discours des informateurs internes clés (utilisateurs ciblés de l'évaluation).

Extrait de verbatim

« *Comment est-ce qu'on peut faire une section sur la section efficience sans parler de données budgétaires moi ça, ça me dépasse complètement. Quels sont les coûts de gestion, quelles sont les dépenses réelles ? Pour rappel l'efficience c'est un examen, c'est l'efficience, c'est un regard sur les coûts, combien ça coûte pour produire un extrant, combien ça coûte pour avoir un certain résultat. Mais pour ça, ça prend des chiffres, on n'a aucun chiffre là-dedans. Tout ce qu'il dit c'est que la X paye, à partir de la littérature et du discours des gestionnaires pour qui l'évaluation représente des enjeux - oui ça je peux théoriquement le croire, mais ça ne me dit pas si le programme est efficient ou non.* » (S2)

Enfin, remarquons que l'évaluateur 'S3' en discutant de la pertinence du devis d'évaluation, explique que parfois les évaluateurs sont en mesure de recourir à une approche méthodologique plus poussée pour évaluer l'efficience et l'économie, mais que cela dépend de la maturité et de la stabilité du programme, ce qui n'était pas le cas ici. Cela fait d'ailleurs écho à l'une de ses précédentes remarques ainsi qu'à la préoccupation de 'S4' en ce qui a trait à « l'obligation » de répondre à des questions d'efficience et d'économie alors que les évaluateurs n'ont pas accès aux données probantes nécessaires.

4.1.2.1.2 L'analyse de contenu du cadre méthodologique

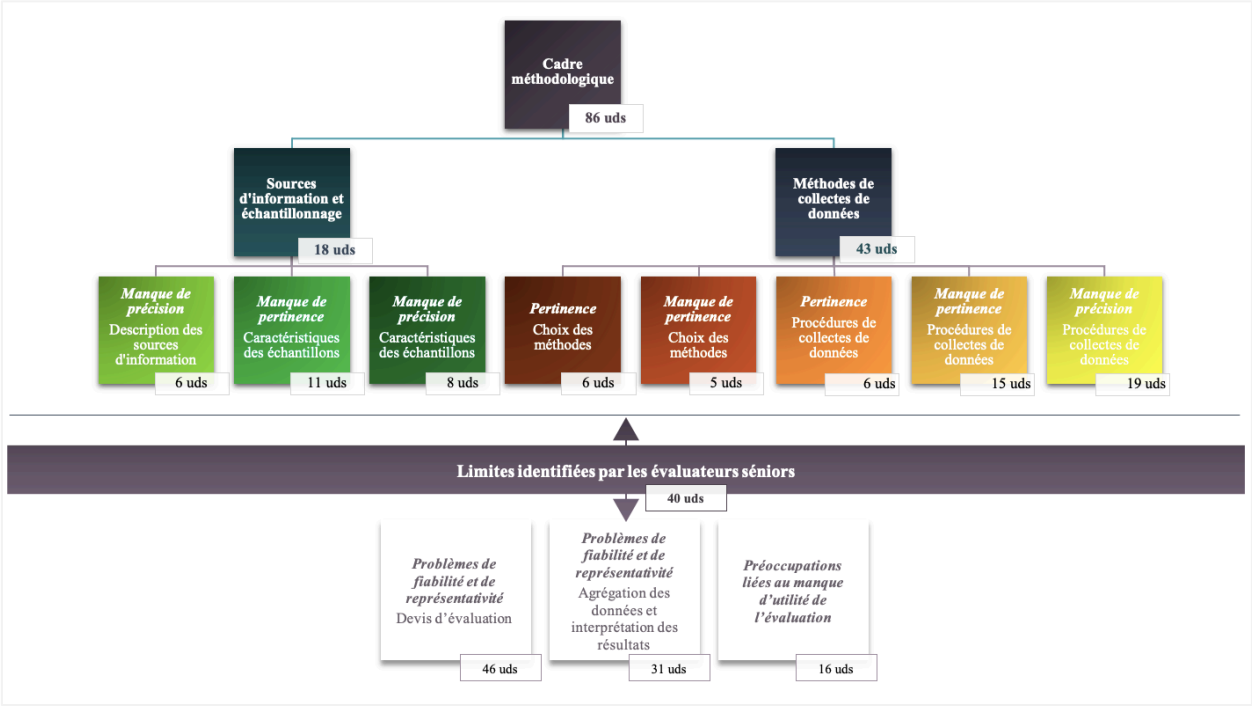


Figure 26. *Principales thématiques et éléments d'analyse du cadre méthodologique (évaluateurs séniors).*

Dans le cadre de *l'appréciation du contenu du cadre méthodologique de l'évaluation* (86 uds), les évaluateurs ont particulièrement concentré leur attention sur les *méthodes de collecte de données* (43 uds), *l'agrégation et l'utilisation des résultats* (28 uds), le tout entrainant l'émergence d'un certain nombre de *limites méthodologiques* (40 uds) qui mettent en exergue les pratiques recommandées par les évaluateurs séniors (Figure 26). Comme nous l'avons mis de l'avant à plusieurs reprises dans la partie portant sur le cadre d'évaluation, les éléments mis en exergues relativement au manque de pertinence et de précision-justesse sur le plan du contenu du cahier des charges demeurent intriqués avec un nombre important de constatations émis sur le plan des limites du cadre méthodologique identifiées par les évaluateurs.

Tout comme ce fut le cas pour le cahier des charges, les groupes de facteurs contextuels ont une *fréquence de cooccurrence tout aussi importante avec les éléments du cadre méthodologique (111 cos)*. En outre, les fréquences desdites cooccurrences sont distribuées de la manière suivante :

- Dyade [Cadre méthodologique - Contraintes de terrain (ESR 1)] : 31.53 % (35 cos).

- Dyade [Cadre méthodologique - Contraintes politiques (ESR 2)] : 46.85 % (52 cos).
- Dyade [Cadre méthodologique - Sensibilité culturelle du milieu] : 21.62 % (24 cos).

Afin d'alléger le texte et éviter la répétition, nous abordons en détail l'influence des divers facteurs sur la pertinence, la précision et la justesse des éléments du cadre méthodologique, dans la section portant sur les limites méthodologiques. En effet, les points critiques saillants du cadre méthodologique, mais aussi du cadre d'évaluation, mettent en exergue les bonnes pratiques des évaluateurs experts et il s'avère très pertinent pour nous d'observer les remarques des participants à la lumière du rôle joué par les facteurs contextuels. Aussi, voici les points critiques saillants du cadre méthodologique mis en exergue par les évaluateurs (Figure 26, p. 194) :

- le manque de pertinence et de précision-justesse de l'échantillonnage des sources d'information (15 uds) ;
- le manque de pertinence des procédures de collectes de l'information (15 uds) ;
- le manque de précision-justesse des procédures de collectes de l'information (19 uds) ; plus particulièrement le manque d'explicitation des procédures (10 uds) ;
- le manque de précision-justesse de l'agrégation des données et de l'interprétation des résultats (21 uds).

En outre, parmi les choix méthodologiques jugés pertinents, les évaluateurs mentionnent :

- le choix des méthodes de collecte de données (6 uds) - leur caractère standard, traditionnel des méthodes choisies (3 uds), et de la complémentarité des méthodes (3 uds) ;
- le déroulement de la collecte de données des études de cas (4 uds) - choix pertinent pour l'accès aux sources d'information, mais aussi culturellement approprié, de sous-traité la tâche à une firme autochtone qui s'est rendue au sein des communautés ;
- et (3) la présence d'une section limite dans le rapport d'évaluation (4 uds).

a. Perceptions de la pertinence et de la précision-justesse des sources d'information et caractéristiques échantillonnales *(16 uds)*

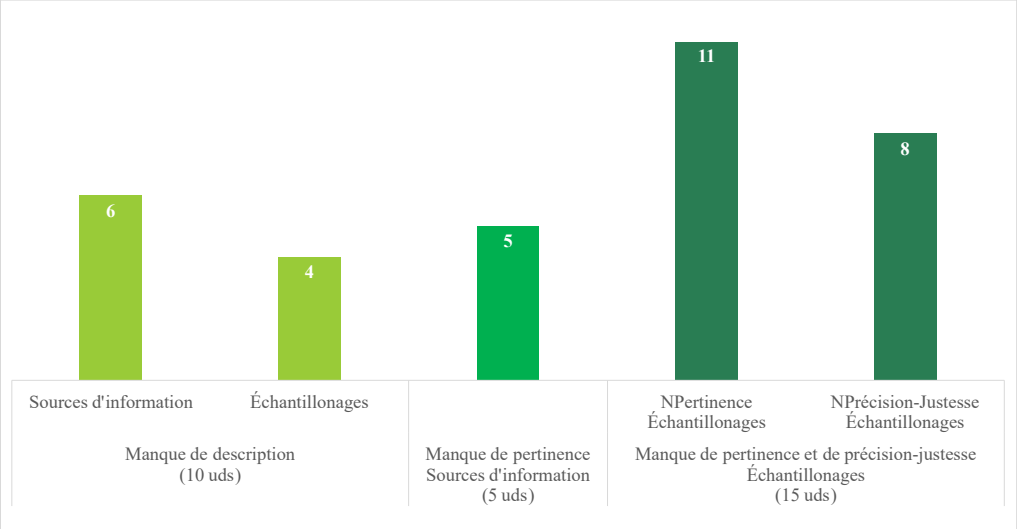


Figure 27. *Thématiques et éléments d'analyse relatifs aux sources d'information et échantillonnages (évaluateurs séniors)*

Le manque de description des sources d'information et de l'échantillonnage (10 uds)

L'évaluateur 'S3' mentionne très tôt dans son analyse que les rapports d'évaluation de programmes fédéraux sont courts et que la section méthodologique excède rarement deux pages. Aussi, si initialement il semblait être satisfait des brèves descriptions des différentes méthodes, il déclare finalement ne pas être en mesure de saisir les procédures d'échantillonnage, ni même de clairement identifier les sources d'information consultées (Figure 27). En effet, celui-ci, de même que 'S1' sont préoccupées par la question de la représentativité des sources d'information, ces derniers ayant noté l'absence de la description du processus de sélection des sources d'information dans le cadre de la recension des écrits scientifiques et des études de cas, de la description des sources d'information et des caractéristiques échantillonnales des différents groupes de participants dans le cadre des entrevues des informateurs clés ainsi que des entrevues réalisées dans le cadre des études de cas. Pour 'S1' l'ajout de la matrice évaluative dans le rapport aurait contrebalancé ce manque de description des sources d'information et de l'échantillonnage.

Extraits de verbatim

"In the Literature review, because there was no reference list attached to the evaluation, it was hard to know how many for instance of these thirty-two documents that they reviewed actually turned out to be relevant enough to make it into the report. And I also thought that they found 32 documents kind of surprising to me given the scope they defined for the literature review. [...]. The scope of it and then the small number of documents just made me kind of curious about the literature. [...]. Often the case with federal evaluations the methodology description is rather short and so it's hard to really don't what people truly did. And it's a little bit. Sometimes they're kind of masking. You know I mean not terribly but you know some of the maybe the limitations of the methodology. I know they discuss them at some points, but some aspects of the methodology are not always fully discussed. Like for the case studies, they were such a key component and it's seemed that the weight evaluation efforts were focused on these four case studies why these four were chosen, I found the description was pretty weak and limited. So, I felt like the justification for why it's focusing on these four was what's kind of missing for me. [...]. They also said for the case studies that they did some interviews individually or in groups in terms of the methods for the case studies. But they didn't say any who was interviewed individually and who was part of a focus group. [...]. I find it interesting that they did not give us a better sense. They interview this number of people in this group individually, and this group with focus groups and why." (S1).

"A couple of comments regarding the description of the key informant interviews. It says that they were conducted with program SH, with different agencies, but it doesn't really tell me you know who was it? Was it the administrative director, was it the manager, was it the staff, the board, etc.? I mean they say it also included partners from the communities, but we don't' clearly see like was it one person that they interview or group of people from each agency and community. And in terms of the case studies, I find that I liked that they put a little rational there to explain why they choose these case studies with that mix of wanting to be representative of the different agencies within the province. But there is also a certain opportunistic element there too. That you want to look at organisations that are, that have data available, that can meet with you. There is that kind of practical aspect there and they mentioned it so that was good. [...]. But they kind of said you know we talk to these kinds of people in general and there was a total of 72. [...]. Who were the 72, what they did, how many per case study, etc.?" (S3).

Le manque de pertinence des sources d'information et de l'échantillonnage (5 uds)

Les différents constats émis par 'S1', 'S2' et 'S3' ont mis en exergue une faiblesse sur le plan de la sélection des sources d'information, et ce, particulièrement sur le plan de l'évaluation du rendement et de l'efficacité du programme. En outre, les évaluateurs regrettent la non-inclusion d'informateurs clés essentiels, tels que :

i. les travailleurs de première ligne - pour 'S1' il s'agit de sources d'information indispensables pour l'évaluation, et mentionne que si ces derniers n'ont pas été consultés par manque d'intrants, ceci aurait pu être palier par le recours à des groupes de discussion, ou encore dans le cadre des études de cas (miser sur la complémentarité des méthodes) ;
ii. les membres de la population ciblée par le programme - 'S2'et 'S3' expliquent que s'il est difficile (voire impossible) d'interroger une catégorie spécifique d'individus de la population autochtone, certains de leurs proches par ailleurs tout aussi concernés par les services offerts par le programme auraient pu, et dû, être consulté dans le cadre de l'évaluation et avoir

l'occasion de faire part de leur expérience et de témoigner directement de sa mise en œuvre et de son efficacité.

iii. les informateurs clés externes - selon 'S3' il aurait été pertinent d'inclure une *« non-program perspectives »,* avec notamment la consultation d'experts dans le domaine ainsi que des organisations gouvernementales et communautaires autochtones.

Extraits de verbatim

"It seems they only spoke with executive directors of different agencies but not to any frontline staff. They could have done focus groups with frontline staff across agencies. Also, since one of the issues was about consistency in how agencies approach things, to me there would be benefits in talking across agencies. But they took things kind of agency by agency and so that seems like a lost opportunity for the evaluation to me." (S1).

« Pourquoi ne pas demander aux communautés autochtones si c'était adapté à leur culture, à leur vision du monde plutôt que de regarder un modèle théorique puis regarder la programmation ? C'est un exercice tout à fait théorique, un plan d'activités, ce n'est pas la réalité. Il faut parler à la population, mais ils ont fait une revue de documents et des écrits. Pourtant, c'est un aspect fondamental. » (S2).

"*I mean, the individual's relatives' perspective, it didn't look like it was incorporated into the case studies in a qualitative way, either through focus groups or interviews. So, I thought, you know, probably for an implementation evaluation committee for prevention plans and activities, especially their kind of community-based awareness, community events and things like that, especially in a topically child welfare, child neglect. Pretty tricky to kind of orchestrate some sort of survey methodology and context. I think."* (S3)

Dans le cas de 'S2' et 'S3', les précédentes constatations émergent, entre autres, de problèmes identifiés au niveau du cahier des charges. À savoir, une dépendance excessive envers les données théoriques comme sources de preuves de mise en œuvre, de rendement et d'efficacité (recension des écrits, plans d'activités) ainsi que l'absence de perspectives comparatives.

Les problèmes de pertinence et de précision-justesse des procédures d'échantillonnage (15 uds)

Dans le cadre de la recension des écrits scientifiques, les évaluateurs 'S1' et 'S4' ont remarqué qu'il y avait un problème d'alignement entre les objectifs de la recension des écrits et les caractéristiques de l'échantillon des écrits recensés. En effet, ces derniers remettent en question la décision de recenser des écrits sur ce que proposent d'autres pays en matière d'intervention au sein des communautés autochtones relativement à la problématique visée par les interventions du programme évalué. Pour 'S1', une telle décision s'avère non justifiée, particulièrement en ce qui a trait à l'observation tant au Canada qu'à l'étranger d'interventions auprès de la population non autochtone. De son côté, 'S4' a remis en question la décision de recenser des écrits scientifiques « *de ce qui se fait dans le monde* », car hormis le manque de pertinence d'un tel choix méthodologique dans le cadre d'une évaluation de la mise en œuvre, ce dernier juge la décision des évaluateurs trop ambitieuse au regard de la complexité de la problématique à l'étude et les

intrants disponibles. Aussi, compte tenu de l'envergure de la problématique visée par les interventions du programme, et la portée attendue de la recension des écrits, 'S1' et 'S4' ont été surpris que seulement une trentaine d'écrits aient été recensés. 'S1' s'interroge d'ailleurs sur le processus de sélection desdits articles. 'S4' va plus loin dans son analyse et explique que le manque de pertinence des objectifs de la recension des écrits a entrainé un manque de précision-justesse sur le plan des procédures et caractéristiques échantillonnales. En effet, l'échantillon des écrits recensés s'avèrent finalement de taille restreinte, et non-aligné avec la portée attendue de la recension des écrits (portée jugée non nécessaire), du fait notamment du nombre très limité d'interventions recensées à l'étranger et du manque de cohérence dans le choix des écrits.

Extraits de verbatim

"In the Literature review, how many for instance of these thirty-two documents that they reviewed actually turned out to be relevant enough to make it into the report. And I also thought that they found 32 documents kind of surprising to me given the scope they defined for the literature review. And the scope that they defined for the literature review to me seemed overly broad. They talked about looking at the state of Aboriginal and not aboriginal child welfare in Canada and abroad. And I was really wondering why some of that would be relevant for this evaluation. For instance, why would a non-Aboriginal intervention in Canada or abroad really be relevant? [...]. It is a bit over broad and unjustified. The scope and the small number of documents just made me kind of curious about the literature." (S1).

« Comment fixe-t-on les paramètres d'une telle recension quand on fait face à une problématique aussi complexe? Et puis là, ils mentionnent au pays et ailleurs dans le monde. Et j'ai trouvé que c'était malheureux, car ils ne devraient pas créer des attentes comme celle-là. Parce qu'ils n'ont pas regardé ce qui s'est fait ailleurs dans le monde. Quelques études aux États-Unis j'ai vu, puis dans un autre pays. C'est correct, mais ce n'est pas une revue de ce qui se fait dans le monde. Ils auraient dû resserrer leur énoncé et dire que le sujet est tellement immense que pour mieux y arriver ils vont se concentrer sur certains points précis et sur des écrits qui sont applicables au Canada. Là, en autochtonie, on a en Australie, aux États-Unis, même en Amérique du Sud. Et puis on parle ici d'une évaluation de la mise en œuvre, donc la revue des écrits n'a pas besoin d'être très grande. Si c'était l'évaluation des résultats, de montrer que les jeunes vont mieux à cause du programme ben là peut-être qu'on peut élargir la recension, aller voir la théorie du changement, les questions les plus fondamentales, sur qu'est-ce qui amène un changement. » (S4)

b. L'appréciation de la pertinence et de la précision-justesse des méthodes de collectes de données *(48 uds)*

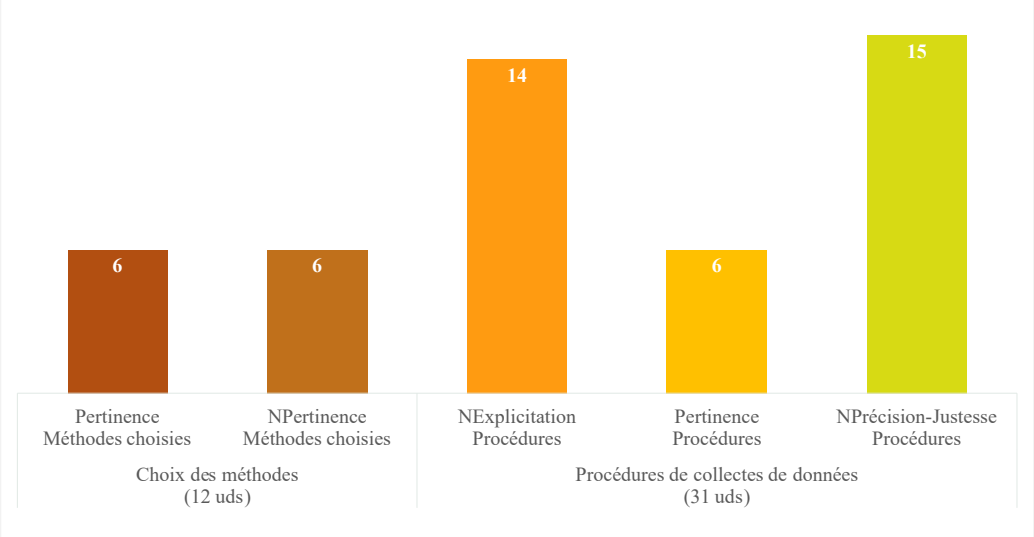


Figure 28. *Thématiques et éléments d'analyse sur les procédures de collectes de données (évaluateurs séniors)*

La pertinence des méthodes choisies (6 uds)

Nous l'avons mentionnée à plusieurs reprises, la plupart des décisions méthodologiques sont prescrites par la politique évaluative, la demande de propositions ou encore les préférences et les attentes des commanditaires. Aussi, il n'est pas surprenant que les évaluateurs séniors aient décrit le choix des méthodes comme traditionnel, et que ces derniers ne se soient guère attardés sur cet élément cette décision. 'S1', 'S2' et 'S3' souligneront tout de même le potentiel pour une triangulation des méthodes afin de pallier notamment les contraintes inhérentes au terrain. Nous verrons par la suite que ce fût une occasion manquée.

La remise en question de la pertinence des méthodes choisies (6 uds)

Selon les évaluateurs 'S2', 'S3' et 'S4', la complexité de la problématique ciblée par les interventions du programme et les caractéristiques du milieu ont une influence assez importante sur la pertinence des méthodes de collectes de données. Aussi, nous avons déjà abordé :

- la difficulté de définir les paramètres d'une recension des écrits scientifiques tout en assurant la pertinence des sources d'information et la précision-justesse des caractéristiques échantillonnales (portée de la recension, critères de sélection, tailles des échantillons, etc.) ;

- le manque de possibilité d'innover, ou encore de recourir à des méthodes plus coûteuses pour tenter de remédier aux problèmes d'accès aux données ;
- ainsi que l'absence de modes de consultation de la population ciblée par le programme ainsi que d'informateurs clés externes notamment au sein d'organismes communautaires et autres organisations.

Extrait de verbatim

"Also, I wondered if there could have been an opportunity for more external perspectives. There were the agency interviews with X individuals, provincial representative in some other organization authorities. What about external like academics, social organizations? Often in these contexts we would also include provincial and federal indigenous organizations, etc. It's some examples of increasing the pool of key informants by including non-program perspectives. Same thing at the community level. And maybe this could have been addressed in the case study again kind of stepping a bit outside of the program to incorporate that external perspective. So, partner organizations they work with, organizations in the community. If they could have done that a little bit in the case studies that would have been good. [...]. Or like they use Dr. X perspective from the literature but it's a good example of a relevant external key informant they could have interviewed, even if they are not experts of the program, they have some knowledge of it." (S3).

Le manque de précision-description des procédures de collectes de données (14 uds)

Dans la lignée des observations concernant le manque de description des sources d'information et procédures d'échantillonnage, les évaluateurs, plus particulièrement 'S1' et 'S3', déplorent *l'absence d'explicitation des procédures de collectes de données (9 uds)* (Figure 28, p. 200). Pour 'S4', si la description des procédures de collectes de données est en effet lacunaire, ce dernier n'en est en revanche pas surpris, puisque le caractère succinct du cadre méthodologique rapporté dans le rapport final est une attente standard des agences fédérales commanditaires. De fait, cet évaluateur sénior précise que selon lui cela ne signifie en aucun cas un manque volontaire de transparence, même s'il reconnait avoir vu, et produit, des rapports fournissant davantage d'explicitation des procédures méthodologique.

Si 'S1' et 'S3' reconnaissent que la section méthodologique dans ce type de rapport final est traditionnellement courte, il n'en demeure pas moins que dans le cas présent, les évaluateurs ils ont eu du mal à apprécier la pertinence et la justesse des décisions méthodologiques. Voir même à cerner la façon dont les informations ont été collectées dans certains cas. Ce problème est d'ailleurs renforcé, selon 'S1', par *l'absence du modèle logique, des matrices (des indicateurs de mesure) et de l'instrumentation (5 uds)*.

Extraits de verbatim

"So, in that way it kind of gets sometimes you sort of not intentionally masking things but you're just trying to be brief and so some things are not fully stated I guess I should say. And with the evaluation matrix it wasn't attached to the report. So, I guess that's not standard or and see you know every department is different. A department that I do a lot of work for always attaches the full matrix, the logic model, the instruments, everything is there. It's a little bit hard to review a report when you don't see the full matrix especially, so you don't see all the indicators they listed the question but the questions, the evaluation questions are always very broad. So, it's hard to really know what they think are evidence for these questions." (S1).

« Que les méthodes d'analyse ne soient pas décrites, ce n'est pas inhabituel. C'est sûr qu'en termes de description des méthodes c'est très laconique, vraiment pas beaucoup de détails. On fait juste nous dire dans le fond, voici ce qu'on a fait, 20 entrevues, tant d'étude de cas, etc. J'ai vu des rapports avec des sections méthodologiques beaucoup plus avancés sur les procédures, les choix, les justifications, etc. Mais là on n'a aucune idée. Est que la firme autochtone a participé aux analyses ? On ne sait pas. Les procédures de triangulation, on ne s'est pas non plus. Comme je l'ai dit, c'est très laconique. On n'a aucune idée là. [...]. C'est probablement parce que le client ne voulait pas plus. Ce n'est pas inhabituel. Au fédéral c'est souvent comme ça. Moi aussi, si personne ne me demande d'élaborer, je mettrais juste ce qu'il faut. Ils n'ont pas caché leur métho, ils ont dit comment ils ont fait. » (S4)

Le choix pertinent de sous-traiter certaines collectes de données à une firme autochtone (4 uds)

Nous l'avons déjà abordé dans le cadre de la discussion sur le cahier des charges, pour l'ensemble des évaluateurs séniors, la décision de sous-traiter la collecte de données de différentes études de cas fut un choix d'une extrême pertinence. Un choix par ailleurs qui n'est pas nécessairement habituel pour 'S2' et 'S4' qui mentionne avoir réalisé de nombreuses collectes de données auprès de communautés autochtones. Aussi, outre la facilitation de l'accès aux sources d'information, ceci renforcerait également l'ampleur ainsi que l'exactitude des données collectées.

Extrait de verbatim

« Ils ont embauché une firme autochtone pour les faire. C'était intéressant parce que ça n'a pas toujours été comme ça. J'en suis la preuve. J'ai souvent fait des études de cas en milieu autochtone. Ça, c'était une bonne pratique. » (S4)

Le manque de pertinence et de précision-justesse des procédures de collectes de données (15 uds)

Une fois de plus, la majorité des éléments mis en exergues ici ont été abordés plus haut, que ce soit au sujet du contenu du cahier des charges, ou encore l'appréciation des sources d'information et des échantillonnages. Aussi, parmi les éléments remettant en question la précision-justesse des procédures de collectes de données (nous abordons leurs conséquences dans la section portant sur les limites identifiées par les évaluateurs), nous retrouvons :

- l'absence de perspective comparative dans le cadre de l'évaluation de la conception et de la mise en œuvre | 'S2' et 'S3' (p. 179-190) ;

- l'absence de perspective de la population ciblée par le programme, de travailleurs de premières lignes et d'informateurs clés externes | 'S1', 'S2' et 'S3' (p. 197-198) ;
- les problèmes d'alignement de la recension des écrits avec les objectifs de la collecte de données - remise en question de la portée irréaliste de la recension des écrits | 'S1' et 'S4' (p. 198-199) ;
- l'absence de recueil de preuves tangibles (données probantes) lors de l'évaluation de l'efficacité - une démonstration théorique de la performance des interventions du programme, ou basée uniquement sur la perception des informateurs clés | 'S2' et 'S3' (p. 190-191) ;
- et, l'absence de recueil de données empiriques sur l'efficience et l'économie - remise en question de la rigueur des procédures méthodologiques | 'S2' (p. 192-193).

'S3' et 'S4' remettent également en question certaines procédures de passation (modes de consultation). Pour 'S3' il s'avérait important au regard de la problématique visée par les interventions du programme d'inclure la consultation d'experts et (ou) d'informateurs clés externes. Selon lui, les évaluateurs auraient pu interviewer les auteurs cités dans la recension d'écrits plutôt que de se concentrer sur un ou deux de leurs articles (surtout lorsqu'ils sont un nombre assez restreint). 'S4', au regard de son expérience, remet quant à lui en question le choix de réaliser des entrevues par téléphone. En effet, si les évaluateurs ne précisent pas dans le rapport s'il s'agissait des entrevues avec les informateurs clés ou les entrevues réalisées dans le cadre des études de cas, il ne s'agit pas d'un mode de passation efficace auprès de la population autochtone, ou les populations sensibles de manière générale – « *leur réalité est tellement complexe, on ne réalise pas vraiment ce qui se passe de l'autre côté de la ligne au téléphone. Les fois où je me suis rendu sur place, on a rencontré les organismes, on était présent dans la communauté, et ils pouvaient vraiment nous parler de leur expérience. On tisse un lien interpersonnel, ça rend les entrevues plus riches. Se déplacer dans les communautés ça permet de bien mieux comprendre leur réalité.* »

L'axe d'analyse sur la synthèse et l'interprétation des résultats a généré des limite importantes, nous l'abordons directement dans la partie suivante qui traite des limites identifiées par les évaluateurs. Il est important de mentionner que les choix méthodologiques remis en question face à l'observation de l'insuffisance voire de l'absence de preuves tangibles sur le plan des constatations soulèvent d'importantes questions : est-ce un problème de communication de l'évaluation ? Est-ce un problème de précision-justesse au moment de la synthèse des résultats

rapportés ? Est-ce uniquement dû à un problème d'accès aux données ? Ou s'agit-il tout bonnement d'un manque de rigueur méthodologique (mauvaise définition des indicateurs et (ou) mauvaise sélection de la documentation administrative, choix de questions d'évaluation auxquelles on ne pas répondre de manière appropriée du fait du manque de données administratives/de monitorage du programme, etc.).

4.1.2.2 Les limites méthodologiques identifiées par les évaluateurs séniors (experts)

La description des limites identifiées constitue un excellent moyen de mettre en exergue les pratiques recommandées par les évaluateurs séniors. On trois axes de préoccupation dans le discours de ces derniers :

- Axe 1 : les préoccupations liées à la fiabilité, validité et représentativité sur le plan du devis d'évaluation - 46 uds.
- Axe 2 : les préoccupations liées à la fiabilité, validité et représentativité sur le plan de la synthèse et de l'interprétation des résultats - 31 uds.
- Axe 3 : les préoccupations liées à l'utilité de l'évaluation, et de ses résultats - 16 uds.

Les évaluateurs séniors n'ont pas fait preuve du même degré de remises en question en ce qui a trait aux différents axes, et ils ne sont pas nécessairement concentrés sur les mêmes éléments. S'ils ont tous évoqué des limites en lien avec l'axe 1, l'axe 2 s'est avéré la plus grosse préoccupation de 'S2' et de 'S3' qui se sont attardés sur le manque de preuves empiriques et la *méreprésentation* des données dans le cadre de la formulation des constatations. L'axe 3 constitue un élément important de l'analyse de 'S1' qui déplore le caractère inadéquat du devis d'évaluation vis-à-vis de la sensibilité culturelle du milieu. Quant à 'S4', nous remarquons qu'il encadre ces constats selon d'une part la culture évaluative fédérale et les utilisateurs ciblés de l'évaluation, et d'autre part, la population du programme et les caractéristiques spécifiques de leur communauté. Un tel processus conduit l'évaluateur à accepter certaines choses dans un cas, tout en les remettant en question dans la perspective du second.

Comme nous l'avons annoncé dans la partie portant sur l'analyse du cadre méthodologique, la description des limites identifiées pour chacun des axes se fait de manière conjointe avec l'examen de leurs cooccurrences avec les différents groupes de facteurs contextuels (Figure 29).

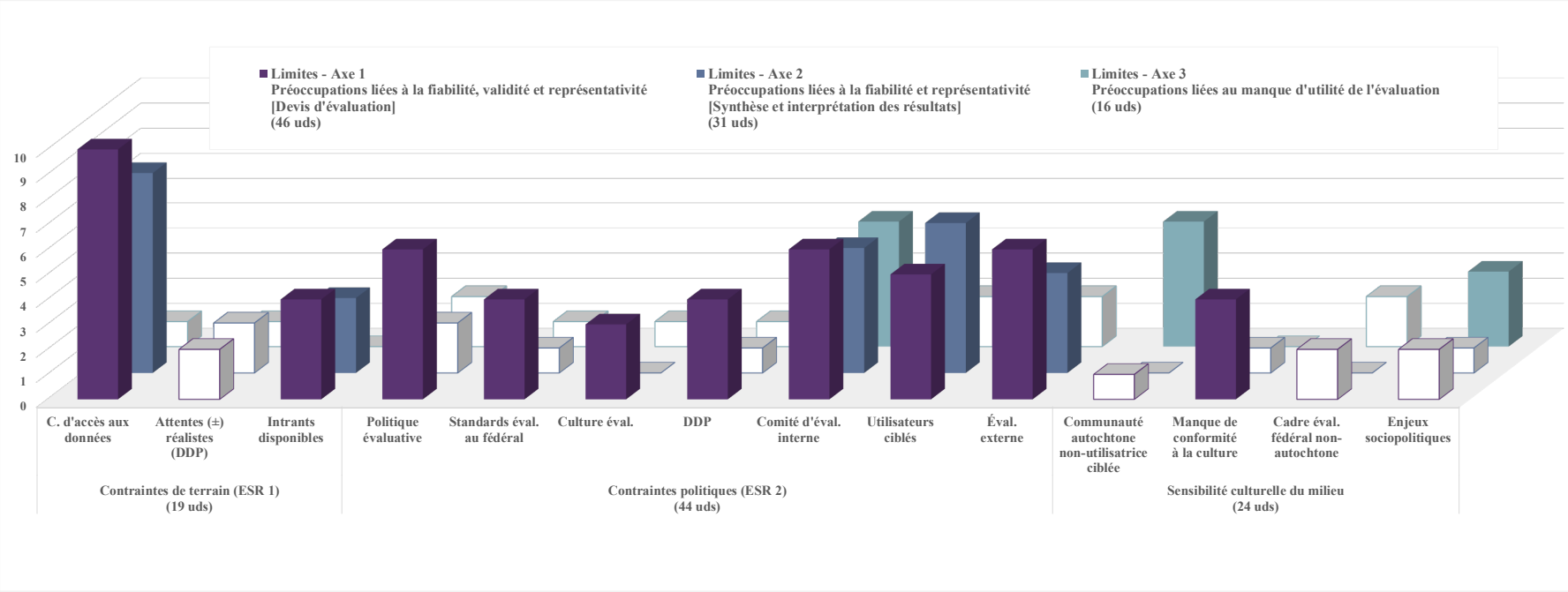


Figure 29. *Distribution des cooccurrences dans le cadre des interactions entre les différents axes de limites et les facteurs contextuels*

a. Limites - Axe 1 : les préoccupations liées à la fiabilité, validité et représentativité sur le plan du devis d'évaluation *(46 uds)*

Ce premier axe de limites identifiées par les évaluateurs séniors n'est autre que la « résultante » des points critiques saillants relatifs aux éléments clés du plan d'évaluation émis par ces derniers. Aussi, cet axe traite : 1) du manque de transparence sur le plan du cadre méthodologique dû aux manques de description et d'explicitation des procédures de sélection des sources d'information, et des procédures de collectes de données ; et 2) du manque de pertinence et de précision-justesse des éléments du devis d'évaluation, particulièrement au niveau du cahier des charges, du cadre méthodologique - le tout remettant ainsi en question de la rigueur du processus évaluatif (potentielle *mésévaluation*). Le Tableau 14 illustre les cooccurrences entre les limites identifiées dans cet *axe 1* et les différents groupes de facteurs contextuels.

Tableau 14. *Description des cooccurrences entre les préoccupations liées à la fiabilité, validité et représentativité sur le plan du devis d'évaluation et les facteurs contextuels (évaluateurs séniors)*

		Manque de transparence		**Manque de pertinence et de précision-justesse**		
		Sources d'information et échantillonnages (8 uds)	Procédures de collectes de données (6 uds)	Cadre d'évaluation (7 uds)	Cadre méthodologique (9 uds)	Mésévaluation (9 uds)
Contraintes ESR	Contraintes d'accès aux données	3	5	3	9	6
	Intrants disponibles	0	0	4	1	2
	Attentes (±) réalistes des commanditaires	0	0	3	2	0
Contraintes politiques	Politique d'évaluation IGF	0	0	2	1	2
	Standard de l'évaluation fédéral	4	4	1	2	0
	Comité d'évaluation interne	2	2	5	3	4
	Résistance des utilisateurs ciblés	0	0	4	3	6
	Évaluateur externe	2	2	2	2	7
SCM	Problèmes de conformité à la culture de la population	1	1	2	2	3
	Enjeux socio-politiques	1	1	2	2	3

Le manque de transparence sur le plan des procédures méthodologiques

Si les évaluateurs séniors ne se sont pas autant penchés sur le manque d'explicitation des procédures du cadre méthodologique comparativement aux évaluateurs juniors, 'S1' et 'S3' se sont tout de même attardés sur le manque de transparence sur le plan des sources d'information, des procédures d'échantillonnage ainsi que des procédures de collectes de données (Figure 30). Un tel constat souligne la présence de potentiels biais. En effet, pour 'S3' ce manque de précision et de clarté indique de potentielles barrières à la méthodologie telles que des problèmes d'accès aux données, ou encore l'influence des utilisateurs ciblés sur certaines prises de décision.

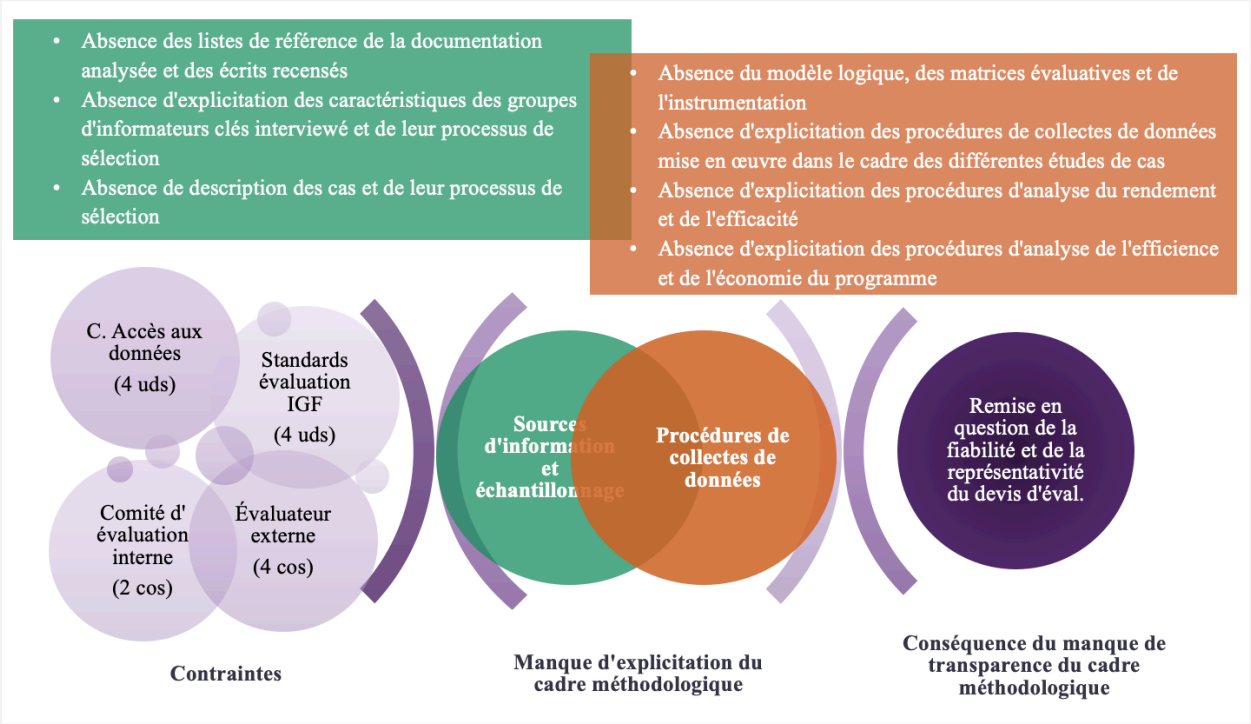


Figure 30. *Thématiques et éléments d'analyse sur la relation entre le manque d'explication du cadre méthodologique (et ses conséquences), et certains facteurs contextuels clés (évaluateurs séniors)*

Enfin, si les évaluateurs ne semblent pas surpris du manque de description sur le plan du cadre méthodologique du fait des standards de communication de l'évaluation des programmes fédéraux, il n'en demeure pas moins qu'un certain niveau de description est essentiel pour assurer la fiabilité et la représentativité de l'évaluation. Nous avons d'ailleurs vu plus tôt, que certains organismes fédéraux exigent l'inclusion du modèle logique, des matrices et de l'instrument dans le rapport final, et qu'à la lumière des informations qu'il a, 'S1' n'est pas en mesure de juger de la qualité de l'évaluation. En revanche, malgré son constat du manque d'explicitation de certaines

procédures méthodologiques, cela ne semble pas être problématique dans le présent contexte, selon 'S4'. En effet, pour lui, les évaluateurs ont répondu aux attentes habituelles des clients et des standards de l'évaluation au fédéral, et il ne s'agit pas d'un manque volontaire de transparence de leur part. En revanche, il mentionne tout de même qu'il aurait été pertinent de rendre compte des efforts de triangulation. Quant à 'S2', nous verrons qu'il s'agit d'une combinaison des attentes des clients et du processus itératif de réécriture initié par ces derniers, et du manque de rigueur des évaluateurs externe.

Extraits de verbatim

"For the full report, it would have been useful to get to see the full evaluation matrix. I don't know if they say you don't attach that to their evaluation reports I find that very helpful to get a better sense of what they're wanting the methods to do. But the indicators are for the evaluation questions. And they also didn't attach the instruments that would have been useful in terms of assessing whether the methodology seemed to work from my perspective. I'm a little limited in the ability to do that. Again, often the case with federal evaluations methodology descriptions are rather short and so it's hard to really don't what people truly did. Sometimes they're kind of masking. I mean not terribly but you know some of the maybe the limitations of the methodology. I know they discuss them at some points, but I found the description was weak and limited." (S1).

« *En termes de description des méthodes, c'est très laconique, vraiment pas beaucoup de détails. On fait juste nous dire dans le fond, voici ce qu'on a fait, 20 entrevues, tant d'étude de cas, etc. J'ai vu des rapports avec des sections méthodologiques beaucoup plus avancés sur les procédures, les choix, les justifications, etc. Mais là ici on n'a aucune idée, les procédures de triangulation, on ne s'est pas non plus. [...]. C'est sûrement parce que le client ne voulait pas plus. Ce n'est pas inhabituel. Au fédéral c'est souvent comme ça. Moi aussi, si personne ne me demande d'élaborer, je mettrais ce qu'il faut. Ils n'ont pas caché leur métho, ils ont dit comment ils ont fait. Au fédéral je suis sûr que ça passe sans problèmes, mais dans d'autres contextes je suis sûr qu'ils diraient que c'est inacceptable on a besoin de beaucoup plus de précision, de détails sur les procédures, etc.* » (S4)

Le manque de pertinence et de précision-justesse du devis d'évaluation

Comme le montre le Tableau 14 (p. 206), il s'agit ici des limites résultantes de l'analyse de la pertinence et de la précision-justesse du cadre d'évaluation (plus particulièrement de son produit final - le cahier des charges), et du cadre méthodologique (plus précisément de la méthodologie des collectes de données). Aussi, la Figure 31 (p. 209) présente les principales remises en question méthodologiques relevées dans le discours des évaluateurs (et explorées jusqu'à présent), et l'influence potentielle des différents groupes de facteurs contextuels.

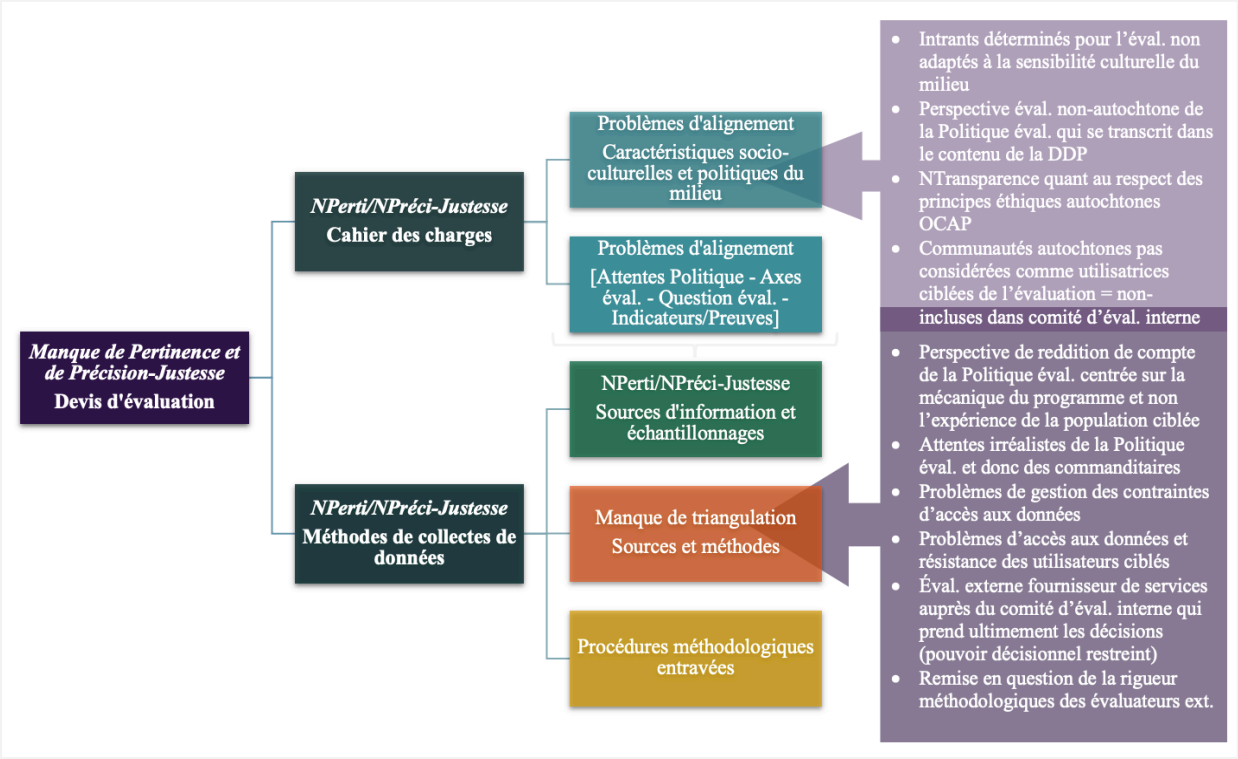


Figure 31. *Thématiques et éléments d'analyse sur la relation entre le manque de pertinence et de précision-justesse sur le plan du contenu du devis d'évaluation et certains facteurs contextuels clés (évaluateurs séniors)*

La préoccupation centrale qui émerge ici porte sur la démonstration théorique de la pertinence du programme ainsi que du rendement et de l'efficacité de la mise en œuvre de ses interventions. En effet, 'S2' et 'S3' ont constaté à plusieurs reprises que les évaluateurs semblaient baser presque exclusivement leur jugement évaluatif sur l'analyse du contenu des plans de développement du programme ainsi que la revue des écrits scientifiques, et non sur des données probantes relatives à la réalisation des activités du programme. 'S2' déclare que « *c'est tout à fait théorique comme exercice, un plan d'activité ce n'est pas la réalité* » et que, « *il faut parler aux autochtones, au lieu de se concentrer sur une revue de documents et des écrits* ». 'S3' le rejoint dans son raisonnement, expliquant que les évaluateurs se sont concentrés sur la théorie du programme et non sur ce qui s'est réellement passé. La principale raison serait ici l'absence de données de monitorage interne de la mise en œuvre des activités du programme et de leur performance. Or, s'il s'agit d'une contrainte d'accès aux données courantes. 'S2' et 'S3's'attendaient à ce que les évaluateurs recourent à la triangulation des sources d'information et des méthodes pour pallier le mieux possible à cette limite; et (ou) qu'ils attestent des pistes de

solution qui ont été envisagées et pourquoi cela fut impossible. Parmi les solutions possibles, 'S2' et 'S3' ont d'ailleurs mentionné :

- La constitution d'une base de données à partir de l'analyse documentaire ;
- La consultation des parties prenantes clés de la communauté, de population ciblée afin qu'ils partagent leur perspective, leur expérience, et que les évaluateurs puissent observer par exemple, dans quelle mesure les services offerts (et la manière dont ils sont mis en œuvre) sont-ils adaptés à leur réalité culturelle, à leur vision du monde ;
- La consultation du personnel de première ligne afin qu'ils rendent compte de l'efficacité des activités du programme du fait de leur expérience sur le terrain (mentionné également par 'S1').

Si pour 'S2' il est question d'un manque de rigueur méthodologique, pour 'S3' il s'agit davantage d'un problème de gestion des contraintes d'accès aux données, particulièrement d'une définition initiale stricte des paramètres d'utilisation des intrants (et des intrants disponibles), par les clients, ce qui aurait pu limiter les possibilités de consolidation du cadre méthodologique.

Aussi, ce dernier rappelle d'ailleurs que les évaluateurs externes sont des fournisseurs de services avec un pouvoir décisionnel restreint, le dernier mot appartenant au comité d'évaluation interne. Il reproche donc davantage un manque de transparence qui entraine la suspicion de biais potentiels. En effet, après réflexion, 'S3', tout comme 'S2', se sont interrogés sur la possible résistance des utilisateurs ciblés, résistance qui aurait pu entraver les procédures de collectes de données, particulièrement l'accès aux sources d'information. Tel quel, le devis d'évaluation ne permet pas d'assurer la représentativité du programme, et ne répond pas aux attentes du mandat. En revanche, nous avons vu que 'S4' ne s'attendait pas à ce que les procédures de collectes de données visent à collecter des données démontrant « *dans quelle mesure l'approche a contribué à l'atteinte des résultats visés* », puisque l'évaluation visait davantage à rendre compte de la bonne conceptualisation, ou non, des interventions du programme, et des facteurs facilitateurs de mises en œuvre. Pour lui, il n'était donc pas question de mesurer le rendement, ou de compter sur le recueil de preuves tangibles et de suffisamment de résultats probants, la mise en œuvre du programme étant trop récente (2 ans). 'S4' ne fut donc pas surpris de l'approche théorique dans le cadre de la démonstration de la pertinence et de l'efficacité du programme.

Dans la même lignée, nous avons vu plus tôt que les évaluateurs ‘S2’ et ‘S4’ ont remarqué que les constatations relatives à l’efficience et l’économie du programme reposaient majoritairement sur le recueil de données théoriques, et non de preuves empiriques. Aussi, pour ‘S4’ l’inexistence de ce type de données probantes sur le terrain est un problème bien connu dans le présent contexte, problèmes que les clients (et la politique de l’IGF) semblent persister à ignorer. Pour lui, ces attentes irréalistes ont poussé les évaluateurs à conceptualiser l’efficience différemment, stipulant que « *faire de la X est moins coûteux que de faire de l’intervention* », et ce, à partir de la recension d’écrit sur la problématique visée par le programme. Or, si ‘S4’ ne dénigre pas la décision des évaluateurs de se reposer sur la collecte de données théoriques, il n’en demeure pas moins qu’ils ne répondent pas au mandat d’évaluation en ce qui a trait aux attentes de la politique de l’IGF en matière d’efficience et d’économie. Il y a donc un problème de fiabilité et de représentativité des procédures de collectes de l’information à ce niveau du devis, mais aussi de l’agrégation des données et de l’interprétation des résultats (*limites - axe 2*).

Extraits de verbatim

“*There is a lack of balance. And so, I found there are a lot of literature and documents in there saying cultural tailored programming is important. But the question is, is this element being used in the design and delivery to help achieved the outcomes. And so, there is at the end of the paragraph a little bit about what’s really going on at the business level. But kind of where the weight is. Like you know I think it’s important to demonstrate that this is important to successful programming. So yeah, I was really surprised by that because they had the opportunity to inform that with the key informant interviews, the case studies. I don’t know if there was something specifically asked about the business plans, but I think they should have said this is how these agencies are culturally tailoring their program. I don’t know how they engaged the program in the community, how the prevention campaign and the material they distributed were culturally appropriate. I guess what their approach was in terms of helping the community, the parents, etc. Where are the examples of how it’s happening?*” (S3).

Le risque de mésévaluation et de mésusage

Le risque de *mésévaluation* émerge assez logiquement des constatations relatives au manque de transparence au niveau des procédures méthodologiques et des raisons sous-jacentes aux choix posés, de leur manque de pertinence et (ou) de précision-justesse. En effet, si dans le cas de ‘S1’ le manque d’information l’empêche de se prononcer sur la qualité de la démarche évaluative, pour ‘S2’ il s’agit tout bonnement d’un manque de qualité de la méthodologie évaluative, voire de compétence et (ou) de rigueur des évaluateurs externes. Ce dernier ajoute d’ailleurs que les problèmes qu’il a identifiés sur le plan de l’exactitude des éléments de définition du programme couplé avec l’absence du modèle logique remettent en question le caractère approprié du cahier des charges et des choix méthodologiques associés. En revanche, pour ‘S3’ il s’agit davantage d’un potentiel problème de résistance des utilisateurs ciblés et du détournement

du processus évaluatif (mésusage). Quant à ‘S4’, nous avons vu à plusieurs reprises que la persistance d’attentes irréalistes de la part de la politique et des commanditaires est l’unique raison pour laquelle les évaluateurs ont détourné l’un des objectifs évaluatifs (forme de mésusage non intentionnel).

Extrait de verbatim

“Federal evaluations are led by internal evaluators outside the program. Even though they're within the department, they're still supposed to be objective, external. Often when facing these types of conundrums, we rely on their help or advice like evaluation colleagues. Like when the program doesn't want us to talk to our stakeholders X, Y and Z here. We can talk to them, ask them why. If there is a legitimate kind of challenge there. Or is it just kind of trying to hide some dirty laundry or whatever? Then you can kind of say, well, now it's important that these people were included for these reasons. And that can be a sort of negotiation. But I mean, I think this is against the federal evaluation. I sometimes find programs. There is a sort of self-interest of the program that directs like I want my program to look great. And it happens. But often we see that program managers are aware of the weaknesses and limitations of their program and want an evaluation to cover that and make recommendations so they can get their program fixed. Sometimes they're saying the same thing to their managers or senior managers, but they're not getting any kind of traction in response to their request. And so, an evaluation, you know, that kind of provides a way to objectively look at their program and make recommendations that often align with things that they know and understand about their program.” (S3).

b. Limites - Axe 2 : les préoccupations liées à la fiabilité et la représentativité sur le plan de la synthèse et de l’interprétation des résultats *(29 uds)*

Il est question ici d’un point saillant de l’analyse critique des évaluateurs ‘S2’ et ‘S3’ - le manque de transparence, de pertinence et de précision-justesse sur le plan des résultats mobilisés pour soutenir les constatations évaluatives (formulation des réponses aux questions d’évaluation). Aussi, ces évaluateurs se sont penchés sur : 1) la faiblesse/l’insuffisance des preuves, et 2) la mérepréstentation et le mésusage des données/résultats, *évaluatives*. Les limites relatives à ce deuxième axe ainsi que leurs cooccurrences avec les différents groupes de facteurs contextuels sont exposées dans le Tableau 15 suivant.

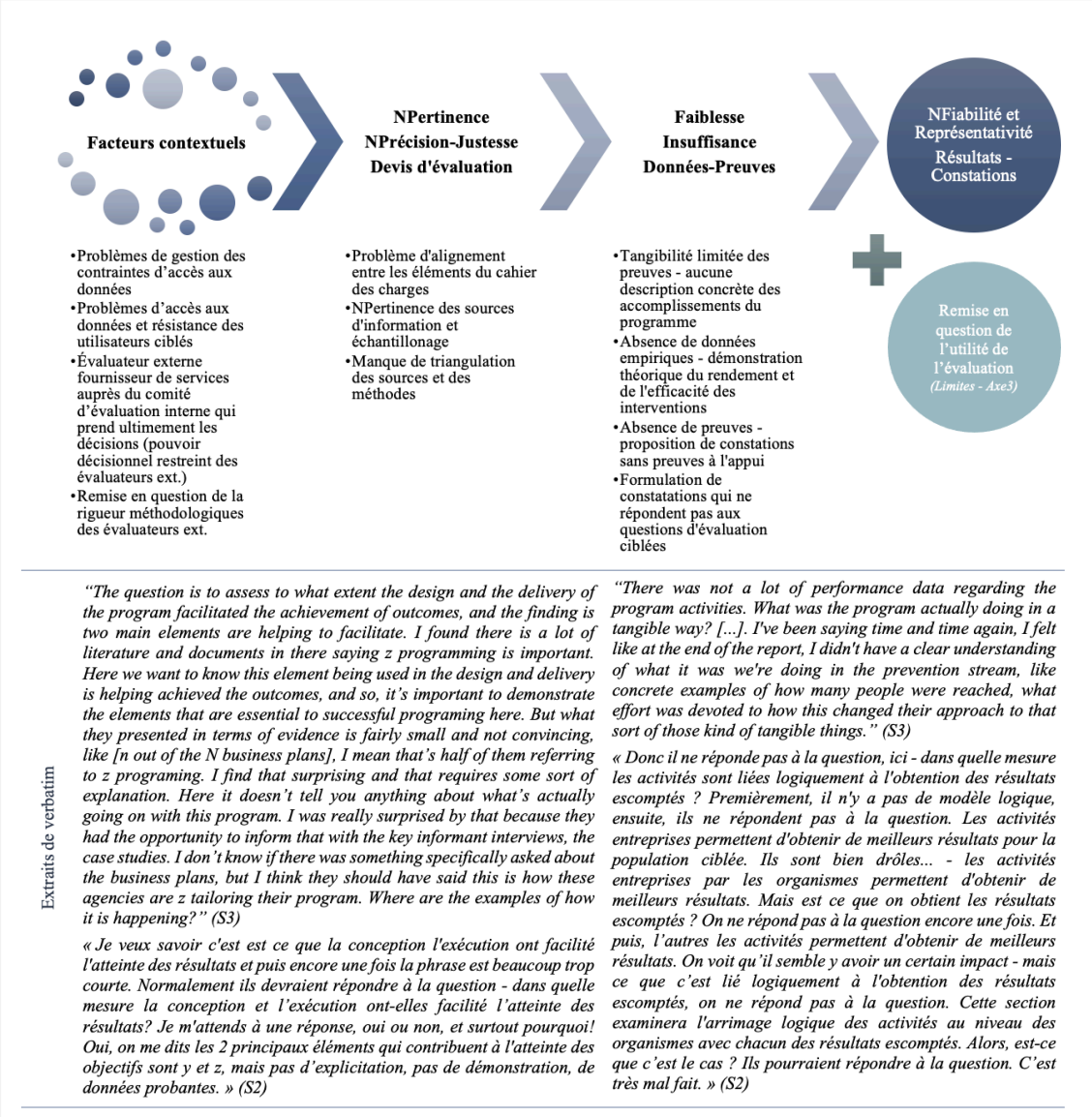
Tableau 15. *Description des cooccurrences entre les préoccupations liées à la fiabilité, validité et représentativité sur le plan de la synthèse et de l'interprétation des résultats, et les facteurs contextuels (évaluateurs séniors 'S2' et 'S3')*

**Limites identifiées principalement par 'S2' et 'S3'*		**Faiblesse/Insuffisance des preuves**				**Manque de transparence sur l'appréciation des résultats (« *Méreprésentation* » des résultats)**		
		NTangibilité des preuves (11 uds)	Absence de données empiriques (8 uds)	Absence de preuves (3 uds)	Résultats ne répondent pas au mandat (13 uds)	Manque de cohérence [données-conclusions] (4 uds)	Exagération/ Manipulation des données-résultats (8 uds)	"*Data cherry-picking*" (4 uds)
Contraintes ESR	Contraintes d'accès aux données	3	2	2	4	0	3	2
	Intrants disponibles	2	0	2	4	0	0	0
Contraintes politiques	Politique d'évaluation IGF	0	0	0	2	0	0	0
	Comité d'évaluation interne	2	2	1	4	2	3	2
	Résistance des utilisateurs ciblés	1	2	1	5	1	3	2
	Évaluateur externe	0	0	2	2	0	2	0
SCM	Problèmes de conformité à la culture de la population	0	1	0	1	0	1	0
	Enjeux socio-politiques	1	1	0	1	1	1	1

La faiblesse et (ou) l'insuffisance des preuves

La manifestation de contraintes importantes liées à des problèmes sur le plan l'existence et l'accessibilité des données probantes, de la gestion de ces données (données de programme) sont à l'origine de multiples écueils méthodologiques selon 'S2' et 'S3' reliés au manque de pertinence et de précision-justesse (NPertinence/NPrécision-Justesse) du devis d'évaluation. Comme le montre l'Encadré 8 (p. 214), les lacunes sur le plan des données remettent en question le caractère approprié du cahier des charges, la sélection d'indicateurs pour lesquels on ne peut pas obtenir de réponse ou on n'a pas pu obtenir de réponse, le manque de pertinence des sources d'information, le manque de triangulation des sources d'information, etc. Ultimement, cela entraine un manque de tangibilité des preuves, la démonstration théorique du rendement du programme (la limite la plus conséquente), la formulation de constats sans preuves à l'appui, etc., et donc un manque de fiabilité des conclusions et la remise en question de l'utilité.

Encadré 8. *Thématiques et éléments d'analyse sur le manque de pertinence et précision-justesse du devis d'évaluation, de la faiblesse des preuves, et du rôle joué par certains facteurs contextuels (évaluateurs 'S2' et 'S3').*



La mérepréstentation des données/résultats d'évaluation

Outre l'observation de l'insuffisance et de la faiblesse des preuves face au manque de preuves tangibles et l'absence de données probantes en ce qui a trait au rendement et à l'efficacité des interventions du programme, 'S2' et 'S3' ont observé un manque de transparence et de précision-justesse sur le plan de la synthèse et de l'interprétation des résultats. Plus particulièrement une *mérepréstentation et un mésusage des données/résultats* sur lesquels s'appuie

la formulation des constatations évaluatives (Tableau 16). Il est question ici de *l'exagération et (ou) de l'impression de manipulation des résultats* (9 uds) ou encore de « *data-cherry picking* » (4 uds) ainsi que la *contradiction dans l'usage des données/résultats* (4 uds).

Tableau 16. *Thématiques et éléments d'analyse relatifs aux différentes formes de mérepréséntation identifiées par les évaluateurs séniors 'S2' et 'S3'*

Formes de *mérepréséntation*		Extraits de verbatim
Utilisation inappropriée de pourcentages	Présentation biaisée des preuves	*"Typically, when its key informant, qualitative data, when we are talking about generally small numbers especially since you don't talk to all key informant overall and no external, so the number is getting smaller and smaller. You know then you did 10 interviews, and you say 40% and it's just 4 people you know for me it's not an appropriate use of percentages. It conveys a kind of precision that's not really there. [...]. With key informant interviews you can't say 42% agreed because you did not present those structured survey type questions to a 100 people, so for me it's an unfair way to introduce your evidence."* (S3).
	Présentation erronée des preuves	« *[D'après les données examinées les organismes ont connu des hausses de frais x qui compte parmi les plus élevés. La principale cause [...] forte augmentation des frais d'exploitation [...] hausse des coûts des placements de 28% durant la période en question.] - On me donne des pourcentages, mais je veux des dollars. [...]. L'efficience c'est un regard sur les coûts, sur combien ça coûte de produire un extrant, de produire un résultat, etc. On n'a pas de chiffres dans cette section, il me faut des montants en dollar. Comment quelqu'un peut faire une analyse de l'efficience, dire que ça coûte plus cher, et fournir aucun détail des coûts.* » (S2).
Manque de cohérence et de clarté dans l'utilisation des données	Contradiction dans l'utilisation des données	« Ça *veut dire qu'ils nous ont caché de l'informations, on n'a pas tous les chiffres... Puis là maintenant ils disent qu'il y a moins de prise en charge ? Non il y a plus de prise en charge (12%). Mais que les dépenses ont augmenté de 28%. Pourquoi ? Je ne sais pas. OK, ensuite dans le rendement, ils disent en efficacité que malgré cela, le programme a permis un meilleur accès aux services - vraiment ? ils disent qu'il y a plus de prise en charge ? - « les indicateurs préliminaires font état d'une diminution du nombre de personne pris en charge » - alors que plus tôt dans l'implantation on nous disait qu'il était de +12% ». Est-ce que j'hallucine ? C'est bien ce qui est écris ici... Moi j'accepte des erreurs, c'est possible des erreurs, mais faut que tu donnes des outils aux lecteurs pour vérifier. Et ils ne les donnent pas les outils. »* (S2).
et l'interprétation des résultats	« *Data cherry-picking* »	*"[Within the first few years of the implementation 43% of agency business plans have noted the decline in the number of persons in care.] - I kind of paused here and thought, well, it looks like they did have some empirical data. It looks like business plans were being updated or are evergreen business plans, and that they are tracking and able to see outcomes. I was curious to see that little nugget there and I wondered and thought that this kind of monitoring data is being collected and they are making the link to the implementation of the program."* (S3).

Dans le dernier axe des limites identifiées par les évaluateurs séniors, nous verrons que le manque de transparence et de représentativité sur le plan du devis ainsi que de l'agrégation des données et leurs mérepréséntations dans le cadre de la formulation des constatations évaluatives, entrainent pour 'S2' et 'S3', un problème d'utilité de l'évaluation et de ses résultats. Enfin, lorsque nous avons interrogé ces derniers quant aux potentielles raisons derrière ces manquements observés, ces derniers nous ont expliqué que le contenu du rapport final était de la responsabilité du comité d'évaluation interne. Aussi, il semblerait que si la première version est rédigée par

l'évaluateur externe, le contenu de ce dernier est retravaillé à la lumière des commentaires, des attentes, mais aussi des préférences des commanditaires. Dans d'autres cas, le rapport serait retravaillé directement par l'évaluation interne et il est remis au comité d'évaluation par le directeur d'évaluation qui est responsable de l'évaluation en interne. Dans tous les cas, le dernier mot appartient aux commanditaires, et les participants expliquent qu'ils ne voient jamais (ou rarement), la version finale du rapport d'évaluation.

Extrait de verbatim	« *Il y a peut-être des raisons politiques et internes, moi j'ai déjà fait des évaluations dans un contexte similaire. C'est bourré de politique, alors les évaluateurs ont peut-être été obligés de travailler, retravailler le texte et à force, souvent il y a des bouts qui ne font plus de sens, il y a des contradictions dans le texte. Mais ça, les points qu'on a discutés, c'est fondamental - des statistiques qui manques, des contradictions dans le texte, des constatations qui sont trop courtes. [R — quand vous dites que c'est possible qu'ils l'aient remanié à la suite de discussions politiques...] - c'est possible. Peut-être aussi que quelqu'un a vu la première version du tableau, il y avait peut-être un premier tableau qui disait que bon voici les statistiques des individus qu'ils ont pris en charge, et puis il y a quelqu'un qui a dit « ces chiffres-là ne sont pas fiables, on va les enlever ». Mais ils les enlèvent surtout parce qu'ils ne paraissaient pas bien pour le programme. C'est peut-être ça qui est arrivé. On ne le sait pas, mais c'est possible. Une chose qui est certaine c'est qu'il y a des trous dans la logique, et si on avait le tableau, il aurait pu au moins garder le tableau et expliquer que les données ne sont pas fiables et pourquoi, si c'est bel et bien la raison de leur retrait.* » (S2)

c. Limites - Axe 3 : les préoccupations liées à l'utilité de l'évaluation, et de ses résultats *(15 uds)*

Ce dernier axe de limites identifiées s'avère être la résultante du manque de fiabilité et de représentativité du devis d'évaluation ainsi que de l'agrégation des données et de l'interprétation des résultats (Tableau 17, p. 217). En effet, face au divers manque de transparence sur le plan des procédures méthodologiques, la non-prise en compte des caractéristiques socioculturelles du milieu dans la démarche et les choix méthodologiques, le manque de données probantes et l'absence de preuves tangibles, la méreprésentation des résultats, etc., poussent les évaluateurs, plus particulièrement 'S1', 'S2' et 'S3', à remettre en question l'utilité de l'évaluation (15 uds).

Tableau 17. *Description des cooccurrences entre la remise en question de l'utilité de l'évaluation et les facteurs contextuels*

		Problèmes de fiabilité et de représentation de la démarche		**Non considération de la sensibilité culturelle du milieu**	
		NPrécision-Justesse [Cahier des charges] (9 uds)	NPrécision-Justesse [Agréation des données - interprétation des résultats] (9 uds)	Approche évaluative non-autochtone (3 uds)	NRespect des principes OCAP (5 uds)
Contraintes ESR	C. d'accès aux données	6	2	0	0
	Intrants disponibles	0	3	0	0
	Attentes (±) réalistes des commanditaires	1	2	0	0
Contraintes politiques	Politique d'évaluation IGF	2	2	1	0
	Standards de l'évaluation au fédéral	1	1	1	1
	Appel d'offre	1	1	1	1
	Comité d'évaluation interne	2	5	0	3
	Résistance des utilisateurs ciblés	4	6	0	1
	Évaluateur externe	3	4	0	2
SCM	PP clés communautés ≠ utilisateurs ciblés	0	0	2	5
	NConformité à la culture de la population ciblée	1	1	0	1
	Enjeux socio-politiques	1	2	1	4

Pour 'S1', la remise en question de l'utilité de l'évaluation repose sur le manque de transparence quant au respect des principes éthiques de recherche autochtones OCAP ainsi que l'absence de collaboration avec les parties prenantes clés de la population tant sur le plan de la communication de l'évaluation et de ses résultats, que sur leur inclusion dans le cadre des prises de décisions relatives au projet d'évaluation. En outre, la non-inclusion de ces derniers dans le comité d'évaluation interne signifie que ces derniers ne sont pas perçus comme des utilisateurs ciblés de l'évaluation. Un constat qui met en lumière l'occasion manquée de réaliser une évaluation d'un programme autochtone de qualité, de redonner possession des données, et des résultats aux communautés ciblées par le programme, et de renforcer les bénéfices de l'évaluation pour la population.

Dans les cas de 'S2' et 'S3', l'un des points déclencheurs de la remise en cause de l'utilité de l'évaluation est le manque de transparence, de cohérence et de clarté sur le plan de la synthèse et de l'interprétation des résultats. Outre les problèmes évidents d'accès aux données, et du potentiel manque de compétences et (ou) de rigueur des évaluateurs externes dans la gestion de cette contrainte de terrain importante, ces derniers ne peuvent s'empêcher de penser que le

processus évaluatif ainsi que la communication de l'évaluation et de ses résultats ont été entravés, détournés ou encore freinés par certains utilisateurs ciblés. En ajoutant à cela le fait que l'évaluation ne répond pas à certaines questions évaluatives, ils expliquent que pour eux la qualité et l'utilité de l'évaluation sont fortement remises en question. Un constat d'autant plus inquiétant du fait de la problématique visée par le programme, une problématique qui vise des enjeux importants en ce qui a trait à la santé (et à la sécurité) physique, psychologique et sociale d'individus extrêmement vulnérables dans un contexte socio-économique et politique très sensible.

Enfin, si pour 'S4' les évaluateurs ont réalisé une évaluation de qualité qui répond aux attentes de la politique évaluative de l'IGF et des clients (une mise au point sur les mécanismes d'implantation du programme et une reddition de compte quant au rendement desdits mécanismes), pour les autres évaluateurs l'absence de retour sur l'expérience de la population est regrettable. En effet, pour ces derniers, il était tout à fait possible d'observer les premiers bénéfices du programme, des effets préliminaires des différents services sur les membres des différentes communautés ciblés. Une observation que 'S4' aurait partagée si l'évaluation s'était déroulée dans un tout autre contexte différent (hors politique évaluative gouvernementale, par exemple).

Extrait de verbatim

« Ce rapport-là, il ne nous dit vraiment rien, on aurait pu faire juste une consultation d'experts. Tout ce qu'il nous dit c'est que ça prend de l'argent pour faire de l'activité X. C'est tout ce que ça dit, mettez de l'argent pour les opérations. On aurait pu interviewer trois experts et ils nous auraient dit ça. Ce n'est vraiment pas fort. Ça ne m'apprend rien. Il y a une meilleure pratique, mais malheureusement elle est enterrée dans le rapport. Il aurait pu dire - on a identifié telle nouvelle pratique, on recommande de faire... - et là le lecteur peut se dire que ça pourrait être utile dans d'autres régions. Mais quand il va vouloir plus d'informations là-dessus, c'est complètement évacué de la conclusion et des recommandations. Je l'ai déjà fait, on a même produit un livret qui décrivait la meilleure pratique puis les recommandations. Ici, ça ne s'est même pas rendu dans le résumé. Et puis, on n'a aucune information sur la mise en œuvre du programme et en plus c'était une évaluation de l'implantation. [...]. Ah non ! Le devis ne répond pas du tout au mandat ! Ça m'a beaucoup laissé sur ma faim. Les évaluations ce n'est jamais parfait, mais je trouve malheureusement que celle-ci est bâclée. Et puis comme on en a discuté, il y avait probablement des acteurs en interne qui ne voulaient pas diffuser certaines informations ou l'information n'était tout simplement pas disponible. Mais quand l'information n'est pas dispo, on ne devrait pas juste recommander de renforcer le partage des informations. C'est ridicule et inutile. Il fallait dire de demander que l'organisme financeur et les hauts responsables fassent une demande de partage d'informations pour établir des rapports clairs sur les délais d'attente, d'intervention ainsi que la durée des services une fois la personne prise en charge. C'est quoi leurs normes de service, c'est quoi les attentes, comment est-ce qu'ils livrent par rapport à ces normes-là ? Ce n'est pas compliqué à faire. [...]. C'est ça qui est la statistique. Il n'y a rien là-dessus et c'est ça qui est l'enjeu principal. » (S2)

4.1.2.3 *Interprétation de la question de recherche tertiaire 1* – les pratiques dites recommandées des évaluateurs experts de programmes publics

Question de recherche tertiaire (1) : Lors de l'analyse critique d'une évaluation de programme tierce, quels sont les éléments mis en exergue par les évaluateurs de programmes publics relativement aux choix méthodologiques, et à l'influence des facteurs contextuels sur ceux-ci ?

Au regard des éléments clés de l'analyse critique réalisée par les différents évaluateurs experts participants à cette recherche doctorale, il apparait que les pratiques recommandées par ces derniers semblent être ancrées ici dans l'évaluation en situation réelle (Bamberger et Mabry, 2020) ainsi que la combinaison de savoirs théoriques/procéduraux et de savoirs pratiques/expérientiels propres aux à chacun des évaluateurs experts (LeBoterf, 2002). Aussi, il est notamment question de l'adaptation de la démarche évaluative à une gamme variée de contraintes (C. politiques, budgétaires, temporelles, d'accès aux données), tout en s'assurant de sa pertinence, de sa rigueur, de son utilité et de sa crédibilité (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019 ; Rossi, Lipsey et Freeman, 2019 ; Ridde et Dagenais, 2012 ; Rog, 2005 ; Darabi, 2002). Les pratiques des évaluateurs experts recouvrent également un nombre important des normes relatives aux standards de précision-justesse (*accuracy*), d'utilité, et de faisabilité de l'évaluation du JCSEE (2011), incluant les préoccupations relatives à la fiabilité, la validité et la représentativité des résultats/conclusions évaluatives. Comme vous pouvez le voir dans le Tableau 18 (p. 221-222) qui décrit les caractéristiques des pratiques recommandées des évaluateurs experts sur le plan de la démarche d'évaluation de programmes fédéraux, les évaluateurs experts se concentrent sur la « pertinence » et la « précision-justesse » du devis d'évaluation, et ce, à la lumière :

- *De son caractère approprié pour répondre aux différents axes de questionnement évaluatifs*
  - Alignement des différentes composantes clés du cahier des charges, c'est-à-dire « objet d'évaluation - questions d'évaluation - indicateurs - méthodes - sources d'information disponible ».
  - Évaluabilité des questions d'évaluation.
  - Justesse et représentativité des indicateurs de mesure.
  - Effort de triangulation des sources de données (méthodes-sources d'information).
  - Pertinence/contribution réelle de chaque méthode choisie.

- Pertinence et représentativité des échantillons de sources d'information consultées.
- Pertinence et ampleur suffisante des données probantes collectées.
- Etc.

- *De la prise en compte du contexte politique, organisationnel et socioculturel*
  - Application des directives de la politique évaluative, du guide de revue d'évaluation fédéral.
  - Considération des besoins des utilisateurs ciblés.
  - Démarche évaluative culturellement appropriée, respect des principes éthiques et déontologiques.
- *De la gestion efficace des contraintes de terrain*
  - Paramètres stricts du cahier des charges prescrits dans la demande de propositions.
  - Intrants disponibles pour l'évaluation (budget, échéancier).
  - Maturité du programme, contraintes d'accès aux données (données administratives - de monitorage du rendement du programme, données secondaires disponibles).

Nous avons vu que le contenu du rapport d'évaluation final est tributaire des préférences de l'organisme fédéral commanditaire et que l'espace plutôt limité à la description, la documentation du devis d'évaluation, particulièrement du cadre méthodologique, semble habituel dans le présent contexte. Malgré tout, à la lumière de leur expérience, de leur pratique professionnelle, les évaluateurs experts s'attendent tout de même à un minimum d'information quant aux procédures méthodologiques, et ce, en vue d'apprécier la fiabilité, la validité et la représentativité des informations collectées, mais aussi l'utilité et la crédibilité de l'évaluation – c'est-à-dire, description des sources d'information et procédures d'échantillonnage, des indicateurs, de différentes méthodes, des efforts de triangulation des sources de données (méthodes et sources d'information), ou encore de l'influence des contraintes ESR sur la méthodologie. Ils insistent à l'unanimité sur l'importance de décrire de manière détaillée les limites méthodologiques et les stratégies mises en œuvre pour les gérer au mieux, particulièrement en ce qui a trait aux contraintes d'accès aux données de rendement.

Enfin, les évaluateurs experts 'S2' et 'S3' se sont penchés sur le manque de transparence et de précision-justesse sur le plan de la synthèse et de l'interprétation des résultats, mettant en exergue l'importance de veiller à la présentation et l'usage juste et équilibré des données/résultats dans le cadre de la formulation des constatations évaluatives. Il est question ici de la fiabilité, de la représentativité et de l'utilité des conclusions évaluatives.

Tableau 18. *Pratiques recommandées par les évaluateurs experts dans le contexte de l'évaluation de programmes du gouvernement fédéral*

		'S1' • Firme [A] • Formation [M.A. SHS] [Juriste droit civ.] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [16 ans]	**'S2'** • Firme [B] • Formation [PhD Sc.Écon] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [16 ans] • Enseigne ÉP niveau univ.	**'S3'** • Firme [B] • Formation [M.A. SHS] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [20 ans] • Membre conseil admin. asso. pro. ÉP	**'S4'** • Firme [A] • Formation [M.A. SHS] [Juriste droit civ.] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [17 ans] • Ancien président asso. pro. ÉP
Cadre de référence		*Politique évaluative, pratiques et expériences professionnelles, principes éthiques de la recherche*	*Politique évaluative, guide revue des évaluations fédérales, pratiques et expériences professionnelles*	*Politique évaluative, pratiques et expériences professionnelles*	*Politique évaluative, guide revue des évaluations fédérales, pratiques et expériences professionnelles*
Cadre d'évaluation	Modèle logique	• Outil essentiel pour assurer la pertinence et la justesse du contenu des questions éval. et du choix des indicateurs (alignement, représentativité) ◦ À inclure dans le rapport.	• Outil de conception et de mise en œuvre du devis d'évaluation + extrant éval. • Indispensable pour fournir des preuves de rendement (efficacité et efficience). • À concecevoir et inclure dans le rapport.		
	Questions d'évaluation	• S'assurer de l'alignement des questions éval. avec les priorités de la politique eval. • Miser sur la complémentarité du modèle logique et de la matrice éval. pour qu'on puisse bien saisir les intentions de l'évaluateur et ce qui sera considérer comme preuves pour les questions éval. qui sont souvent très large.	• S'assurer de formuler des questions claires et précises. ◦ Veiller à formuler des questions d'évaluation et non des questions de recherche. ◦ S'assurer qu'on soit en mesure de saisir l'intention derrière chaque question d'évaluation et la manière dont on peut y répondre.	• Suivre les directives de la politique évaluative. ◦ Formulation de questions d'évaluation standards.	• S'assurer de l'alignement des questions d'évaluation avec les axes de questionnement, les attentes de la politique évaluative.
	Matrice éval.	• Outil essentiel pour assurer la pertinence et la justesse du contenu des questions d'éval, du choix des indicateurs et des méthodes et sources d'information. • À inclure dans le rapport.	• S'assurer que le devis d'évaluation répond de manière appropriée aux questions fondamentales de la politique évaluative. *Éval. conception-implantation prog. gouv. d'aide social avec des enjeux sociopolitiques complexes (contexte autochtone)* • Réaliser une comparaison interjuridictionnelle (prog. d'autres provinces, ou pays du Commonwealth). • Comparer les services offerts à la population autochtone dans les réserves, en dehors des réserves, et au reste de la population de la province. *Évaluation du rendement - efficacité* • S'assurer de la justesse et de la représentativité des indicateurs de mesure de rendement. • Recueillir des données probantes, preuves tangibles. • Miser sur la triangulation des sources de données pour pallier le manque de données de rendement. *Évaluation de l'efficience et économie* • Prioriser la démonstration empirique de l'efficience. • S'assurer de la justesse des indicateurs économiques.	• S'assurer de la pertinence et faisabilité du plan d'éval. à la lumière des contraintes ESR (politique éval., demande de propositions avec paramètres stricts en termes d'attentes métho. et d'intrants disponibles, maturité du programme, problèmes d'accès aux données, etc.). *Éval. conception-implantation prog. gouv. d'aide social avec des enjeux sociopolitiques complexes* • Réaliser une comparaison interjuridictionnelle (prog. d'autres provinces du pays, ou pays du Commonwealth). *Évaluation du rendement - efficacité* • Recueillir des données probantes, preuves tangibles. • Identifier rapidement l'existence de bases de données. ◦ Tenter de construire une base de données de rendement via l'analyse documentaire. • Miser sur la triangulation des sources de données pour pallier le manque de données. • Expliquer les stratégies employées et les contraintes rencontrer pour pallier le manque de données de prog.	• S'assurer que le devis d'évaluation répond de manière appropriée aux questions fondamentales de la politique évaluative. • S'assurer que le devis d'éval. répond aux besoins des utilisateurs ciblés de l'évaluation/objectif de l'évaluation. Soit l'éval. du rendement et de l'efficacité: ◦ dans une perspective de reddition de compte ◦ dans une perspective de résultats - montrer les effets du prog. sur la population (p.ex.). • S'assurer de la pertinence et faisabilité du plan d'évaluation à la lumière des contraintes ESR (politique éval., demande de proposition avec paramètres strictes en termes d'attentes méthodologique et de répartition des intrants, maturité du programme, problèmes d'accès aux données). • Rendre compte de l'effort de triangulation des sources de données.
	Plan dévaluation/cahier des charges	• S'assurer que la démarche évaluative soit culturellement appropriée. Dans le contexte prog. gouv. autochtone: ◦ Inclure les parties prenantes clés des communautés dans le comité d'évaluation interne qui valide, entre autres, les choix méthodologiques (cahier des charges, sources d'information, contenu des instruments, procédures de passation, etc.) ◦ Adapter les intrants de l'éval. à la réalité culturelle du milieu, faire particulièrement preuve de souplesse en terme d'échéancier. ◦ Recourir à des méthodes culturellement appropriée. ◦ Respecter les principes de l'éthique de la recherche OCAP des Premières nations.			

		'S1'	'S2'	'S3'	'S4'
		• Firme [A] • Formation [M.A. SHS] [Juriste droit civ.] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [16 ans]	• Firme [B] • Formation [PhD Sc.Écon] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [16 ans] • Enseigne ÉP niveau univ.	• Firme [B] • Formation [M.A. SHS] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [20 ans] • Membre conseil admin. asso. pro. ÉP	• Firme [A] • Formation [M.A. SHS] [Juriste droit civ.] • Évaluateur qualifié [SCÉ] • Expérience pro. ÉP [17 ans] • Ancien président asso. pro. ÉP
Cadre de référence		*Politique évaluative, pratiques et expériences professionnelles, principes éthiques de la recherche*	*Politique évaluative, guide revue des évaluations fédérales, pratiques et expériences professionnelles*	*Politique évaluative, pratiques et expériences professionnelles*	*Politique évaluative, guide revue des évaluations fédérales, pratiques et expériences professionnelles*
Cadre d'évaluation	Sources d'information et échantillonnages	• Décrire (et justifier) les sources d'information sélectionnées. ◦ Référencer la documentation, les écrits scientifiques. ◦ Détailler la composition/les caractéristiques des groupes d'informateurs clés interviewés. ◦ Expliciter la démarche de sélection des cas. • Veiller à la pertinence, représentativité des échantillons à la lumière des objectifs des méthodes de collectes de données et du programme (taille, étendu, caractéristiques des sources d'information). ◦ Consulter un échantillon représentatif (en termes de qualité de « couverture », non de quantité) de sites du programme (agences, bureaux, centres, etc.) ◦ Recueillir la perspective des employés de première ligne.	• Veiller à la pertinence, représentativité des sources d'information sélectionnées. ◦ Recueillir la perspective des membres de la communauté, de la population ciblée par prog. (éval. pertinence-efficacité). ◦ Rassembler l'ensemble de la documentation sources de données admin., secondaires sur le rendement du prog. ▪ Mentionner les contraintes existantes.	• Décrire (et justifier) les sources d'information sélectionnées. ◦ Détailler la composition/les caractéristiques des groupes d'informateurs clés interviewés. • Veiller à la pertinence, représentativité des sources d'information sélectionnées. ◦ Inclure le recueil de perspectives externes au prog. dans la sélection des informateurs clés (i.e., experts, org. communautaires, leaders de la communauté). ◦ Recueillir la perspective des communautés, de la population ciblée par le prog. ◦ Sélectionner les cas selon leur pertinence, mais aussi la disponibilité des informateurs clés (contraintes de temps) et l'ampleur des données de prog. accessibles. ◦ Rassembler l'ensemble de la documentation sources de données admin., secondaires sur le rendement du prog.	• Veiller à la pertinence, représentativité des échantillons à la lumière des objectifs des méthodes de collectes de données et du programme (taille, étendu, caractéristiques des sources d'information). ◦ Mentionner les contraintes existantes.
	Procédures de collectes de données	• Décrire et documenter avec précision les différentes méthodes de collectes de données. ◦ Inclure le modèle logique, la matrice éval. et l'instrumentation dans le rapport. ◦ Décrire (et justifier) la répartition des informateurs clés entre entrevue individuelle et entrevue de groupe. ◦ Décrire les méthodes de collectes de données utilisées dans les études de cas. • Recourir à des entrevues de groupe, utiliser les études de cas (complémentarité des méthodes) afin d'optimiser l'allocation du nombre d'entrevues budgétés et de renforcer la pertinence des des groupes d'informateurs clés consultés.	• S'assurer de la pertinence et de la faisabilité des méthodes sélectionnées au regard des caractéristiques des informateurs clés ciblés. • Miser sur la triangulation des méthodes pour pallier les biais. ◦ *p.ex.* combiner l'enquête auprès des professionnels de première ligne avec l'analyse documentaire (éval. efficacité). • Décrire et documenter le plan d'analyse de l'efficience.	• Choisir la méthode la plus appropriée dans le cadre de la consultation des sources d'information. ◦ Interviewer les experts plutôt que d'analyser un nombre restreints de leurs publications. • Décrire les méthode de collecte de données utilisée. ◦ Préciser s'il s'agit d'entrevues individuelles ou de groupes (nombre d'entrevues réalisées versus nombre de personnes interviewées).	• S'assurer de la pertinence du choix des méthodes à la lumière de la complexité/ de l'ampleur de la probématique ciblée par le programme, des ressources disponibles pour l'éval., et de leur réelle utilité/contribution au devis d'évaluation. • Tenir compte de la sensibilité socioculturelle du milieu dans la collectes de données. Dans le contexte autochtone, ◦ Se déplacer dans les communautés afin de réaliser les entrevues en personne - renforce la pertinence des informations collectés, le partage des expériences des individus, de la communauté. ◦ Sous-traiter la passation des entrevues des études de cas à une firme autochtone.
	Limites	• Énoncer les limites méthodologiques. • S'assurer de la pertinence et de la clarté des limites présentées.	• Énoncer les limites rencontrées. • Expliciter clairement l'implication des limites énoncés, particulièrement les écueils face au contraintes d'accès aux données (données admin./monitorage, données secondaires) • Intégrer des recommandations à destination des clients pour remédier aux contraintes d'accès aux données.	• Énoncer les limites rencontrées. • Expliciter clairement l'implication des limites énoncés, particulièrement les écueils face au contraintes d'accès aux données (données admin./monitorage, données secondaires), et l'impact des ressources disponibles dans la gestion efficace dedites contraintes.	• Énoncer les limites méthodologiques. • S'assurer de la pertinence et de la clarté des limites présentées.

4.2 Les informations fournies par les évaluateurs pour expliquer et décrire les choix méthodologiques entrepris durant l'évaluation (*activité d'explicitation*)

Dans le cadre de l'activité d'explicitation, la chercheuse a réalisé une analyse critique du devis d'évaluation présenté dans le rapport d'une évaluation auxquelles les évaluateurs participants ont contribué, et ce, afin de recueillir des informations quant à leurs pratiques méthodologiques réelles dans le cadre d'évaluation de programmes.

4.2.1 Le cas des évaluateurs juniors (novices)

L'un des principaux objectifs de l'activité d'explicitation était d'amener les évaluateurs novices à nous expliquer les choix et les processus mis en œuvre dans le cadre des tâches méthodologiques auxquelles ils ont pris part. En outre, dans le contexte des évaluations faisant l'objet de l'entrevue d'explicitation, il importe de préciser que, quel que soit le statut de l'évaluateur novice au sein de l'équipe évaluative externe, qu'il soit *junior exécutant* (J2), *juniors lead exécutant*s ('J3' et 'J4'), ou *junior lead collaborateur* (J1), l'ensemble des décisions pratiques et choix méthodologiques reposaient sur les épaules de l'évaluateur expert chargé du projet d'évaluation (développement de la matrice évaluative, conception de l'instrumentation, planification des procédures de collecte et d'analyse de données, synthèse des résultats, communication de l'évaluation, etc.). Comme vous pouvez le voir dans le Tableau 19 (p. 224) qui résume les activités auxquelles ont pris part les différents évaluateurs novices participants, 'J2', 'J3' et 'J4' ont été intégrés au projet d'évaluation au moment de la phase de collecte et d'analyse de données ; tandis que 'J1' nous explique qu'il a pu participer et contribuer de façon significative à l'ensemble des étapes du projet d'évaluation, de la rencontre initiale avec les clients à la rédaction de la première ébauche du rapport final. Si nous ne pouvons pas parler des choix méthodologiques des évaluateurs novices en tant que tels, l'activité nous a tout de même permis de recueillir leurs connaissances et leur compréhension, leurs perceptions et suppositions, sur le plan du processus de conception et de mise en œuvre du devis de l'évaluation à laquelle ils ont participé ; et les facteurs contextuels ayant pu influencer ces processus.

Tableau 19. *Activités réalisées par les évaluateurs novices dans le cadre de leur projet d'évaluation respectif*

'J2' **Junior exécutant** Firme [A] Formation [PhD SHS] Expérience pro. ÉP [-1 an/Act.1] [16 mois/Act.2]	A travaillé principalement sur le profil du programme, son historique, ses activités, etc. - Examen de la documentation, des données de programme (codage, analyse, synthèse des informations), rédaction d'éléments sur le profil du programme en collaboration avec l'évaluateur sénior chargé du projet à l'externe (à des fins d'intégration dans le rapport final). - Recherche d'écrits scientifiques. - Ébauche de résumé des résultats d'enquête.
'J4' **Junior lead exécutant** Firme [B] Formation [M.P.P.A] [DES É.P] Expérience pro. ÉP [2 ans - 2 ans 1/2]	A participé à l'ensemble des collectes de données qualitatives, à l'analyse de données qualitatives ainsi qu'à la rédaction de rapports techniques. - Conduite de l'analyse documentaire – gestion et organisation de la documentation, codage des informations (incluant la supervision des assistants juniors). - Passation d'entrevues. - Analyse de données qualitatives (analyse documentaire, entrevues). - Rédaction de rapports techniques - matrice de résultats de l'analyse documentaire et des entrevues d'informateurs clés, rapport de cas à partir d'un gabarit.
'J3' **Junior lead exécutant** (quelques expériences en tant que junior collaborateur) Firme [A] Formation [M.Écon.] Expérience pro. ÉP [4 ans - 4 ans1/2]	A participé à l'ensemble des collectes de données qualitatives, à l'analyse de données qualitatives ainsi qu'à la rédaction de rapports techniques. - Conduite de la recension des écrits (études de cas inclus) – gestion et mise en œuvre du processus de recherche/sélection/codage des écrits (incluant la supervision des assistants de recherche). - Conduite du recueil et l'analyse de données financières. - Passation d'entrevues. - Analyse de données qualitatives (recension des écrits, analyse documentaire). - Rédaction de rapports techniques (matrice de réponse aux questions d'évaluation) – recension des écrits, entrevues d'informateurs clés, rapports financiers.
'J1' **Junior collaborateur principal** (1re expérience dans ce rôle) Firme [B] Formation [PhD SHS] [DES ÉP] Expérience pro. ÉP [18 mois/Act.1] [28 mois/Act.2]	A participé à l'ensemble des étapes du projet d'évaluation. - Participation à l'ensemble des rencontres de travail de conception du devis d'éval. (rencontre initiale avec clients inclus). - Révision de l'instrumentation. - Coordination des activités de collecte et d'analyse de données (incluant liaison avec l'évaluateur interne, la supervision des juniors exécutants). - Participation à l'ensemble des activités de collecte de données - analyse documentaire, entrevue. - Analyse de données qualitatives provenant de chaque méthode (préparation des rapports techniques/matrices de résultats). - Synthèse des résultats et formulation des constatations évaluatives (matrice de preuves) - accord interévaluateurs avec l'évaluateur sénior chargé du projet d'évaluation externe. - Rédaction du rapport d'éval. final (chargé de la 1re version) et conception d'outils d'implantation (extrants du projet).

4.2.1.1 Contenu du cadre d'évaluation : perceptions et explicitations ancrées dans l'influence de la gouvernance de l'évaluation et du cadre de référence gouvernemental

Dans le cadre de l'activité d'analyse critique, nous avions été surpris de l'absence d'observations de la part des évaluateurs juniors en ce qui a trait au cadre de l'évaluation, ou au choix même des méthodes de collectes de données. Ceci peut s'expliquer par leur statut d'évaluateur *junior exécutant* ainsi que leur perception du poids des standards de l'évaluation au fédéral, et de la gestion des attentes de l'IGF en termes de pratiques méthodologiques et de communication de l'évaluation. L'expérience des évaluateurs juniors dans le cadre de leur projet d'évaluation respectif s'inscrit dans trois contextes qui possèdent à la fois des caractéristiques communes (p. ex. la gouvernance du projet d'évaluation, la complexité de telles entreprises évaluatives), mais qui présentent des variations quant à l'émergence du contenu du cadre d'évaluation.

Contexte évaluatif numéro 1 - les cas de 'J3' et 'J4'

'J3' et 'J4' furent seulement impliqués durant la phase de collecte et d'analyse de données de leur évaluation respective, évaluation réalisée en réponse à la politique d'évaluation fédérale qui exige que les programmes fédéraux soient évalués tous les cinq ans.

Pour 'J4', le contenu du cadre d'évaluation est guidé par la politique évaluative fédérale, déterminé par le comité d'évaluation interne (représentants du programme, de l'agence commanditaire, et division interne de l'évaluation) et prescrit au sein de la demande de propositions. En effet, sur la base de la politique d'évaluation fédérale, les parties prenantes clés détermineraient les objectifs et les questions d'évaluation ainsi que la manière dont on doit y répondre. Selon 'J4', l'équipe évaluative externe n'a pas vraiment de pouvoir décisionnel à ce niveau de la conception du devis d'évaluation, le cahier des charges étant prescrit par la demande de propositions, et ce, de manière plutôt rigide. Il cite l'exemple de questions standards d'évaluation de la pertinence qu'il juge inutiles, notamment celles relatives à l'adéquation du programme avec les priorités du gouvernement où la réponse est toujours « oui », mais à laquelle l'évaluation se doit de toujours répondre.

"Like most federal initiatives or program, most evaluations are driven by the 2009 evaluation policy stating that the program must be evaluated every five years. So, I mean a lot of it is driven by the policy. [...]. The federal department usually when they come out with their RFP, they already identified the evaluation questions. It's quite prescriptive. They tell you exactly the questions they want answered, they tell you how they want you to answer those questions to some extent. They define the evaluation questions. Sometimes there is a bit of leeway in terms of... like some department like to ask for innovative approaches in the methodology but as far as I heard, again just based on the fact, looking at the technical reports and the methodology, the internal evaluation committee chose the evaluation methodology. They chose the methodology. [...]. Je ne veux pas dire qu'on [contractant] ne peut pas apporter de modifications, mais ce n'est pas quelque chose qui se fait d'habitude. Ce sont des questions qui sont déjà établies par un comité interne, ce sont des individus des programmes et de l'agence, de la direction de l'évaluation. Donc ils décident d'abord sur la base de la politique, ils décident quels programmes vont être évalués, les questions, la méthodologie like les méthodes, sources d'information, etc. et quand on reçoit l'appel à soumission, tout est déjà bien défini. [...]. Some of these [evaluation] questions you still have to go through the motions, but they aren't useful. Plus, it's not very efficient to use resources to answer these questions. It meets the requirements of the policy, but the answer is always yes, why are we asking." (J4)

Si les propos de 'J3' s'alignent avec ceux de 'J4' en ce qui a trait à l'ancrage politique du plan d'évaluation, il inclut en revanche une autre entité prescriptive de la démarche à suivre afin de répondre de manière appropriée à la politique évaluative fédérale, à savoir les directives évaluatives de l'IGF. Selon 'J3', celles-ci prescrivent un cadre d'évaluation arrimé aux préoccupations et questions d'évaluation fondamentales de l'IGF, qui visent à rendre compte de la pertinence et du rendement (efficacité, efficience et relation coûts-bénéfices) du programme. En outre, ce dernier fait preuve d'une connaissance très précise du contenu des différents axes de questionnements (objets d'évaluation) et des questions d'évaluation prescrites par les directives de l'IGF pour y répondre.

"There is a sort of overall framework for doing this and this is informed by the IGF. It talks about the way that evaluations should be done, how we should go about answering the mandate, the questions. Like looking at relevance and performance, these are the things that have been singled out by [IGF]. These evaluation directives, gives the core evaluation issues, and for each core issue a set of questions and the evaluation is linked to these directives. So, when talk about relevance and performance. These are usually you know these are the things that have been singled out by TBS, these are the things that they really want to know all about. Relevance broadly speaking is - why does this program exist? And what it is for? You will also look at - how is the program aligned the government objectives, priorities? How is it aligned with the departmental mandate of the agency? For example. So, these are typical questions regarding relevance. For performance, it usually refers, on the one hand, to the actual achievement of outcomes - were things implemented the way it should have been implemented? Were the activities planned the activities carried out? Did we see the early outcomes that we expected to see? Intermediate outcomes? Final outcomes? And the second half of the performance questions relate to efficiency and economy. And those are questions that relate more to resource use. So, given what the program was trying to accomplish was the use of resources reasonable? Could they have used the resources in a better way? Is there another program designed that could have been better situated to achieve these types of outcomes? Etc." (J3)

Comme se fût le cas pour l'exercice d'analyse critique, 'J3' et 'J4' semblent avoir une nouvelle fois une perception plutôt divergente en ce qui concerne l'influence politique sur l'évaluation externe de programmes fédéraux. En effet, pour 'J4', tout semble se faire dans l'obligation - obligation d'évaluer, obligation de poser telles questions d'évaluation, obligation de répondre aux questions d'évaluation de telle manière. L'évaluation de programmes fédéraux deviendrait, pour lui mécanique, la démarche évaluative systématique et non innovante. Pour 'J3', les directives évaluatives émises par l'IGF répondraient davantage à une question légale. S'il n'est pas en mesure de se prononcer sur la dimension obligatoire de la démarche d'évaluation à suivre, il suggère fortement qu'il s'agit d'un modèle auquel les agences gouvernementales adhèrent et aspirent, et ce, pour répondre aux mieux aux attentes de l'IGF et des hauts décideurs. En revanche, si pour lui, c'est synonyme de rigueur et de qualité, il conçoit que cela laisse peu d'espace à la négociation du contenu du devis d'évaluation.

De plus, l'évaluateur 'J4' perçoit l'équipe évaluative externe comme exécutante du plan d'évaluation prescrit par le comité d'évaluation interne, alors que 'J3' reconnait le rôle que ladite équipe a à jouer dans l'opérationnalisation des paramètres de l'évaluation prescrite dans la demande de propositions, avec par exemple la contextualisation du contenu des questions d'évaluation prescrites afin qu'il reflète les caractéristiques du programme, ou encore la détermination du contenu de la matrice évaluative (« *evaluation crosswalk* »), l'organisation des méthodes, etc. Il aborde ici la phase d'évaluation de l'évaluabilité mise en œuvre par l'équipe évaluative externe (examen de la documentation et des données administratives disponibles). 'J3' perçoit la division de l'évaluation interne comme un point de contact, un agent de liaison avec les parties prenantes clés de l'évaluation (reste des membres du comité d'évaluation interne), un conseiller et une source de rétroaction méthodologique. Aussi, contrairement à 'J4', une dimension collaborative entre l'équipe évaluative externe et le comité d'évaluation interne sur le plan de la conception du plan d'évaluation émerge dans le discours de 'J3'.

"The [IGF] directive, the request for proposals, gives for the core evaluation issues, and for each core issue describe a set of questions. So we [the contractor] would link our evaluation to these directives. We would adapt the questions to the context of our program but using the [IGF] guidelines when it comes to the content and structure of the questions. [...]. We are the ones who are doing the work. I mean early one we talk about what the evaluation should look like with the evaluation division, and you know they certainly have some input in terms of what questions we ask and how we should go about it, the methods we should, we can use to answer each question. They are our key point of contact in completing the work, they liaise, they give their feedback. I do believe they review the deliverables as well [methodology and technical reports], but ultimately, we are in charge of the design of the evaluation plan, what we call here an evaluation crosswalk. But yes, of course, within the parameters defined in the RFP. [...]. The evaluation design and the evaluation matrix are usually developed based on the evaluability assessment. It's like a preliminary review of documents and administrative data that are available to us early on. The evaluation crosswalk or matrix if you prefer was designed to answer like eight or ten main evaluation questions and a number of subquestions, indicators related to program relevance and performance in accordance with the federal evaluation policy." ('J3')

Enfin, l'une des particularités des projets sur lesquels 'J3' et 'J4' ont respectivement travaillé portait sur la décision d'évaluer deux programmes de manière conjointe.

- Dans le cas de 'J4', il s'agissait de la volonté des commanditaires de combiner l'évaluation de deux programmes sur une même demande de propositions et de produire un rapport final unique. Aussi, ce dernier nous explique que les départements sont sous l'obligation d'évaluer leurs programmes tous les cinq ans, mais que ceci n'est pas reflété dans le budget disponible pour leurs activités évaluatives. - "*We see that a lot. They need to check those boxes and decide to join the evaluation of two programs.".* Dans le cas présent, il s'agissait pour 'J4' d'une initiative pertinente, car les programmes étaient similaires, mais que parfois cela complique l'évaluation. D'ailleurs, lorsque nous lui avons fait remarquer qu'un des programmes semblait avoir pris le dessus sur l'autre en ce qui a trait à l'ampleur des informations collectées et des résultats rapportés, il émit l'hypothèse qu'un des programmes devait avoir moins de ressources financières. Aussi, cela fait échos à sa remarque sur la nécessité de « *cocher une case* », ce qui remet pour nous en cause l'utilité de l'évaluation du deuxième programme.
- En ce qui a trait au projet de 'J3', la décision a été prise après l'obtention successive de deux contrats d'évaluation pour des programmes d'un même département, qui adressent une problématique similaire et possèdent une structure semblable. En effet, les équipes évaluatives ont réalisé qu'elles avaient beaucoup de sources d'informations en commun et qu'elles pourraient utiliser les ressources de manière plus efficace en combinant les

procédures de collectes de données. Les deux évaluations partagent les mêmes rapports techniques, mais chacune a fait l'objet de son propre rapport d'évaluation finale.

Contexte numéro 2 - le cas de 'J2'

Il s'agissait d'un des premiers projets d'évaluation auquel 'J2' a participé à titre d'exécutant dans le cadre de certaines activités de collecte et d'analyse de données. Aussi, si 'J2' mentionne à plusieurs reprises qu'il est arrivé tard sur le projet, soit une fois que tout était fait sur le plan de la conception du protocole méthodologique, il est tout de même en mesure de nous faire part d'une caractéristique bien particulière vis-à-vis des attentes des parties prenantes clés de l'évaluation. En effet, si les activités du programme évalué ne faisaient pas l'objet d'une obligation formelle d'évaluation sous le cadre prescriptif de la politique évaluative fédérale, son évaluation a été mandatée comme si elles l'étaient - "[*B*]*ecause they are funded by Ministry X, I think that they probably had some impact on what shape that would take*.". En effet, il suppose que le contenu du mandat d'évaluation est calqué sur le modèle fédéral standard afin de faciliter le déroulement de l'évaluation et de son utilisation par le ministère financeur. Ce désir de conformité s'étend d'ailleurs jusqu'au contenu même du rapport d'évaluation. En outre, s'il ne fait jamais référence au contenu de la demande de propositions, les considérations méthodologiques initiales semblent s'inscrire dans la démarche évaluative standard et les attentes des clients. Pour 'J2', le processus décisionnel engagé dans la conception du cadre d'évaluation s'avère finalement presque par défaut, et ce, pour le bien de l'évaluation. Aussi, s'il semble cerner l'ancrage du cadre d'évaluation dans le modèle fédéral standard, 'J2' ne mentionne pas de manière explicite les directives de l'IGF, même si c'est ce dont il s'agit. Il ne fait pas non plus de liens entre l'utilisation d'un tel cadre de référence et la formulation des questions d'évaluation. Enfin, il n'est pas en mesure de nous décrire la relation de travail entre le comité d'évaluation interne et l'évaluateur sénior en charge de l'équipe évaluative externe. Il parle des décisions d'une entité pluriel « ils », mais nous ne sommes pas en mesure de discerner de qui il s'agit, ni de la dynamique de pouvoir décisionnel entre les différentes parties concernées.

Ce qui est intéressant ici c'est que 'J2' nous affirme à plusieurs reprises que la division de l'évaluation de programme n'était pas impliquée dans le projet d'évaluation, puisqu'il s'agissait d'une évaluation indépendante. Cependant, le rapport disponible en ligne est attribué à cette dernière, comme si elle en était la commanditaire.

"*I think basically they took it sort of cookie cutter from what a standard federal evaluation would look like in the sense that they look at effectiveness, relevance. You know economy and efficiency and should also look at value for money. These are all sorts of standards, things they put into a federal evaluation. So, I think that they consciously tried to follow the federal model, answer to these standard evaluation questions, using a standard design, method, just so that the report would have a form that Ministry X would be able to process I guess more easily... I don't think they technically fall under that legislation, so they're not required to report it in exactly that way, but I think they just made the choice to comply with that to make it easier for everyone involved*." ('J2').

Contexte numéro 3 - le cas de 'J1'

L'évaluateur 'J1' était membre de l'équipe évaluative externe à titre de *lead junior*, collaborateur principal de l'évaluateur sénior partenaire du cabinet-conseil qui est chargé du projet d'évaluation. 'J1' explique qu'il s'agissait de sa première occasion de participer et de contribuer à l'ensemble des étapes d'une évaluation de programme, de la rencontre initiale avec les clients pour travailler sur le contenu du cahier des charges à la rédaction de la première ébauche du rapport d'évaluation final.

L'évaluation à laquelle a participé 'J1' s'inscrit dans un projet d'étude de l'IGF visant à soutenir l'implantation et le monitorage d'une approche au sein des différentes agences et départements du gouvernement fédéral. Aussi, si le présent projet est complètement différent de ceux des évaluateurs 'J2', 'J3' et 'J4', la gouvernance de l'évaluation demeure la même : 1) l'évaluation est commanditée par un comité consultatif interne constituée de parties prenantes clés du projet d'étude incluant des membres de la division interne de l'évaluation de l'IGF, et 2) un évaluateur de la division interne de l'évaluation tient le rôle de consultant, d'agent de liaison entre le comité consultatif et l'équipe évaluative externe.

Sur le plan du contenu du cadre d'évaluation, 'J1' ne fait aucunement référence à la demande de propositions ou aux directives des clients en ce qui a trait au contenu du cahier des charges de l'évaluation. L'entrée de 'J1' sur le projet d'évaluation semble avoir été au moment de la rencontre initiale avec le comité consultatif interne où le contenu initial du plan d'évaluation est présenté, discuté, affiné, particulièrement avec l'appui de la division interne de l'évaluation. Aussi, celui-ci nous fait part du caractère essentiel de cette rencontre, et ce notamment, pour les raisons suivantes :

- Cerner les attentes/besoins des parties prenantes clés du projet d'étude/d'évaluation (comité consultatif interne) - l'un des extrants de l'évaluation ici est un plan d'implantation

de l'initiative adaptable selon les caractéristiques des agences gouvernementales, accompagné d'une boite à outils de monitorage de l'initiative.

- Explorer la manière dont les différentes parties prenantes clés définissent le concept central de l'initiative à l'étude, les facettes clés de sa mise en œuvre, etc.
- Définir/affiner/modifier les questions d'évaluation, ou plutôt thèmes d'investigation dans le cas présent du fait de la complexité de l'initiative évaluée (nous élaborons davantage ce point dans la section sur les répercussions de la complexité de l'initiative évaluée sur le devis d'évaluation).
- Recommander/identifier des écrits pertinents, une première vague d'informateurs clés.

Enfin, 'J1' explique que le processus de développement du cahier des charges fut un effort collaboratif entre l'équipe évaluative externe, plus particulièrement l'évaluateur sénior responsable du projet d'évaluation à l'externe, et la division de l'évaluation interne. Il précise que selon lui l'évaluateur chargé de l'évaluation à l'interne n'était pas là pour dicter la marche à suivre, mais bien pour conseiller les évaluateurs externes sur la faisabilité des propositions méthodologiques au regard, notamment, de leur expertise en ce qui a trait aux possibilités d'accès aux données dans un tel milieu évaluatif.

Veuillez noter que si les éléments clés du cadre d'évaluation sont résumés dans le Tableau 20 à la page suivante, nous revenons plus en détail sur les particularités du mandat d'évaluation dans la section portant sur les répercussions de la complexité de l'initiative évaluée sur le devis d'évaluation.

Tableau 20. *Explications ou perceptions du processus de définition du cadre d'évaluation (évaluateurs juniors)*

		Contexte numéro 1		Contexte numéro 2	Contexte numéro 3
		'J4'	'J3'	'J2'	'J1'
Gouvernance de l' évaluation	*Entités évaluées*	Programmes d'un ministère du gouvernement fédéral (GF).		Activités d'une ONG financée par un ministère GF.	Agences et départements du GF.
	Initiative évaluative	Politique évaluative fédérale.		Indépendante.	Projet d'étude d'un ministère GF.
	Parties prenantes clés	Comité d'évaluation interne (CMI) - division de l'évaluation IGF et membres de la direction du programme.		Membres de la direction du programme et du ministère financeur.	CMI - Comité consultatif chargé de l'étude de la problématique et division de l'évaluation IGF.
	Équipe évaluative externe	Évaluateur responsable - sénior partenaire ; 'J3' et 'J4' - **juniors lead exécutants** d'un nombre important d'activités de collecte et (ou) d'analyse de données.		Évaluateur responsable - sénior partenaire ; 'J2' - **junior exécutant** d'un nombre limité d'activités de collecte et (ou) d'analyse de données.	Évaluateur responsable - sénior partenaire ; 'J1' - **lead junior, principal collaborateur** de l'évaluateur sénior chargé du projet d'évaluation externe.
Contenu du cadre évaluatif	*Objectifs évaluatifs*	• Objectifs standards GF - amélioration du programme, reddition de compte, et monitorage la progression des cas étudiés lors de la précédente évaluation.	• Objectifs fondamentaux des directives de la politique éval. de l'IGF - rendre compte de la pertinence et du rendement du programme.	• Objectifs standards du GF.	• Objectifs non standards d'une évaluation GF - utilisation des résultats évaluatifs afin de guider l'implantation et le monitorage de l'approche X au sein des différentes agences du GF.
	Type et fonction de l'évaluation	• Prescrits par le CMI dans la demande de propositions (DDP) à partir de la politique éval. GF - éval. de la pertinence et du rendement.	• Déterminés par le CMI dans le DDP à partir des directives de la politique éval. de l'IGF - éval. de la pertinence, l'efficacité, l'efficience et la relation coûts-bénéfices.	• Application du modèle évaluatif standard GF - évaluation de la pertinence, l'efficacité, l'efficience et la relation coûts-bénéfices.	• Perspective davantage formative qui se manifeste par l'absence de la dimension évaluative habituelle du rendement - évaluation de la mise en œuvre et des résultats.
	Questions évaluatives	• Prescrites par le CMI dans la DDP à partir de la politique évaluative GF.	• Formulées par l'éval. sénior chargé de l'éval. à l'ext. qui contextualise le contenu des questions standards du cadre de réf. de l'IGF et prescrites dans la DDP, au regard des composantes du prog. • Validées auprès du CMI.	• Hypothèse 1 - conçues durant une évaluation antérieure du programme. • Hypothèse 2 - conçues par l'EC à la lumière des attentes du CMI, des objectifs du programme.	• Formulées par l'éval. sénior chargé du projet d'éval. ext. en collaboration avec CMI durant la rencontre initiale - définition de thème de questionnement et non de questions standards GF. • En raison de la complexité de la problématique ciblée, conception itérative de sous-thèmes d'investigation au fil du recueil et de l'analyse des données.
	Cahier des charges/ Plan d'évaluation	• Prescrit dans le DDP par le CMI qui détaillent la façon de répondre aux questions d'éval. (méthodes, sources d'information - nombre d'entrevues à réaliser, cas à étudier). • Réutilisation du devis de l'évaluation précédente du prog. (hypothèse)	• Prescrit dans le DDP qui détaille les sources de données et d'informations, et caractéristiques échantillonnales (p.ex. nombre d'entrevues et de cas, paramètres de recension). • À partir des paramètres de la DDP, de l'analyse profonde de la logique du prog., et des ressources disponibles, l'éval. sénior chargé de l'éval. à l'ext. propose son plan d'éval. - « *evaluation crosswalk* ». • Collaboration de la division interne de l'éval. • Validation par le CMI.	• Aucune référence à la DDP ou au processus de proposition du plan méthodologique. • Suivis des lignes directrices de la démarche évaluative propre au modèle fédéral. • Proposition de méthodes de collecte de données standards du modèle fédéral.	• *Aucune référence à la DDP.* • L'éval. sénior chargé du projet d'éval. ext. travaille de concert avec la division de l'évaluation IGF sur la proposition de leur plan méthodologique. • La division de l'évaluation IGF, conseillée de l'équipe éval. ext. veille à la faisabilité des choix proposés à la lumière, entre autres, des contraintes d'accès aux données, et relaye les décisions prises, les besoins d'information, etc. auprès du CMI.
	Matrice éval.	Mentionnée comme principal outil de conception de l'instrumentation, d'analyse et d'agrégations des données.		Mentionnée comme source d'indicateurs à utiliser pour collecter et analyser les données.	• Développement itératif - échantillonnages de type boule de neige avec définition, et affinement des thèmes et sous-thèmes d'investigation, au fil des informations recueillies.

Notons qu'aucun des évaluateurs juniors ne fait mention de la fonction de l'évaluation. Ils se réfèrent toujours aux objectifs fondamentaux de l'évaluation - soit la nécessité de rendre compte de la pertinence, de l'efficacité, de l'efficience et du rapport coûts-bénéfices du programme. Selon les propos de 'J1', le seul à avoir tenu un rôle *de collaborateur,* les évaluations de programmes fédéraux sont généralement sommatives. Aussi, ceci constitue son point de référence quand il explique les objectifs de son évaluation, qui correspondent davantage à ce qu'on attend d'une évaluation formative, à savoir la mise au point sur la pertinence, la mise en œuvre et les résultats d'un programme moins mature. L'ajout des composantes liées à l'évaluation du rendement (efficacité, efficience et rapport coûts-bénéfices) représente pour lui ce qui caractérise une évaluation sommative, ce qui est l'objectif de la politique évaluative.

4.2.1.2 Communication du devis d'évaluation : perceptions et explicitations des carences observées sur le plan du contenu rapporté

À l'instar des observations émises par les évaluateurs novices durant l'activité d'analyse critique, nous avons observé dans le cadre de la préparation de l'activité d'explicitation un manque de précision sur le plan du contenu du devis d'évaluation présenté dans les rapports d'évaluation fournis par les évaluateurs participants. En effet, comme vous pouvez le voir dans le Tableau 21 (p. 234), on remarque notamment que le cadre d'évaluation est présenté très succinctement. On note plus particulièrement l'absence des questions d'évaluation dans la plupart des cas, celles-ci se trouvant en annexe dans la matrice évaluative, et (ou) étant abordées dans le corps du rapport en tant que thèmes et sous-thèmes de la section sur la présentation et l'interprétation des résultats. Et un seul rapport contient la matrice évaluative ('J3'). En outre, si le caractère succinct du cadre d'évaluation, et plus généralement du devis d'évaluation, est attribué au désire de produite des rapports d'évaluation moins techniques, centrés sur les résultats, la matrice évaluative en elle-même offre une présentation complète du plan d'évaluation, et demeure selon nous indispensable au contenu du rapport final, particulièrement face au manque de description de certains éléments clés du devis d'évaluation.

Tableau 21. *Suppositions et explicitations des évaluateurs juniors quant au manque de précision sur le plan du contenu du cadre d'évaluation rapporté*

Critères principaux	Critères détaillés	Respect des critères			Suppositions et explicitations
		Non	Part.	Oui	
Présentation du modèle logique		J1	J3	J2 J4	• Le programme évalué par l'équipe de 'J1' n'avait pas de modèle logique, mais le rapport contient une représentation graphique du plan d'implantation de l'approche (produit à partir des données, des résultats de l'évaluation). • Le rapport de 'J3' contient une section « modèle logique » où sont décrit sous forme textuelle les activités, les extrants et les différents niveaux de résultats attendus, le modèle logique étant accessible uniquement via une demande par courriel à l'agence commanditaire. "*It's odd not to include the logic model but it's probably in the same line of consideration than for the evaluation questions. Also, we prepare the report, but the evaluation division could've decided that it shouldn't be included, pulled it out still recognizing people can always email to get it. I don't know who made that decision. One of two things I think they figured that it wasn't essential information, that the description of activities, outputs and outcomes would suffice. Or the complexity of the program would have made it complicated to present the model in that format.*" (J3).
Présentation des questions éval.	Dans le devis d'éval.	J1 J2 J3		J4 (-)	• Le rapport de 'J4' est le seul qui présente les questions dans la section « objet et portée de l'évaluation ». Les questions sont généralement précises, cependant certaines nécessitaient un découpage en sous-questions (la matrice éval. n'était pas inclus dans le rapport). • Dans chacun des rapports, la section sur les résultats est organisée selon les questions/sous-questions ou plutôt thème/sous-thèmes d'éval. Aussi, dans les cas de 'J1 et 'J2' nous n'avons pas été en mesure de consulter les questions d'éval. à proprement parlé (pas de matrice éval. en annexe). "*Honestly that decision would have been made without me, I don't know why the evaluation questions are not included in the report. My assumption is that that would be just to keep it less technical. I think they were trying to sort of give the high-level picture. You know like here's what's going on with the program. Those questions are also a bit artificial, very formulaic. Things like what would happen if this organization didn't exist or what is the unique value that this organization adds. They're obviously important questions that need to be addressed directly but in terms of forming a narrative about the program they're just very staccato and bullet point. It's difficult to wrap prose around it and make it sound interesting.*" (J2). *"I guess the tools weren't attached to here like the evaluation matrix, the instruments. Well, I guess like we always put the headers in the findings section kind of as our questions or guiding theme. [...]. It's not a typical evaluation report which is great because the evaluation doesn't follow the standard template, but this is kind of rare. I liked it. They didn't want this like straightforward structural typical report. They wanted it to look neat and different and be easy to read, more focused. that's why I didn't follow the flow. Normally in one of these reports you would have the executive summary and it has the evaluation questions, the objectives, the scope. But here the client wanted us to make it a little bit more creative like the theme of the evaluation [fostering change and creativity]. The focus of the report were the findings, their interest was focused on the tool that we came up with as result of the evaluation project."* (J1).
	Dans la section résultats		J1 J2 J3 J4		
	En annexe	J1 J2 J4		J3	• Le rapport de 'J3' présente un éventail de questions et sous-questions d'évaluation très précises dans la matrice évaluative inclus en annexe. *"I didn't draft this report so I can't say for sure, but I would bet that the rationale for introducing the questions in the appendix [in the matrix] is, it's just a lot of the information for people to absorb. The trend has been to do short final report rather than longer ones. There is a lot of questions more like main questions with a lot of sub questions. The only way would have been to maybe summarize them, giving only the main one but then you lose in term of precision."* (J3).
Présentation de la matrice éval.		J1 J2 J4		J3	• Si les équipes éval. ext. semblent répondre au désire de produire des rapports moins techniques, plus courts et centrés sur les constats évaluatifs, l'inclusion de la matrice éval. à elle seule permettait de présenter les questions/sous-questions d'éval., indicateurs de mesure et sources de données mobilisées pour y répondre. « *Je pense que vous pourriez faire une demande officielle à [l'IGF]pour obtenir la matrice évaluative. I think to some extent they don't include it in here because of privacy concerns as well. In some of the technical report we have identified individuals and organizations, etc. We guaranteed anonymity to our participants. But a lot of work actually went into it. Also, there are not a lot of people interested in this and I think that's why the report is maybe so general. [...]. Bien entendu une matrice a été développée avec les questions et les indicateurs précis ainsi que les méthodes. Je ne sais pas pourquoi ce n'est pas dans le rapport. C'est aussi le comité d'évaluation interne du gouvernement qui décide du contenu et puis moi je n'étais pas impliqué dans la rédaction. [...]. Je ne sais pas pourquoi ils ont choisi de mettre juste un sommaire de la métho en annexe. Après il y a des rapports techniques plus détaillés pour le client.* » (J4).

En ce qui concerne le cadre méthodologique, dans deux rapports, celui-ci est présenté en annexe (‘J2’ et ‘J4’), et son contenu fait entre une demi-page (‘J1’) et quatre pages (‘J2’ et ‘J3’). Parmi les carences observées (Tableau 22, p. 236), on pouvait trouver : 1) l’absence de justification des choix de méthodes de collectes de données; 2) le manque de description des procédures de sélection de certaines sources d’information (échantillonnages); 3) le manque de description des procédures de collectes de données et de conception de l’instrumentation; et (ou) 4) le manque, voir l’absence de description des procédures d’analyse et d’agrégation des données (dans certains cas, l’analyse et l’agrégation des données ne sont pas abordées dans le cadre méthodologique).

Que ce soit dans le cadre de l’explicitation de ‘J1’, ou les hypothèses émises par les autres évaluateurs novices, la principale raison évoquée quant à l’absence de ces différents éléments a trait aux attentes des clients, au modèle de chaque organisme fédéral commanditaire en ce qui a trait au contenu méthodologique rapporté. D’après les évaluateurs novices, le rapport d’évaluation final doit être court, et il doit se concentrer sur les résultats de l’évaluation, soit sur ce qui semble le plus intéressant pour le lecteur, plus précisément les utilisateurs ciblés. De plus, toutes décisions finales concernant le contenu de ce rapport semblent appartenir aux clients, plus précisément à la division interne de l’évaluation, selon ‘J1’ et ‘J3’. L’existence d’un rapport méthodologique contenant des éléments essentiels tels que la matrice évaluative, l’instrumentation, le protocole de collecte et d’analyse des données, et de rapports techniques (‘J1’, ‘J3’ et ‘J4’), ainsi que les propos recueillis auprès des évaluateurs novices nous montrent que le manque de précision ne signifie pas nécessairement un manque de transparence volontaire, et (ou), un manque de rigueur méthodologique. Enfin, deux rapports possèdent une section limite assez riche, qui rend compte de certains écueils (également observés par la chercheuse), liés à l’évaluation de l’efficacité-efficience et la gestion des contraintes d’accès aux données de rendement du programme (p. ex. absence de système d’information de gestion et de production systématique de rapports de rendement, données manquantes, problème d’alignement entre la meure de rendement et le modèle logique)

Tableau 22. *Suppositions et explicitations des évaluateurs juniors quant au manque de précision sur le plan du contenu du cadre méthodologique rapporté*

Critères principaux	Critères détaillés	Respect des critères			Suppositions et explicitations
		Non	Part.	Oui	
Description de la méthodologie	Description des sources d'information et échantillonnages			J3 J4 (+) J2 J1(-)	• Espace accordé au cadre méthodologique restreint (entre une demi-page et quatre pages). Le contenu du rapport de 'J1' est le plus lacunaire avec très peu voir aucun effort de description - les deux méthodes employées ont été décrites en cinq lignes au total accompagnées de la liste des informateurs clés interviewés. Dans les autres cas, la méthodologie évaluative mobilisait au moins quatre sources de données. Dans les rapports de 'J2' et 'J4', elle est présentée en annexe, en deux pages pour 'S4' et quatre pages pour 'J2. Comme dans le cas de 'S1', la méthodologie fait l'objet d'une section dans le corps du rapport de 'S3'(quatre pages). • Le choix des méthodes n'est pas justifié. En revanche, dans le cas de 'J3' nous avons la matrice éval. ainsi qu'une brève description du processus de définition du plan d'évaluation ; et dans le cas de 'J4', il est fait mention du recours à une approche multiméthodes, misant sur l'équilibre entre des méthodes quantitatives et qualitatives, où les méthodes qualitatives visent à fournir une description/explication des données quantitatives. • Pour chaque méthode, l'accent est particulièrement mis sur la description des sources d'information et échantillonnages, avec un effort plus marqué dans les rapports de 'J3' et 'J4'. On trouve aussi une brève description quant aux objectifs de la collecte d'informations, et quelques éléments procéduraux relatifs au recueil des données dans les cas de 'J2', 'J3' et 'J4'. • Nous n'avons observé aucune mention d'éléments sur la conception de l'instrumentation. • Le rapport de 'J2' est le seul à contenir une description partielle des procédures d'analyse, plus précisément des analyses statistiques des données d'enquête. Dans les cas de 'J3' et 'J4' on retrouve quelques lignes sur le procédé de synthèse et de triangulation des résultats pour répondre aux questions d'éval.; et le rapport de 'J3' inclut, par ailleurs, la matrice de réponse aux questions d'évaluation. Aucune mention des analyses n'est faite dans le rapport de 'J1'. "*You know, there would've been a more detailed discussion in the technical report that wouldn't come here. So yeah, like a methodological report, technical reports with details regarding data analysis procedures and results for the different line of evidence, you know for the government agency, our client. But yeah, I mean, the final report, there just isn't enough space to do that.*" (J3). "*The clients here, was like make it simple, concise. I think we submitted like not attached to this - our evidence matrix for each method for people who do want to see how we a came up with these themes since we had begun with a very general evaluation matrix that we fine-tuned along the way. Usually in a report we would describe in detail. There are other attachments that didn't make it into the final report, I guess. Again, a reflection of him saying - don't make this a typical government report. This is more... also even that is short compared to a community-based report. They always want them 10-20 pages and you must cut. But we always back up our evidence. We are able to give the evidence matrix, so they don't think that we just kind of pull them out of nowhere. I'm finding that more and more in a lot of evaluations people just want to... because leaders and decision makers at the admin level are now reading these reports, are getting more involved, that's a very prominent thing that's happening. They want a concise, to the point document. These attachments are more for the program level, people who care about it. We have a very detailed methodology report with everything in it, matrix, instruments, data collection and analysis procedures, etc. and a technical report for each method.*" (J1). • Les rapports de 'J3' et 'J2' sont les seuls à offrir une description assez détaillée des limites méthodologiques et des efforts entrepris pour les contrecarrer. Dans les cas de 'J1' et 'J4' celles-ci ne sont pas mentionnées.
	Description des indicateurs de mesure	J1 J2 J4		J3	
	Justification du choix des méthodes	J1 J2	J3 J4 (-)		
	Description des procédures de collecte de données		J1 (-)	J3 J4 J2 (-)	
	Description des procédures de conception de l'instrumentation	J1 J2 J3 J4			
	Description des procédures d'analyse de données / synthèse des résultats	J1	J2 J3 J4 (-)		
	Description des limites méthodologiques	J1 J4		J2 J3	

4.2.1.3 Conception du cadre méthodologique : perceptions et explicitations ancrées dans les décisions de l'évaluateur sénior chargé du projet à l'externe et le comité d'évaluation interne

a. Les enjeux méthodologiques face aux contraintes de terrain et la complexité du programme

Perception quant aux répercussions de la complexité du programme sur la conception et la mise en œuvre du cadre méthodologique - Les cas des évaluateurs juniors exécutants 'J2', 'J3' et 'J4'

- Un nombre important de composantes principales et une quantité d'activité presque trop importante pour évaluer les résultats (mentionné dans les limites).
- Un large éventail de parties prenantes clés du programme impliquées dans la gestion de la mise en œuvre des activités du programme et de la production des extrants.
- Des difficultés à cerner la contribution de chacune des composantes du programme ainsi que les résultats attendus.
- Un nombre de sites et de groupes d'intérêt ciblé très large (multisite à travers le pays dans certains cas).

Le cas de 'J2'

À la lumière des discussions qu'il a eues au cours du projet, le principal défi reposait selon lui sur le fait que l'organisme dont les activités font l'objet de l'évaluation a pour mission de desservir de manière plus ou moins importante différents systèmes opérationnels à travers le Canada, dans le domaine ciblé. Aussi s'agissait-il : 1) d'offrir à l'ensemble des groupes d'intérêts l'occasion de faire part de leur expérience et de leur rétroaction sur le programme, et ce, de manière significative ; et 2) de recueillir le maximum d'information auprès du plus grand nombre de groupes et d'individus possible, pour un moindre coût. Pour ce faire ils ont recouru à l'enquête, et 'J2' précise que pour tout ce qui concerne les activités et les différents extrants, le devis s'appuyait principalement sur l'examen de la documentation, les sources importantes de données secondaires (p. ex. rapports d'activité, rapports de rendement), ce qui palliait la difficulté d'interagir avec l'ensemble des groupes d'intérêts.

Extrait de verbatim

"I think one of the challenges is that there are like in different X systems in operation across Canada. And this organization is tasked to sort of service all of them to some greater or lesser extent. So, a big part of trying to include all those stakeholder groups was thinking about you know how you could sort of reach them in a meaningful way. I didn't design this methodology, I just participated in its execution, I'm basing this on some things that we've talked about during the project. So, one of them was by using the program contact database. So, the survey component for example was designed for many different groups. So, for both [clients] and [professionals/staffs] and administrators and different stakeholder groups and so it was a long and complex survey, portions of which would be skipped depending on which group you belong to. One way that they sort of got around the complexity and talk to all the people you need to talk to is by doing a survey where we collect a lot of responses for a relatively low cost. [...]. Trying to reach all the sorts of people that they're supposed to be interfacing with was quite difficult, a lot of it had to be done through documentation. So, we might not have actually effectively reached all groups but a lot of the stuff about activities and outputs, we were mainly accessing it through documents that were given to us saying that this is sort of what they had done in terms of developing curriculum or that kind of stuff like annual reports, performance reports, sources of secondary data." ('J2')

Par ailleurs, face à la limite énoncée suivante, « *[d]ans de nombreuses évaluations, l'attribution [des résultats] constitue un défi fréquent, particulièrement en ce qui concerne les résultats à moyen et à long terme, étant donné qu'une multitude de facteurs peuvent influencer l'atteinte des résultats et il n'est habituellement pas possible de concevoir des évaluations expérimentales ou rigoureuses (p. ex. des évaluations qui incluent des hypothèses).* », 'J2' explique que mesurer l'efficacité implique d'être en mesure de prouver que le programme à des effets sur l'ensemble des territoires, des systèmes dans le domaine concerné à travers le Canada, et que c'est une chose impossible. De plus, il précise que les activités de l'organisme évalué ne représentent qu'une portion des services visant la gestion de la problématique ciblée, aussi prouver que le programme a bel et bien des effets est difficile. En outre, si pour lui la solution était de faire une comparaison entre un groupe expérimental et un groupe contrôle, il reconnait que le recours à un tel devis est beaucoup trop coûteux en termes de ressources. En revanche, il n'est pas en mesure de décrire la manière dont l'équipe évaluative s'y est prise pour gérer ces défis dans le cadre de l'évaluation de l'efficacité du programme. Remarquons que dans le cadre de la présentation des limites méthodologiques dans le rapport d'évaluation final, il est fait état de stratégies d'atténuation mobilisées pour être en mesure de rendre compte dans quelle mesure le programme a contribué à l'atteinte des résultats – soit la réalisation d'une analyse de contribution telle que définit par l'IGF.

Extrait de verbatim

"I think it is just coming down to the fact that you know at its core the evaluation is asking you know is this effective. But this effort is spread so widely across the [X] system that it's almost impossible to say whether or not something is safe. Even if you can observe change in which they were involved right because what [the organization] do is only a small part of the puzzle. So, if we can look at whatever small piece and like look at in one [site] at once you know like oh well they implemented [the organization] tools let's say and then there [clients] safety increased well. That clearly wasn't the only thing that [site] did during that time right. They've obviously you know someone has a group that said we're going to implement this and how they implemented it mattered, what else they implemented too. They have some funding level. They exist in some [X] system. And so, I think if they want to truly answer if things are being effective, what they need to do is pick two sites that share common characteristics and compare them to each other over time and match those characteristics. I think that's what they are trying to sort of driving. But typically, evaluations don't do that because, you know, it's extremely resource intensive." (J2)

Si l'impossibilité d'entreprendre un devis expérimental dans le cadre de l'évaluation d'un programme fédéral semble être une opinion commune, 'J2' n'est pas en mesure de nous expliquer pourquoi, selon lui, la complexité du programme empêche la réalisation d'une évaluation rigoureuse. Or, nous avons discuté de certains freins auxquels il aurait pu faire référence et qui pourraient justifier ce propos (questionnaire trop long et compliqué, faible taux de réponse à l'enquête, contraintes ESR - problèmes d'accès aux données de rendement dû à l'absence d'un système de suivi, contraintes de temps et d'argent, etc.).

Le cas de 'J3'

Pour 'J3', la complexité repose principalement sur la bonne compréhension et appréhension des rouages du programme évalué afin d'être en mesure de bien observer les activités et les résultats ciblés. Des programmes gouvernementaux comme celui décrit par 'J3' sont d'un niveau de sophistication très élevé, mobilisant un nombre très important de partenaires responsables de différents secteurs, volets et activités. Tout ceci implique donc de comprendre et d'être en mesure de décrire la logique du programme (les interactions entre les différents partenaires et activités, etc.), et ce, afin d'être en mesure de l'évaluer. Par exemple, « *looking at the interaction between two agencies – how does that specific interaction provoke the program intended outcome? How does it facilitate the program achieving its goals? You've got a lot of moving parts, so you need a deep understanding of the mechanics of it all.* » (J3). Aussi, l'évaluateur novice aborde le coût important du devis en termes de ressources (c.-à-d. temps, argent et main-d'œuvre) face à l'ampleur des données à recueillir, des sources d'informations à consulter, etc. S'il ne le mentionne pas comme tel, il s'agit ici de veiller à la fiabilité, la représentativité et l'utilité des résultats. Ainsi, le devis d'évaluation comprenait l'examen de plus de trois mille

documents (documentation du programme, divers rapports sources de données secondaires sur les activités, les extrants et résultats atteints, littérature grise, écrits scientifiques, etc.), l'entrevue de plus de près de 60 informateurs clés ainsi que la réalisation de deux enquêtes.

Extrait de verbatim

"The necessary deep level of understanding of the mechanics, the role and responsibility of each partner, etc. I'm thinking here more in terms of resource requirements, it does impact the design. The large number of partners means you need to collect a lot more data. Whenever you write a statement in the report, you need to think if it's true not just with reference to the one partner but with reference to all of them. There is a lot of complexity involved. For example, if you find that a program component is not working how it should, you need to diagnose the problem and identify where it comes from. You need to think carefully about how each individual partner plays a role in the overall functioning of the program and how potentially one or more than one, potentially all of them can be responsible, and why and how to help them improve. So yeah, the document and literature review were pretty massive, we interviewed more than 60 key informants, and also had two surveys." (J3).

Par ailleurs, en référence par exemple aux quelques milliers de documents analysés, 'J3' nous explique qu'une telle entreprise est possible aujourd'hui grâce à l'utilisation d'un logiciel de gestion de référence qui permet non seulement de gérer et d'organiser la documentation recueillie, mais aussi d'effectuer des analyses de contenu. En effet, le recours à un tel outil semble augmenter significativement le rapport coût-bénéfice d'un tel choix méthodologique ainsi que la qualité du devis d'évaluation, et la rigueur méthodologique.

Extrait de verbatim

"We had thousands of documents and literature so, the use of bibliographic software like Zotero to manage, classify and analyze these documents saved a lot of resources. Before, we would not have bothered to go this deep in terms of data sources because it cost too much in terms of time, budget, but also manpower. We would have just created a file structure and we would have tried to organize the documents data into those files, duplicate them if they belong to multiple files which meant reanalyzing them sometimes... Not a good way to do things. With Zotero, in my opinion, it's a big improvement. Basically, you could take all three thousand documents, put them in one place, and you can sort of accumulate all the bibliographic data, which isn't necessarily a big problem upfront, but when you start the final report, you have to cite a thousand different sources. And that's really important for transparency, but also reliability, you need to be able to say where the evidence you are using is coming from. It's an organizational tool where you can have a document that exist simultaneously in multiple files, and I have these sorts of tags applied so that you can actually search on a theme that is of interest. It allows you to do some fairly refined searches on the content that is in those files, so you can actually run the keyword searches, for example, all journal articles that were written within a certain span of time that has a particular word in them. So, our document review is more systematic, rigorous. It's quite a sophisticated tool, not just for collecting bibliographic data, but also for managing the enormous number of documents that are involved here. It reinforces the quality of the analysis process, the efficacy of data management, etc. It strengthens the design." (J3)

<u>*Le cas de 'J4'*</u>

Le premier élément de complexité abordé par l'évaluateur junior est la phase d'acquisition de nombreuses connaissances, dans un temps assez court, sur les différents corps professionnels ciblés par le programme, et ce, afin de saisir pleinement les enjeux spécifiques à chacun et être à

même d'avoir une conversation constructive avec les différents professionnels consultés dans le cadre d'entrevues. En effet, 'J4' nous explique que, dans le présent contexte, les intervieweurs ne pouvaient pas s'appuyer uniquement sur leurs connaissances techniques en tant qu'évaluateurs pour mener à bien cette activité de collecte de données. Ils ont misé sur l'utilité des études de cas qui sont un très bon outil pour acquérir une connaissance et une compréhension assez profondes des composantes clés des projets financés, de leurs spécificités, la manière dont ils opèrent, leur fonctionnement avec le programme, etc. Ainsi, 'J4' souligne le caractère pertinent de répartir les différentes entrevues d'informateurs clés en fonction de l'attribution des études de cas.

Extrait de verbatim

"I think there was a bit of a learning curve and so what we did given that, we assigned specific occupations to specific members of the evaluation team. Some were in charge of specific interviews per occupation. Like maybe three people. Case studies were also very useful to prepare with their document analysis component, and we really dig deep, try to have a good understanding about the projects that are being funded, their key aspects, the mechanism, its relation to the program, etc. Some evaluation you can just get by with your technical knowledge and you being an evaluator, but to have a meaningful conversation with someone in some specific occupation with critical aspects in terms of characteristics, you need to understand the challenges in depth. So, the challenge here was that there was a lot of knowledge for us to soak up in a somewhat limited timeframe. I think we did a good job assigning interviews for example based on how the case studies were assigned as well. I did all the interviews for the case study and key informant interviews related to the [X] occupation." (J4).

'J4' explique que le programme finance des projets d'implantation d'un modèle au sein de différents corps professionnels qui sont chacun sous la gouvernance d'un organisme à l'échelle pancanadienne et d'organismes provinciaux, et qu'on observe par ailleurs des différences sur le plan de l'implantation du modèle d'une province à l'autre. À la lumière de son expérience de rédaction du rapport sur un corps professionnel (rapport de cas), il explique que face à la diversité sur le plan des corps professionnels financés, de la gouvernance et de l'implantation du modèle, il peut s'être avéré difficile d'émettre des conclusions évaluatives générales (ou « *blanket statements* »), représentatives de l'efficacité à l'échelle du programme.

Extrait de verbatim

"It was quite complex for sure. We had to speak to provincial bodies and pan-Canadian bodies as well. I don't think that was proved to be too difficult but again thinking of the case study I wrote; you could talk to one occupation like [X] profession and you could talk to a national body or something and few provincial bodies as well. In terms of implementation, there were differences between provinces which highlight the level of complexity. I didn't read the report in depth, but it was difficult to make a blanket statement given the variety of diversity across occupations and across provinces. So, it's difficult to make a blanket statement like - the program was successful because blah." (J4).

Explicitations des répercussions de la complexité du programme sur la conception et la mise en œuvre du devis d'évaluation - Le cas de 'J1'

- Un mandat évaluatif qui manque de clarté et de précision
- Une approche dont la théorie et la mise en application demeurent plutôt abstraites
- Une problématique ciblée qui s'avère compliquée à définir
- Un échéancier très serré au regard de la complexité du projet évaluatif

'J1' en tant que lead junior sur ce projet d'évaluation fut le collaborateur principal de l'évaluateur sénior responsable de l'évaluation externe, et il a participé à l'ensemble des discussions dans le cadre de la conception du devis d'évaluation ainsi que des différentes étapes méthodologiques. Aussi, il a été en mesure de nous expliquer de manière claire et précise en quoi la complexité du projet et de son évaluation a influencé le développement et la mise en œuvre du devis d'évaluation.

'J1' nous rappelle à plusieurs reprises qu'il ne s'agissait pas d'une évaluation typique du gouvernement fédéral en ce sens que l'approche évaluée au sein des organisations ne répondait pas à une théorie de programme concrète. Inscrite dans une perspective qui semble à la fois formative et exploratoire ainsi qu'un désir des clients de « *sortir des sentiers battus* », l'évaluation avait pour mandat d'offrir un modèle logique de l'approche avec des composantes précises et clairement définies, et une feuille de route ainsi qu'une boite à outils (adaptable selon le contexte organisationnel) visant à guider l'implantation et le monitorage de cette dernière au sein des agences et des départements du gouvernement fédéral.

Aussi, pour l'évaluateur junior, le principal défi reposait sur la tâche ardue de définir et de circonscrire les dimensions clés de l'approche sachant que son appréhension et implantation variait d'un contexte, d'une culture organisationnelle à une autre. En gardant cela à l'esprit, nous avons vu que les paramètres de l'évaluation définis par les évaluateurs étaient assez larges, et que le cadre méthodologique a été conçu à partir d'une matrice dont le contenu était organisé par thèmes d'investigation plutôt généraux (absence de questions d'évaluation, d'indicateurs précis). Toujours en réponse à ce désire et à cette nécessité de demeurer ouvert à la multitude de possibilités en termes de pratiques et à la diversité sur le plan des organisations, les évaluateurs ont misé sur la complémentarité de deux méthodes - l'entrevue d'informateurs clés (de type non dirigé) et l'analyse documentaire ainsi que le recours à l'échantillonnage de type boule de neige (chaque interviewé partage sa documentation et recommande des organisations ou des individus qu'il

reconnait pour leur expertise ou leurs bonnes pratiques), avec à mi-parcours, ou plutôt une fois que des tendances commençaient à émerger, un affinage des thèmes d'investigation avec la définition de sous-thèmes, une révision des contenus de l'instrumentation plus particulièrement du gabarit d'analyse de la documentation (ou matrice de données/preuves) avec notamment un réexamen des documents analysés jusque-là.

Extrait de verbatim

"We used document review and key informant interview, and well, they are kind of tangled together. [...]. Remember for typical evaluation you would have clear indicators. But this one we didn't want to narrow ourselves because we didn't know what was out there. We started with general investigative themes, we looked at certain things like culture and practices. At first, we reached out to key informants from federal agencies recommended by the clients. The interviews we conducted were more conversational style, so we wouldn't limit ourselves. We would also ask them to share all relevant internal documentation they could about what their practices, as well as any available gray and scientific/academic literature. We asked if they were some research or researcher that they like and rely on. And also, we would ask them to recommend other organizations or practitioners, in the public or private sector, whom they recognize for their expertise with regard to the approach X. We kept snowballing like that, kept going until the timeline it. [...]. So, yeah, as we started doing more interviews and kind of fine-tuning those matrix headers and their themes, it helped us kind of narrow it down especially for the document literature review which, you know, can get really big, really fast. So yes progressively... Based on the content of the document review data matrix, and the interview data matrix, we would reassess the content of the evaluation matrix midway through the interviews, fine tune in those questions that guided us, define more specific indicators/headers/sub-headers. Then we would complete the data collections. For the document literature review, it also meant digging deeper into the documentation already collected and reviewed." (J1).

L'évaluateur junior 'J1' souligne l'importance de la souplesse des clients, qui habituellement dans le cadre d'évaluations au fédéral fournissent une liste complète (préétablit) d'informateurs clés avec un nombre précis d'entrevues préapprouvé. Ici, les évaluateurs avaient en quelque sorte carte blanche, avec finalement pour seule contrainte, le temps. Aussi, en dépit de son caractère chronophage, la décision de recourir à l'échantillonnage de type boule de neige constituait pour 'J1' le meilleur choix possible, un choix qui en valait la peine puisqu'il a permis l'obtention d'un échantillon qu'il qualifie de bonne qualité avec un spectre plus élargi que s'il avait opéré de la manière standard. Il réitère à plusieurs reprises que dans un tel contexte, le succès de l'évaluation était fortement tributaire de la collaboration étroite, et de la communication constante avec les clients, plus particulièrement l'équipe d'évaluation interne qui facilitait grandement l'accès aux sources d'information.

b. Les choix de méthodes de collecte de données

Selon les évaluateurs juniors, les différentes méthodes de collectes de l'information sélectionnées correspondent aux sources de données standards dans le cadre de l'évaluation de

programmes au gouvernement fédéral. Il semble s'agir de choix "par défaut" et (ou) prescrits dans la demande de propositions ('J4'), dans la proposition initiale de la division de l'évaluation suivie de la réponse de l'équipe évaluative externe ('J3'), ou encore dans le mandat d'une évaluation indépendante des obligations formelles de la politique évaluative ('J2') (Tableau 23, p. 245). Ce qui est intéressant ici, c'est que 'J1', qui a travaillé en tant que collaborateur sur un projet d'évaluation hors politique évaluative et commanditée par des clients ayant un désire d'approche innovante, fait tout de même référence aux méthodes standards quand il explique qu'il s'agit d'une évaluation plus petite et qu'ils n'ont pas eu recours à l'ensemble de sources de données habituelles, seulement à certaines d'entre elles. Pour les évaluateurs juniors ayant une formation initiale en recherche et (ou) en évaluation de programme, outre un choix de méthodes qui est standard ou par défaut, il est également question de miser sur : la complémentarité des méthodes (J1); la combinaison de méthodes quantitatives et qualitatives (J4); ou encore la triangulation et la maximisation des preuves (J2). Ce qu'il est intéressant de noter ici, c'est que malgré son rôle d'exécutant de collectes de données, 'J4' connait (comprends) l'intention derrière la triangulation des sources de données (et par extension des sources d'information), et formule une réflexion pertinente en ce qui a trait à la complémentarité entre les méthodes qualitatives et les méthodes quantitatives. Dans le cas de 'J2', lorsqu'il mentionne le recours à plusieurs sources afin d'accéder de différentes façons à la même information, il semblerait qu'il soit question de sa perception de l'intention de triangulation, et ce, sans qu'il ait nécessairement connaissance de ladite intention (il semble émettre ici une hypothèse). En revanche, nous avons vu dans le cadre de la discussion relative aux enjeux méthodologiques face à la complexité du programme que 'J2', témoin de plusieurs discussions au cours du projet, avait été en mesure de nous expliquer selon lui les raisons quant au recours à l'enquête, à savoir rejoindre le plus de parties prenantes clés du programme (utilisateurs ciblés inclus), et ce, à un moindre coût.

Enfin, il ressort des propos des évaluateurs juniors 'J2', 'J3' et 'J4', qu'en raison de l'envergure importante des différents programmes évalués, l'analyse de documents et de données secondaires constituait la source principale de données pour rendre compte du rendement du programme (efficacité et économie).

Tableau 23. *Explications ou perceptions du choix des méthodes de collectes de données (évaluateurs juniors)*

	Contexte numéro 1		Contexte numéro 2	Contexte numéro 3
	'J4'	'J3'	'J2'	'J1'
	Juniors lead exécutants d'un nombre important d'activités de collecte et (ou) d'analyse de données.		**Junior exécutant** d'un nombre limité d'activités de collecte et (ou) d'analyse de données.	**Lead junior, principal collaborateur** de l'évaluateur sénior partenaire chargé du projet d'évaluation externe.
Recours à de multiples sources de données	• Choix des clients - prescrit dans la demande de propositions. • Par défaut - méthodes standards typiques d'une évaluation d'un programme fédéral. • Recours au devis d'évaluation antérieure. • Complémentarité des méthodes quantitatives et qualitatives.	• Choix de méthodes standards pour l'évaluation d'un programme fédéral. • Choix des clients dans la demande de propositions.	• Par défaut - méthodes standards, typiques d'une évaluation d'un programme fédéral. • Afin de trouver plusieurs manières d'accéder à la même information. • Afin de collecter le plus de preuves, possible.	• Choix de méthodes standards pour l'évaluation d'un programme fédéral. • Complémentarité des méthodes - utilisation de deux différentes méthodes de manière séquentielle, la première ayant pour extrant les sources d'information consultées dans le cadre de la deuxième.
Choix des méthodes	Examen de dossiers et de documents • Collecter des données secondaires pour évaluer la pertinence, l'impact et la rentabilité du prog. de financement. • Recueillir des données sur les activités financées par le prog., leurs extrants et les résultats associés afin de prouver l'efficacité, le rendement du prog. de financement. Entrevues d'informateurs clés • Couvrir l'ensemble des questions d'éval. Enquête auprès des bénéficiaires d'un financement • Répondre aux questions sur l'efficacité du prog. Études de cas : • Observer l'efficacité et le rendement à l'échelle des projets financés (données secondaires et perception des informateurs clés).	Analyse documentaire • Dresser le profil du programme. • Recueillir des données secondaires pour l'éval. du rendement du prog. Recension des écrits • Situer le prog. dans les écrits. • Faire une comparaison interjuridictionnelle. • Évaluer le besoin continu du prog. Enquête auprès de membres de l'industrie • Répondre aux questions d'éval. de rendement. Enquête auprès des professionnels utilisateurs/prescripteurs et consommateurs • Recueillir des informations sur la pertinence et l'efficacité. Entrevues d'informateurs clés • Couvrir l'ensemble des questions d'évaluation au regard de la perspective des différents groupes de parties prenantes clés du prog. Études de cas • Toujours incluses dans une évaluation.	Analyse documentaire • Répondre à l'ensemble des questions d'éval. • Dresser le profil et l'historique du programme. • Recueillir des données secondaires comme preuve de l'efficacité (palier au non-recours au devis expérimental). Recension des écrits • Situer le prog. dans les écrits, et ce qui est recommandé en termes d'aspects clés du prog. et de ses services. • Faire une comparaison interjuridictionnelle. Entrevues d'informateurs clés • Obtenir la perspective des parties prenantes du prog. Enquête auprès des parties prenantes clés • Recueillir la perspective d'une grande d'informateurs clés internes et externes sur la pertinence et l'efficacité du programme. • Pallier le non-recours à un devis expérimental ou quasi expérimental, la complexité du programme (envergure). Études de cas • Toujours proposées, mais pas toujours incluses dans le devis final en raison de son caractère chronophage et laborieux (rapport coût-bénéfice).	Entrevues d'informateurs clés (internes et externes) • Comprendre la manière dont est définie, conceptualisée, implantée et monitorée l'approche X au sein de l'organisation. • Recueillir des exemples de bonnes pratiques. • Recueillir l'ensemble de la documentation interne et littérature grise générée/ utilisée. • Identifier les écrits/auteurs de référence sur lesquelles l'organisation appuie son approche. • Identifier des experts/organisations « modèles » qu'il serait pertinent de consulter. Analyse documentaire/Recension des écrits • Évaluer les preuves de l'approche X au sein des organisations. • Offrir des pistes de recommandation sur le monitorage de l'implantation de l'approche X au sein d'une organisation gouvernementale.

Extrait de verbatim

« O*n aime utiliser un processus de triangulation, donc on va utiliser plusieurs méthodes qui vont chacune produire une matrice de données, de preuves. Et ensuite on va les comparer, voir ce que les différentes sources ont à nous dire, les répondants from different sites, beneficiaries, experts, people from program management, admin data, etc. We look at what each has to say à propos du programme. We compare, contrast, combine. Quantitative data doesn't tell you the whole story, you can't see some things... you paint a picture but you don't necessarily have the details and with such a complex program with so many occupations, context, sites, stakeholders, you can look at the big numbers and be like this is how they spend the funds, they spend this much, etc. but without the story, the description, the explanations, I don't think you have much of an evaluation. That is where the balance comes into play. Using the quantitative data but enriching it with the qualitative data that really tells the story.* » (J4).

Extrait de verbatim

"I think one of the challenges is that there are like in different X systems in operation across Canada. And this organization is tasked to sort of service all of them to some greater or lesser extent. So, a big part of trying to include all those stakeholder groups was thinking about you know how you could sort of reach them in a meaningful way. I didn't design this methodology, I just participated in its execution, I'm basing this on some things that we've talked about during the project. So, one of them was by using the program contact database. So, the survey component for example was designed for many different groups. So, for both [clients] and [professionals/staffs] and administrators and different stakeholder groups and so it was a long and complex survey, portions of which would be skipped depending on which group you belong to. One way that they sort of got around the complexity and talk to all the people you need to talk to is, by doing a survey where we collect a lot of responses for a relatively low cost. [...]. Trying to reach all the sort of people that they're supposed to be interfacing with was quite difficult, a lot of it had to be done through documentation. So, we might not have actually effectively reached all groups but a lot of the stuff about activities and outputs, we were mainly accessing it through documents that were given to us saying that this is sort of what they had done in terms of developing the curriculum or that kind of stuff like annual reports, performance reports, sources of secondary data." ('J2').

c. La sélection des sources d'information et procédures d'échantillonnage

Pour les évaluateurs juniors exécutants ('J2', 'J3' et 'J4'), cette composante du cadre méthodologique repose presque entièrement sur les choix du comité d'évaluation interne, plus particulièrement des gestionnaires du programme, et selon 'J3' sous les conseils de la division interne de l'évaluation. Certains détails notamment en ce qui concerne la portée de la recension des écrits (comparaison interjuridictionnelle, etc.), le nombre d'informateurs clés à interviewer et le nombre de cas à étudier sont souvent mentionnés dans la demande de propositions. Dans le cadre de la planification des procédures de collectes de données, les clients semblent fournir la liste des informateurs clés à consulter, l'ensemble de la documentation, et déterminent les banques de sources d'information à consulter (bases de données, listes de diffusion, etc.). En outre, comme on peut le voir dans le Tableau 24 (p. 247), lorsque nous avons relancé ces évaluateurs juniors sur le sujet, ils étaient en mesure de fournir des informations sur les critères de sélection, les procédures d'identification et de sélection des sources d'information. Aussi, les décisions semblaient finalement répondre à un compromis entre les critères d'échantillonnage déterminés par l'équipe évaluative externe (et le contenu de la matrice évaluative), le choix des clients (le comité d'évaluation interne), l'existence de bases de données, de listes de contact de parties prenantes clés et groupes ciblés par le programme ainsi que les intrants disponibles.

Tableau 24. *Perceptions des procédures d'identification et de sélection des sources d'information (évaluateurs juniors 'J2', 'J3' et 'J4')*

	Contexte numéro 1		Contexte numéro 2
	'J4'	'J3'	'J2'
	Juniors lead exécutants d'un nombre important d'activités de collecte et (ou) d'analyse de données.		**Junior exécutant** d'un nombre limité d'activités de collecte et (ou) d'analyse de données.
Documentation	• À l'aide de la matrice évaluative, les « gens du programme » font parvenir l'ensemble de la documentation jugée pertinente (rapports d'activités et financiers, descriptions de programme, bases de données, politiques, etc.) o Au fil du cycle d'analyse documentaire, l'équipe évaluative externe relance les clients si un complément de documentation est nécessaire. o Importance d'obtenir des données administratives, données secondaires permettant de rendre compte du rendement du programme (principales sources de données éval. efficacité, efficience et économie)		
Écrits recensés	• Mise à jour de la recension des écrits de la précédente évaluation. • Utilisation d'un moteur de recherche.	• Critères de sélection o Écrits scientifiques et académiques. o Littérature grise. o Liste de recherche par mots-clés. • Utilisation d'un moteur de recherche et d'un logiciel de gestion de réf.	• Critères de sélection o Écrits scientifiques - revues systématiques, méta-analyses, comparaisons interjuridictionnelles sur les thématiques centrales ciblées par le prog. • Portée de l'échantillon limitée par le budget et l'échéancier. • Utilisation d'un moteur de recherche et d'un logiciel de gestion de référence - sélection des articles à partir du contenu de la section résumé.
Informateurs clés	• Taille échantillon déterminée selon le budget éval. et l'accès aux données (complémentarité des sources d'information). • Critères de sélection - « *broad coverage* » + pertinence/utilité. o Gestionnaires, employés, clients/parties prenantes (autres départements, organismes financés dans secteurs ciblés, experts, etc.). • Liste des informateurs clés fournie par gestionnaires de prog. - possible entente entre l'équipe éval. ext. et les clients pour valider la liste finale et le nombre d'entrevues par groupe ciblé.	• Notification de la liste des informateurs clés dans la demande de propositions, ou au moment de la validation du devis d'éval. par le CMI.	• Critères de sélection : o Triangulation des sources o Consultation des individus à chaque niveau de l'organisation et (ou) avec un niveau d'implication différent dans l'organisation. • Liste d'informateurs clés fournie par les clients.
Informateurs sondés	• Critère de sélection : o Organismes financés par le programme (ceux ayant fait l'objet d'entrevue exclus) • Pas d'échantillonnage – enquête disséminé via Listserv.	• Critères de sélection : o *Enquête 1* - membres de l'industrie ayant manifesté intérêt ou participé au prog. o *Enquête 2* - prof. utilisateurs et indiv. consommateurs de produits. • Échantillon non aléatoire – Listserv. « parties prenantes clés » de l'agence gouv.	• Critères de sélection : o Triangulation des sources. o Consultation des professionnels et membres de la population ciblée par prog. • Questionnaire envoyé à l'ensemble des adresses courriel enregistrées dans bases de données d'informateurs clés externes - échantillonnage via fonction 'branchement conditionnel'.
Cas	• Échantillonnage non aléatoire : o Sélection des cas ayant participé à la précédente évaluation du programme - perspective longitudinale.	• Critères de sélection déterminés par l'EC à partir de l'analyse documentaire, le modèle logique du programme, les objectifs évaluatifs, etc. • Identification initiale des cas par l'équipe éval. ext., puis consultation avec l'éval. interne. Ultimement, la décision appartient au CMI.	• Identification et recrutement des cas par le CMI.

Dans le cas de 'J1', *lead junior* (collaborateur principal), les décisions furent prises en collaboration étroite avec le comité d'évaluation interne, avec l'équipe évaluative interne qui s'assurait que les choix proposés étaient faisables, et qui facilitait l'identification, et l'accès aux diverses sources d'information. La complexité de l'initiative à l'étude et la nécessité, pour le bien de l'évaluation, de rester ouvert à la multitude de possibilités en termes de pratique ainsi qu'à la

diversité sur le plan des organisations, a conduit les évaluateurs à utiliser l'échantillonnage de type boule de neige - chaque informateur clé interviewé partage l'ensemble de la documentation pertinente (documents internes, littérature grise, et/ou écrits scientifiques sur lesquels ils se sont appuyés), et recommande des organisations ou des individus qu'il reconnaisse pour leur expertise ou leurs bonnes pratiques (Tableau 25). 'J1' reconnait l'importance de la souplesse des clients dans le présent projet, puisque ce type d'échantillonnage n'est pas une pratique standard dans le contexte de l'évaluation au fédéral, les clients fournissant normalement une liste complète préétablie d'informateurs clés avec un nombre précis d'entrevues prédéterminé.

Tableau 25. *Explicitations des procédures d'identification et de sélection des sources d'information (évaluateur junior 'J1')*

Sources d'information	Critères de sélection des sources d'information	Procédures de sélection des sources d'information
Informateurs clés	Chercheurs, praticiens, gestionnaires (service public ou secteur privé). • Identifiés comme utilisant ou connaissant des modèles/processus/outils liés à l'approche X dans les organisations. *Ou* • Qui ont développé des modèles/processus/outils liés à la promotion et à l'évaluation de l'approche X dans les organisations.	Échantillonnage de type boule de neige • À l'appui des thèmes d'investigation généraux (matrice), le gestionnaire de l'évaluation externe en collaboration avec le comité consultatif interne identifie une première vague d'organisations internes (gouv. fédéral) et leurs informateurs clés. • Chaque informateur clé est invité à o Partager l'ensemble de la documentation pertinente disponible. o Recommander d'autres organisations/informateurs clés, qu'ils reconnaissent pour leur expertise ou bonnes pratiques. o Identifier des auteurs et (ou) écrits/ouvrage de référence. • Utilisation d'un moteur de recherche en ligne pour compléter la recherche de littérature grise et d'écrits scientifiques (en lien avec les informations/suggestions fournies par les informateurs clés). Taille des échantillons déterminée par la limite de temps - « *we kept going until our timeline hit* » • 18 informateurs clés interviewés, 80 documents analysés.
Documentation	L'ensemble des sources documentaires permettant d'évaluer les preuves de l'approche X au sein des organisations. • *Documents internes* - rapports, présentations sur l'approche X, etc. • *Littérature grise* - pages web, livres, présentations, critiques et synthèses de livre, etc. • *Articles scientifiques*	

d. Les procédures de collecte et d'analyse de données

Selon les évaluateurs juniors exécutants, l'ensemble des procédures de collecte et d'analyse de données sont déterminées par l'évaluateur sénior responsable de l'équipe évaluative externe, qui soumet des rapports techniques au comité d'évaluation interne, plus précisément de la division de l'évaluation dans le cas de 'J3', pour validation.

Si les évaluateurs juniors exécutants ('J2', 'J3' et 'J4') ne sont pas en mesure d'expliquer les divers choix méthodologiques, ils demeurent capables de décrire avec précision les objectifs

des collectes de données auxquelles ils ont participé. En revanche, tous ne font pas preuve du même degré de clarté et de précision en ce qui a trait à l'instrumentation et les procédures de passation.

'J2' réitère ici qu'il travaille en silos sur l'exécution de tâches de collectes et (ou) d'analyse qui lui sont assignées, sans savoir (la plupart du temps) les processus décisionnels et méthodologiques exacts, entrepris en amont. Il parle clairement d'hypothèse lorsqu'il aborde la conception du questionnaire d'enquête. Il mentionne l'utilisation d'un logiciel de gestion de référence pour la recension des écrits et l'analyse documentaire pour organiser la documentation et coder l'information, mais n'aborde pas les raisons quant à ce choix.

Dans les cas de 'J3' et 'J4', il s'est avéré difficile de faire la distinction entre leurs connaissances concrètes concernant les motivations sous-jacentes quant à la planification des procédures de collectes de données, à la démarche de conception de l'instrumentation, les procédures de passation, dans le cadre de leur évaluation respective, et leurs hypothèses (suppositions) basées sur des évaluations antérieures, ou encore ce qu'ils perçoivent comme des habitudes méthodologiques, des procédures systématiques, routinières.

Dans le cas de 'J1', en tant que collaborateur, ce dernier a participé à l'ensemble des discussions et des différentes étapes dans le cadre du développement et de la mise en œuvre de la méthodologie évaluative. Il est en mesure d'expliciter clairement et avec précision chacun des éléments de procédures méthodologiques. Il nous explique que c'est toujours l'évaluateur sénior (responsable du projet d'évaluation) qui construit l'ensemble des instruments de collectes de données, mais qu'il a été sollicité pour réviser leurs contenus. Il aborde également la raison pour laquelle le guide d'entrevue est non directif ainsi que la rencontre de préparation de l'équipe d'interviewers, une phase qu'il décrit comme particulièrement essentielle compte tenu notamment du fait que ce sont généralement des évaluateurs plus juniors intégrés au projet à la dernière minute (au moment de la collecte des données).

Enfin, on retrouve au cœur du discours des évaluateurs juniors, qu'ils soient exécutants ou collaborateurs, un élément clé, la matrice évaluative, qui est perçue à juste titre comme le principal outil et point de référence méthodologique dans la conception des instruments, la gestion et l'organisation des données, et les procédures d'analyses de données et de synthèse des résultats (Tableau 26, p. 250 ; et Tableau 27, p. 251).

Tableau 26. Description ou explicitation des procédures de conception et de passation de l'instrumentation *(évaluateurs juniors)*

	Contexte numéro 1		Contexte numéro 2	Contexte numéro 3
	'J4'	'J3'	'J2'	'J1'
	Juniors lead exécutants d'un nombre important d'activités de collecte et (ou) d'analyse de données.		**Junior exécutant** d'un nombre limité d'activités de collecte et (ou) d'analyse de données.	**Lead junior, principal collaborateur** de l'évaluateur sénior partenaire chargé du projet d'évaluation externe.
Instrumentation	Conception par l'éval. sénior partenaire chargé de l'éval. ext. avec rétroaction puis validation du contenu par les clients pour s'assurer que tout est couvert. Entrevue d'informateurs (suppositions) • Extraction des sous-questions d'évaluation et des indicateurs correspondants répertoriés dans matrice éval. • Contenu du schéma d'entrevue semi-dirigé adapté aux caractéristiques du répondant ou groupe de répondants ciblé (*conception de plusieurs guides - axes de questionnement similaires*) • Préparation de guides « externes » (version abrégée - sans les questions secondaires, les indicateurs, etc.). • Utilisation d'un gabarit pour retranscrire les notes d'entrevue - matrice évaluative intégrée dans un fichier de feuille de calcul en tant que matrices de données/preuves. Examen des fichiers et documents • Utilisation d'un gabarit - matrice de données (chiffrier électronique).	Conception par l'éval. sénior chargé de l'éval. ext. et validation par le comité d'évaluation interne. Recension des écrits et de la documentation • Intégration des axes de questionnements et indicateurs du « *workplan* » dans un logiciel de gestion de références. • Recours à un logiciel de gestion de référence essentiel vu l'envergure de l'échantillon (n= +3000) – rigueur gestion des informations, profondeur du processus d'examen du contenu de la documentation collectée. Guide d'entrevue des informateurs clés et enquêtes (supposition) • Contenu développé en anglais puis traduit en français par un membre de la firme francophone ou qui parle le français (selon les extrants disponibles, et le souhait des clients, possible recours à un traducteur professionnel).	Conception par l'évaluateur sénior chargé de l'éval. ext. Questionnaire (hypothèses) • Négociation avec les clients - acte d'équilibre entre i) prioriser les items qui obtiendront le plus haut taux de réponse ; ii) s'assurer que les items choisis permettent l'obtention des informations ciblées ; iii) maintenir l'attention des clients sur les questions d'éval. et non sur d'autres informations qu'ils désirent obtenir ; iv) optimiser le temps de réponse. • Items conçus anglais puis traduits en français par un éval. de l'équipe ext. francophone, par un traducteur professionnel, ou encore par les clients avec validation par l'équipe éval. ext. (choix des clients). Recension des écrits et analyse documentaire • Utilisation d'un logiciel de gestion de références où sont intégrés les questions d'évaluation, les thèmes d'investigation.	Conception à partir des thèmes d'investigation définis pendant et à la suite de la rencontre de travail initial avec le comité consultatif interne. Conception par l'éval. sénior chargé de l'éval. ext., avec révision du contenu par l'éval. *lead junior* (J1). Entrevue d'informateurs clés • Utilisation d'un guide d'entrevue non directif – thèmes généraux de discussion pour ne pas limiter l'étendue des informations collectées (complexité de l'approche évaluée – diversité sur le plan de sa modélisation, de son implantation (différents types d'organisation, différentes cultures organisationnelles, etc.). • Utilisation d'un gabarit pour retranscrire les notes - la matrice éval. intégrée dans un chiffrier électronique en tant que matrice de données. Recension des écrits et de la documentation • Utilisation d'un gabarit - matrice de données (chiffrier électronique). • Affinage des thèmes d'investigation, définition de sous-thèmes/indicateurs à mi-parcours à la lumière des informations obtenues dans le cadre des différentes collectes de données.
Passation	Examen de la documentation • Extraction et transcription des informations pertinentes dans la matrice de données. • Souvent réalisé par les évaluateurs plus juniors. Entrevue • Mobilisation d'une équipe d'interviewer (souvent composés d'évaluateurs plus juniors) o Distribution des entrevues par champ professionnel ciblé par le programme • Pas d'enregistrement des entrevues — prise de notes (idéalement directement dans la matrice de données) o Difficultés d'obtention du consentement des informateurs ciblés. o Potentiel objet de demande d'autorisation spéciale. o Transcription des verbatim trop chronophage. • Envoie du guide d'entrevue externe à l'interviewé avant la passation. Enquête (supposition) • Dissémination du lien du questionnaire en ligne par courriel – utilisation des listes de diffusions internes (*listerv.*).	Analyse documentaire et recension des écrits • Codage de l'information dans le logiciel de gestion de références où l'ensemble de la documentation a été téléchargée et classée. • Gestion, organisation de la documentation et des données via logiciel de gestion de référence. • Souvent réalisés par les évaluateurs plus juniors. Entrevue • Mobilisation des évaluateurs de la firme qui parlent français si des informateurs ciblés sont francophones. • Passation par téléphone de la majorité des entrevues. • Pas d'enregistrement des entrevues — prise de notes. Enquête (supposition) • Dissémination du lien du questionnaire en ligne par courriel – utilisation des listes de diffusions internes (*listerv.*).	Analyse documentaire et recension des écrits • Codage de l'information dans le logiciel de gestion de références où l'ensemble de la documentation a été téléchargée et classée. • Souvent réalisés par les évaluateurs juniors. Enquête • Dissémination du lien d'accès au questionnaire en ligne via o les listes de diffusion internes (supposition). o les tableaux de mémos internes (hypothèse).	Coordination du déroulement des collectes de données par l'éval. lead junior (J1) sous la supervision de l'éval. sénior chargé projet éval. ext. Analyse documentaire/Recension des écrits • Extraction et transcription des informations pertinentes dans la matrice de données. Entrevue • Envoie d'une invitation générale à participer à l'évaluation par l'équipe d'éval. interne pour faciliter l'accès aux informateurs clés. • Prise de rendez-vous par la firme consultante avec guide d'entrevue intégré au courriel. • Mobilisation d'une équipe d'interviewers. • Pas d'enregistrement des entrevues — prise de notes et retranscription dans la matrice de données d'entrevues. • Organisation d'une rencontre de préparation de l'équipe d'interviewers, le plus souvent des évaluateurs juniors sollicités au dernier moment à titre d'exécutants.

En ce qui a trait au processus d'analyse des données qualitatives, la procédure la plus utilisée semble être le recours aux matrices de résultats (Tableau 27). Le recours, ou non, à un

logiciel d'analyse de données qualitatives (logiciel QDA) semble être dépendant du choix des clients et de la perception selon laquelle il est moins chronophage, et plus simple, de réaliser les analyses « manuellement » avec un tableau de feuilles de calcul. Dans le cas de 'J3', le recours à un logiciel QDA était selon lui indispensable pour la rigueur et la profondeur des analyses dans le cadre du traitement des données d'entrevue compte tenu de la taille importante de l'échantillon (60 entrevues réalisées). Idem, pour le recours à un logiciel de gestion de recherche qui permettait de classer et coder le contenu de la documentation examinée (3500 documents).

Tableau 27. *Explicitations ou perceptions sur les procédures d'analyse de données et de synthèse des résultats (évaluateurs juniors)*

	Contexte numéro 1		**Contexte numéro 2**	**Contexte numéro 3**
	'J4'	**'J3'**	**'J2'**	**'J1'**
	Juniors lead exécutants d'activités de collecte, et (ou) d'analyse de données.		**Junior exécutant** d'activités de collecte, et (ou) d'analyse de données.	**Lead junior, principal collaborateur** de l'évaluateur en charge.
Traitement des données qualitatives	<u>Écrits recensés, documentation collectée et notes d'entrevues retranscrites</u> • Utilisation d'un chiffrier électronique où sont intégrés les matrices de données - moins coûteux en termes de ressources et simple d'utilisation. o Extraction des informations, des extraits de notes d'entrevue considérées pertinentes et intégration dans les matrices de données à titre de citations. • Analyse à partir de la matrice de données qui devient une matrice de résultats. o Consultation des clients pour connaître leurs attentes quant à la formulation des résultats. o Utilisation de quantificateurs tels que l'ensemble, la plupart ou quelques répondants pour résumer les résultats. o Organisation des résultats par groupe de répondants/type d'informateur clé, et par question d'évaluation - comparer et contraster les résultats.	<u>Écrits recensés et documentation collectés</u> • Codage thématique à l'aide d'un logiciel de gestion des références basé sur le contenu de l'« *evaluation workplan* » - renforce la rigueur et la profondeur des analyses. <u>Notes d'entrevues retranscrites</u> • Préparation des données pour analyse o Traduction en anglais des notes prises en français - soit, par l'interviewer, soit par un traducteur professionnel ; <u>ou</u> l'interviewer réalise les analyses en français et résume les résultats clés en anglais. • Recours à un logiciel QDA pour synthétiser les données, classer les informations et apporter des réponses aux questions d'évaluation. o Idéalement, l'interviewer analyse ses propres données, mais les contraintes de temps font que le corpus de notes retranscrites est souvent partagé entre différents membres de l'équipe.	<u>Écrits recensés et documentation collectés</u> • Codage thématique à l'aide d'un logiciel de gestion des références. <u>Notes d'entrevues retranscrites</u> (hypothèses) • Si échantillon et étendue limités, les interviewers préparent un sommaire narratif. • Dans les cas plus complexes, analyse de contenu thématique avec processus de codage soutenu par l'utilisation d'un logiciel QDA. Ultimement le choix d'utiliser un logiciel QDA dépend : o des intrants disponibles o de l'ampleur du projet évaluatif o de la volonté des clients de pouvoir consulter les analyses, vérifier que toutes les informations utiles et pertinentes ont bien été codées, etc.	<u>Écrits recensés, documentation collectée et notes d'entrevues retranscrites</u> • Cycle itératif d'analyse thématique et de consolidation des thèmes et indicateurs des matrices de résultats. • Utilisation d'un chiffrier électronique où sont intégrées les matrices de résultats - outil d'analyse privilégié des clients. • Analyses des données supervisées par l'éval. sénior partenaire chargé du projet d'éval. ext. : o Distribution du corpus de notes, d'écrits et documents recensés, et répartition des thèmes d'analyse entre les membres de l'équipe évaluative - gestion des contraintes de temps. o Extraction et intégration des informations pertinentes, des preuves, dans les matrices de résultats (réponses aux questions d'éval.). o Validation du contenu des analyses, des résultats/interprétation par l'éval. sénior partenaire chargé du projet d'éval. ext.
Traitement des données quantitatives	• Recours à un statisticien pour les analyses statistiques.	• Recours à un logiciel d'analyse statistique.	• Recours à un technicien spécialiste en analyse quantitative et utilisation d'un logiciel d'analyse statistique.	N/A.
Synthèse des résultats et interprétation	• Utilisation d'un gabarit de rédaction d'étude de cas. • Synthèse des résultats et interprétation réalisées par l'éval. sénior chargé du projet d'éval. ext. à partir des matrices de preuves/ résultats dans le cadre de la phase de rédaction du rapport d'éval. final.	• Organisation des résultats dans des matrices de preuves/résultats afin de répondre aux questions d'évaluation. • Réaliser par l'éval. sénior chargé du projet d'éval. à l'externe.	• Réalisée par l'éval. sénior chargé du projet. • *"I don't know how it was done but the basic framework is you're trying to answer the questions of courses. But in terms of how to determine the right level of detail, I really don't know."*	• 1re ébauche réalisée par 'J1' (*lead junior*), et révisée par l'évaluateur sénior partenaire chargée du projet. • Synthèse des résultats dans une matrice de preuves conçue à partir de la combinaison de la matrice de résultats de chaque source de données afin de faire émerger des réponses aux questions d'éval. qui sont claires, précises et représentatives du programme (triangulation des preuves).

Enfin, les matrices de résultats (ou de preuves) semblent constituer une forme de rapport technique pour chaque méthode employée, rapports sur lesquels se basent la synthèse et l'interprétation des résultats et la formulation des conclusions évaluatives ('J1', 'J3' et 'J4'). 'J1' en tant que rédacteur de la version initiale du rapport d'évaluation finale, nous explique que la formulation des constatations, des réponses aux questions d'évaluation, repose sur l'utilisation, la combinaison/triangulation du contenu des matrices de résultats des différentes sources de données, un processus qui renforce la fiabilité, la validité et la représentativité des conclusions évaluatives.

Retour sur l'obtention de faibles taux de réponse aux enquêtes - les cas de 'J2' et 'J3'

La plus grande limite méthodologique encourue dans le cadre des évaluations auxquelles les évaluateurs 'J2' et 'J4' ont respectivement participé fut le faible taux de réponse aux différentes enquêtes, à savoir des taux compris entre 2,6% et 6,6%. Le Tableau 28 suivant rend compte de la perception et des interprétations des évaluateurs quant aux éléments ayant engendré un manque conséquent de données, et de représentativité des résultats :

Tableau 28. *Perceptions et interprétations sur l'obtention de faibles taux de réponse aux enquêtes (évaluateurs juniors 'J2' et 'J3')*

	Évaluateur junior 'J3'	Évaluateur junior 'J2'
Limites observées	• Enquête auprès d'informateurs clés externes - *taux de réponse de 5.8%.* • Échantillons des sous-groupes de répondants non homogènes (nombre trop restreint, voir absence, de répondants pour certains sous-groupes). o Impossibilité de faire une comparaison intergroupes. o Manque de représentativité des résultats - manque de triangulation avec l'expérience de la population ciblée prog.	• Enquête 1 auprès des industries - *taux de réponse de 6,6 %.* • Enquête 2 auprès des utilisateurs - *taux de réponse de 2,6 %.* • Échantillon dans le cadre de chaque enquête non aléatoire et non représentatif de la population ciblée par le programme.
Perceptions, hypothèses et/ou explications	• Contenus problématiques de la base de données contenant les listes de contacts : o Base de données trop vaste et désorganisée. o Listes des adresses électroniques des répondants ciblés non mises à jour. • Échantillonnage stratifié de la population ciblée au moment de la passation du questionnaire par le biais d'items ayant une fonction '*skip logic*' : o Questionnaire trop long pour certains sous-groupes de répondants (jusqu'à 81 questions, soit une passation de plus 15 minutes), et donc non adapté à la réalité du contexte évaluatif. o Persistance des clients dans l'ajout de questions au moment de la conception du questionnaire, et ce, malgré les recommandations de l'éval. en charge de l'éval. ext. de conserver un temps de passation court du fait des caractéristiques de certains sous-groupes de répondants.	• Enquête 1 - choix des clients de recourir uniquement à la base de données d'entreprises et d'individus ayant contacté leur département et non de l'ensemble de la population du domaine concerné à travers le Canada. • Enquête 2 - réticence des clients à partager la liste de diffusion d'annonces électroniques *(listserv)* avec éval. ext. o Préoccupation en matière de protection de la vie privée de leurs souscripteurs et de la perte de ces derniers s'ils estiment être contactés inutilement. o Liste des répondants ciblés extraite par les clients au sein de la base de données principale des parties prenantes clés. • Enquêtes 1 et 2 - manque de facilitation de la passation des questionnaires par les gestionnaires et administrateurs du programme : o Absence d'envoi d'un courriel notifiant la population ciblée de l'évaluation en cours. o Listes des adresses électroniques des répondants ciblés non mises à jour.

Retour sur le partage des guides d'entrevues avant la passation - les cas de 'J4' et 'J1'

Dans le cadre de leurs évaluations respectives, les évaluateurs juniors 'J1' et 'J4' ont mentionné que les guides d'entrevues étaient envoyés aux informateurs clés au moment de la prise de rendez-vous. Lorsqu'interrogés sur ce sujet, ils ont déclaré que c'était une chose tout à fait

ordinaire, et ça allait de soi. 'J4' nous a toutefois expliqué qu'il s'agissait d'une version épurée du guide, d'un « *guide externe* » qui ne contenait pas les liens entre les questions d'entrevues et les indicateurs des matrices, ou encore les objets de relances. S'il ne fut pas en mesure de nous expliquer une telle décision, pour lui : « *they can prepare and read the blurb about the purpose of the interview, etc., the section that talks about privacy* ». Dans le cas de 'J1', à titre de « *junior lead »* de collectes de données durant de précédentes évaluations, il savait que les guides d'entrevues étaient toujours partagés avant la passation, dans le cas présent, en tant que collaborateur principal de l'évaluateur sénior responsable de l'évaluation externe, il a une meilleure connaissance des raisons sous-jacentes d'une telle initiative - « *When we book the interview, we attach the interview questions so they can think about generally what we're going to ask. It makes them feel more comfortable. No one likes to be unsure about what they are being asked of their program. Right. It reduces the anxiety one can feel when asked to be interviewed for an evaluation. ».* Aussi, pour 'J1', cela ne semble pas représenter de potentiel biais, principalement parce qu'ils ont communiqué les grands axes de questionnement clé, les éléments qui les intéressent plus particulièrement, et que les sous-questions et relances n'étaient pas préétablies.

Retour sur la préparation « hâtive » des juniors exécutants - le cas de 'J1'

Nous avons vu un peu plus tôt que les entrevues semi-dirigées (ou autre collecte de données) sont réalisées par les évaluateurs juniors, très souvent dans une capacité d'exécutants invités à se joindre à l'équipe évaluative dans le cadre de la phase de collectes de données et (ou) d'analyse de données. Aussi, 'J1' revient sur son expérience de junior exécutant, et les multiples fois où il s'est retrouvé « *parachuté* » dans une réunion de préparation des collectes de données d'un projet d'évaluation pour lequel il ne faisait pas partie de l'équipe évaluative initiale. Cette réunion vise à se familiariser avec le projet, les éléments de réponses, les résultats obtenus jusque-là par le biais des autres sources de données, etc. Le gestionnaire de l'évaluation insiste sur les thèmes clés sur lesquels il est essentiel d'obtenir des réponses dans le cadre des entrevues, il parcourt le guide avec l'équipe d'interviewer. 'J1' explique que cela nécessite que les évaluateurs juniors soient capables d'assimiler et de maitrise beaucoup d'information, et ce, très rapidement, afin d'être en mesure de mener des entrevues de qualité. Pour lui, c'est une manière de renforcer le développement de compétences clés chez les évaluateurs les plus juniors qui sont indispensables dans un contexte de pratique professionnelle où tout se passe très vite. En revanche, il décrit sa première expérience en tant que collaborateur impliqué dans chacune des étapes de conception du

devis d'évaluation comme un moment charnière dans son développement professionnel. En effet, il explique avoir eu le sentiment de contribuer de manière significative au projet d'évaluation, d'être en pleine maitrise des informations, des éléments clés, du projet, plus en contrôle durant les collectes de données, etc. (notons que nous revenons davantage sur ce propos dans la section portant sur la synthèse des pratiques réelles des évaluateurs juniors).

"*Oftentimes junior consultants are just invited like okay we are doing key informant interview so here get up to speed! [...]. I've been pulled into meetings for evaluations I was not involved in. And at the last minute. Like there's a lot of interviews and a lot of topics to cover. So, I'm coming into the meeting to really get up to speed on the project, what are the themes that you're hearing, what's happening with the survey, what lines of evidence are happening, etc. It's a lot to process. Like you can imagine jumping into any project midway through. And that often happens as a junior staff and that's just the reality of it when you're trying to gain your skills and you're not leading projects. And here you are building capacity in different areas, so you must quickly get up to speed on things, you must be really good at doing it or else you won't make it. You need to be good at very quickly absorbing information like that because if I were going to do these interviews and if I couldn't absorb what's happening with this or I didn't have the attention span to do it, the interviews would be really bad. Everything is going at a very fast pace from the moment we get the contract to the final report delivery. [...]. But here, I was involved in the design, so it was easier, I was able to give feedback on the guide and understand what we were asking and had the opportunity to ask questions before we launched. That's important for our growth and confidence actually like it did for me. To feel I had a meaningful contribution to the project, to have more agency, to be more in control during data collections."* (J1).

e. Les procédés d'assurance de la fiabilité et de la validité

On retrouve dans le Tableau 29 (p. 255), les éléments de réponses fournis par les évaluateurs juniors lorsque nous les avons questionnés sur les procédures mises en place afin de veiller à fiabilité et la validité des résultats d'évaluation. Celles-ci reposent, entre autres, sur :

- La réalisation des tâches méthodologiques « critiques » par l'évaluateur sénior chargé de l'évaluation externe (conception du plan d'évaluation, de l'instrumentation, du protocole d'analyse des données, supervision des procédures, synthèse des résultats, rédaction des conclusions, etc.), avec l'appui du comité d'évaluation interne, plus précisément de la division de l'évaluation interne qui tient également le rôle d'entité « *validatrice* » des choix méthodologiques, et des contenus rapportés dans les différents rapports d'évaluation. – 'J3'.
- Les efforts de triangulation des sources de données, soit l'assurance que les informations rapportées comme preuves sont bel et bien représentatives du programme évalué. - 'J1', 'J2' et 'J4'. Par ailleurs pour 'J1' ce procédé est soutenu par la conception d'une matrice de preuves qui regroupe les matrices de résultats des différentes sources de données.
- La rigueur et la transparence dans la communication de la démarche méthodologique, la synthèse des résultats et la formulation des constats évaluatifs. Les outils clés sont ici le rapport

méthodologique contenant notamment la matrice évaluative et l'instrumentation, et le rapport technique de chaque source de données (matrice de résultats pour chaque méthode) – 'J1' et 'J4'. 'J1' aborde également ici l'importance de veiller à la réplicabilité du devis d'évaluation.

- L'accord interévaluateurs dans le cadre de la formulation des constats évaluatifs, l'évaluateur sénior chargé du projet d'évaluation à l'externe veillant à l'intégrité de l'utilisation et de l'interprétation des résultats. – 'J1'.

Tableau 29. *Explicitations ou perceptions/suppositions sur les procédures d'assurance de la fiabilité et de la scientificité (évaluateurs juniors)*

Contexte numéro 1		**Contexte numéro 2**	**Contexte numéro 3**
'J4'	**'J3'**	**'J2'**	**'J1'**
Juniors lead exécutants d'un nombre important d'activités de collecte et (ou) d'analyse de données.		**Juniors lead exécutants** d'un nombre limité d'activités de collecte et (ou) d'analyse de données.	**lead junior, principal collaborateur** de l'évaluateur sénior chargé du projet d'évaluation externe.
• Rigueur et transparence dans la description de la méthodologie - rapport méthodologique très détaillé, contient matrice éval. complète avec description des sources d'information, instrumentation, protocole de passation, plan d'analyse, etc. • Recours à la triangulation des sources et des méthodes afin d'assurer la représentativité des résultats. • Conception d'une matrice de résultats pour chaque méthode (inclus données, sources d'information, résultats d'analyse pour chaque question éval.) - rapports techniques à partir desquels l'éval. sénior chargé de l'éval. à l'ext. rédigent la synthèse des résultats/conclusions éval. dans le rapport final.	• Principal rôle de l'évaluateur sénior chargé de l'éval. à l'ext. épaulé par une collaboration étroite avec le comité d'évaluation interne. o L'éval. sénior chargé de l'éval. à l'ext. a pour responsabilité de définir le devis d'évaluation, de concevoir l'instrumentation, de superviser la collecte et l'analyse des données, de synthétiser les résultats et de formuler les constatations, et de rédiger les différents rapports d'éval. o Le comité d'évaluation interne, plus particulièrement la division d'évaluation interne guide, conseille et valide les prises de décisions méthodologiques (cahier des charges, matrice éval., échantillons, instrumentation, etc.), et la rédaction des différents rapports d'évaluation afin d'assurer la rigueur de l'évaluation et de sa communication.	• Recours à la triangulation des sources et des méthodes en vue notamment de gérer les potentiels biais ou représentations erronées des informations - une procédure de contrôle qu'il explique lui avoir été présenté comme fondamental pour ce type d'évaluation dans le présent contexte.	*Réplicabilité* • Effort de rigueur dans la description de la méthodologie évaluative, des procédures d'échantillonnage, dans le développement des indicateurs, des instruments, etc. (rapport méthodologique avec matrice éval., rapport technique pour chaque source de données, etc.). *Fiabilité et représentativité* • Conception matrice de résultats pour chaque méthode (inclus données, sources, résultats d'analyse pour chaque question éval.) pour s'assurer de la possession de l'ensemble des preuves nécessaires pour soutenir les constatations éval. • Conception matrice de preuves à partir des matrices de résultats (inclus données/résultats, sources, interprétation pour chaque question) afin d'assurer/démontrer l'effort de triangulation des preuves sur lesquelles reposent les constations éval. • Recours au système de matrices permet d'assurer le recours à une démarche systématique et rigoureuse, de montrer clairement le raisonnement éval., et de vérifier la provenance des preuves et la représentativité des résultats utilisés pour formuler les conclusions éval. • Recours à l'accord interévaluateurs dans l'interprétation des résultats et la formulation des conclusions - Éval. sénior partenaire chargé du projet d'éval. ext. doit s'assurer de l'intégrité de l'utilisation et de l'interprétation des résultats.

4.2.1.4 *Interprétation de la question de recherche tertiaire 2* - les pratiques réelles des évaluateurs novices de programmes de programmes publics

Question de recherche tertiaire (2) : Quelles sont les informations fournies par les évaluateurs, pour expliquer et décrire les choix méthodologiques opérés dans le cadre de leurs évaluations de programme?

L'activité d'explicitation a permis de mettre en exergue les pratiques « réelles » des évaluateurs novices participants à cette recherche doctorale. Des pratiques qui s'inscrivent dans le présent contexte : l'évaluation d'un programme public (gouvernement fédéral), commanditée par un comité d'évaluation interne, et réalisée par une équipe évaluative externe dirigée par un évaluateur sénior partenaire du cabinet-conseil contractant.

Remarquons que le « statut » de nos évaluateurs novices au sein de leur équipe évaluative respective et l'ampleur de leur contribution à l'évaluation différait d'un cas à l'autre. La plus grande différence sur le plan des pratiques réelles de nos évaluateurs novices repose ici sur le fait que :

- à titre de juniors exécutants, les évaluateurs 'J2', 'J3' et 'J4' travaillent en silos, ces derniers étant intégrés à l'équipe évaluative au moment de la collecte et (ou) de l'analyse des données ;

- à titre d'évaluateur junior collaborateur, 'J1' fut impliqué dans toutes les étapes de la démarche d'évaluation, de la rencontre de travail initiale avec les clients à la rédaction de l'ébauche du rapport final.

En revanche, qu'ils soient *junior exécutant* (J2), *juniors lead exécutants* ('J3' et 'J4'), ou *lead junior collaborateur principal* (J1), l'ensemble des décisions pratiques et méthodologiques reposent ultimement sur les épaules de l'évaluateur sénior chargé du projet d'évaluation (développement de la matrice évaluative et de l'instrumentation, planification des procédures de collecte et d'analyse de données, synthèse des résultats, communication de l'évaluation, etc.).

Si on se réfère à la taxonomie des compétences essentielles des évaluateurs de programme (Stevahn, King, Ghere et Minnema, 2005 ; King, Stevhan, Ghere et Minnema, 2001), cadre de référence des socles de compétences définies par les Sociétés canadienne et américaine en évaluation de programme (2018), quel que soit le profil des évaluateurs juniors au moment où ils débutent, leurs pratiques réelles, les premières années, sont ancrées principalement dans le *volet*

technique (« *systematic inquiry* »). Dans un premier temps, il s'agit de la réalisation de tâches de collecte et d'analyse de données ainsi que de synthèse, interprétation et communication écrite des résultats. Le tout à partir de protocoles, d'instruments, de gabarits rédactionnels préparés par l'évaluateur sénior chargé du projet d'évaluation à l'externe. Puis, dans un deuxième temps, lorsque les évaluateurs juniors commencent à tenir le rôle de collaborateur sur certains projets, il est question de la participation aux activités de conception et de mise en œuvre du devis d'évaluation. Remarquons que compte de tenu de la diversité sur le plan des profils des évaluateurs qui entrent sur le terrain (Hatry, Newcomer et Wholey, 2015; Stevahn, King, Ghere et Minnema, 2005), à l'instar des participants de cette thèse doctorale, il n'est pas surprenant que les premières années d'expérience professionnelle des évaluateurs priorisent, comme la formation des jeunes évaluateurs (Chouinard et *coll.*, 2017), la dimension technique de l'activité évaluative, soit l'acquisition, et le renforcement de connaissances et compétences méthodologiques essentielles pour réaliser des évaluations rigoureuses.

Comme on peut le voir dans les Encadrés 9, 10 et 11 (p. 267-269) qui résument les caractéristiques des pratiques décrites par les évaluateurs juniors exécutants, on observe une évolution dans les rôles, tâches et responsabilités octroyés aux évaluateurs juniors selon leur nombre d'années d'expérience professionnelle au sein de leur firme respective (où chacun débute sa carrière professionnelle). En effet, 'J2', pour qui il s'agissait de l'une de ses premières expériences professionnelles en évaluation de programme, a principalement travaillé sur le profil du programme, soit le recueil et l'examen d'informations, de données de base sur son historique, son bien-fondé, les activités mises en œuvre, etc. (codage et synthèse des informations). Alors que les évaluateurs 'J3' et 'J4' qui avaient respectivement (au moins) trois ans et deux ans d'expérience au moment de leur projet d'évaluation ont participé à l'ensemble des collectes de données qualitatives ; et ils furent d'ailleurs responsables de la conduite de la recension des écrits (J3) et de l'analyse documentaire (J4), ce qui implique notamment la gestion des sources documentaires, du processus de codage des informations, du contenu de la matrice de données ainsi que la supervision des évaluateurs subalternes (qu'ils qualifient parfois d'assistants de recherche). Ils ont participé à certaines activités d'analyse de données qualitatives ainsi qu'à la rédaction de rapports techniques (matrices de réponses aux questions d'évaluation sur lesquels se basent la rédaction de la synthèse des résultats, et la formulation des constatations évaluatives). 'J3' et 'J4' expliquent que si c'est l'évaluateur sénior chargé du projet d'évaluation à l'externe qui rédige le rapport d'évaluation

final, il est arrivé qu'ils soient invités à rédiger certaines sections de la partie sur la présentation des résultats directement dans l'ébauche du rapport final. Ainsi, au gré de leur expérience, les évaluateurs juniors exécutants sont invités à contribuer de manière de plus en plus significative à la collecte des informations, puis à leur analyse, et enfin à la synthèse et l'interprétation des résultats ainsi qu'à leur communication écrite. Notons que si 'J2 ne fait aucune mention explicite des rapports techniques, il explique que dans le cadre de l'écriture du rapport final, l'évaluateur sénior partenaire a intégré et édité de manière collaborative les éléments clés du profil du programme et de ses activités, rédigés par l'évaluateur junior. Il mentionne avoir également brièvement travaillé sur l'ébauche de résumé des résultats d'enquête.

Dans le cadre des premières années d'expérience des évaluateurs novices, on semble donc cibler l'acquisition et le renforcement de savoirs techniques opérationnels, soit la maitrise de l'utilisation pratique des différentes méthodes, procédures, et instruments de collecte et d'analyse de données, et de synthèse/interprétation/communication des résultats (Le Boterf, 2002). Selon nous, il est également question du développement du volet *compétence interpersonnelle*, plus particulièrement des compétences relatives à la communication écrite, la communication verbale/l'écoute ainsi que le travail en équipe (SCÉ, 2018 ; Stevahn, King, Ghere et Minnema, 2005).

Par ailleurs, on observe qu'un accent tout particulier est mis sur le développement de la capacité analytique des évaluateurs subalternes, une compétence transversale qui est essentielle aux volets i) *analyse situationnelle* (description et évaluation de l'évaluabilité d'un programme), ii) *technique/opérationnel* (conception du devis d'évaluation, collecte/analyse de données, interprétation des résultats, mobilisation du jugement évaluatif, etc.) de la pratique des évaluateurs de programme. En outre, l'évaluateur 'J1' explique qu'à travers ces expériences d'exécutant de tâches de collecte/analyse de données sur une multitude de projets d'évaluation (souvent concomitants), il s'agit pour les évaluateurs juniors de développer des compétences qu'il qualifie de cruciales, notamment la capacité d'absorber et d'assimiler rapidement une quantité importante d'information (programme, objets d'évaluation, enjeux critiques, etc.), et de maitriser l'instrumentation, afin d'être en mesure de collecter des informations utiles et de qualité.

Extrait de verbatim

"Oftentimes, especially the first couple years, junior staff are just pulled in last minute to conduct literature, document review, collect information. So, we are just pulled in for interviews, like it happened to me today actually, I was pulled in a project that's been going on for like six months where they are starting interviews, and I'm not sure what's going on, between me and you. I have to read the scope, read the RFP, the evaluation, and data matrices. It takes a long time to get up to speed, to jump in on that project last minute and be doing interviews in a couple of days. You are coming into the meeting to really get up to speed on the project, what are the themes that you're hearing, what's happening with the survey, what lines of evidence are happening. Who are we speaking to... it's a lot to process... you can imagine jumping into any project midway through. That often happens as a junior staff and that's just the reality of it, when you're trying to gain your skills because you're not leading projects, here you are building capacity in different areas like learning to get up to speed on things, absorb tones of information very quickly and understands elements that are essential to be able to have a meaningful conversation during interviews, critical thinking to be on top of things and be able to rebound on things and dig deeper. You have to be really good at doing it or else you are not going to make it. In the context we practice in, with all the different projects going on at once, the tight timelines, etc., you need to be good at very quickly absorbing, analyzing, understanding the program, the evaluation object, information's collected so far, because if you were going to do these interviews and weren't able to absorb what's happening with the program, the evaluation, or didn't have the attention span to do it, those interviews would be really bad. So, you also need to manage your time efficiently to be able to set aside time to do that, to prepare. It's really about gaining essential skills, technical skills but also analytical." (J1).

En dépit du fait que les évaluateurs juniors exécutants arrivent une fois que la méthodologie évaluative est conçue, ces derniers semblent avoir une bonne connaissance globale du projet et des spécificités de l'environnement évaluatif (complexité du programme, contraintes de terrain, etc.). En revanche, en ce qui a trait au processus décisionnel, à la démarche entreprise, dans le cadre de la conception du devis d'évaluation, il n'est pas toujours facile de dissocier, particulièrement chez les participants qui ont le plus d'expérience ('J3' et 'J4'), si les informations transmises sont fondées sur une connaissance concrète de ce qui s'est passé ou si elles sont basées sur des projets d'évaluation antérieurs, sur ce qu'ils perçoivent comme des procédés systématiques ou encore des pratiques « prescrites » par le contexte évaluatif. Au demeurant, au cœur des propos tenus par ces évaluateurs se trouve principalement le facteur politique, ce qu'ils qualifient de pratique standard de l'évaluation dans le contexte fédéral et leur connaissance/perception/compréhension de la gouvernance de projets d'évaluation, du pouvoir décisionnel de l'équipe évaluative externe dans ledit contexte. Comme on peut le voir dans les Encadrés 9, 10 et 11 (p. 267-269), selon l'expérience des évaluateurs juniors exécutants, on observe une augmentation de la précision et de la justesse sur le plan de leurs connaissances en ce qui a trait à la démarche de conception du devis d'évaluation, aux spécificités des éléments contextuels d'ordre politique et organisationnel, et l'influence que cela exerce sur les choix méthodologiques opérés par l'équipe évaluative externe. L'un des éléments saillants repose sur la perception, et le niveau de connaissance et de compréhension de l'ampleur des responsabilités de l'équipe évaluative externe et du processus

décisionnel engagé entre cette dernière et le comité d'évaluation interne (client) sur le plan de la planification de la méthodologie évaluative. Ainsi :

'J2' (12 - 16 mois d'expérience) a une assez bonne connaissance globale du cahier des charges et comprend que le devis d'évaluation est conçu selon un modèle standard du gouvernement fédéral dont il a une connaissance superficielle. Il n'a pas connaissance ni de la démarche de conception du plan d'évaluation ni de la dynamique relationnelle/décisionnelle entre l'équipe évaluative externe et parties prenantes clés commanditaires de l'évaluation. Il ne fait aucunement mention de la division interne de l'évaluation, malgré le fait que le rapport est publié en ligne sous son nom. S'il réitère à plusieurs reprises n'avoir aucune connaissance des choix méthodologiques opérés au moment de la conception du devis d'évaluation (du fait de son statut d'exécutant), il est tout de même en mesure de discuter de certaines décisions telles que le recours à l'enquête.

'J4' (2 ans - 2 ans ½ d'expérience) démontre une bonne connaissance globale du cahier des charges et de son ancrage politique (politique évaluative fédérale, plan d'évaluation quinquennale des programmes de l'agence commanditaire). Pour lui, l'équipe évaluative externe est exécutante du plan d'évaluation prescrit par le comité d'évaluation interne composé des parties prenantes clés et la division interne de l'évaluation, dans la demande de propositions (prescription des objets/questions d'évaluation à traiter, méthodes à utiliser, sources d'information à consulter). Il fait référence à un climat d'obligation, ce qui semble biaiser sa perception du poids du rôle/des responsabilités de l'équipe évaluative externe sur le plan de la planification du devis méthodologique, occultant ainsi tout le travail entrepris par cette dernière sur le plan conception du protocole d'évaluation (matrice d'évaluation). Il fait preuve d'un esprit critique face à l'utilisation mécanique de l'évaluation de programme dans le contexte du gouvernement fédéral. S'il semble percevoir le rôle d'agent de liaison à l'interne de la division interne de l'évaluation, il ne fait pas d'autre distinction entre cette dernière et les autres parties prenantes clés de l'évaluation qui sont également membres du comité d'évaluation interne (hauts gestionnaires de programme).

'J3' (4 ans – 4 ans ½ d'expérience) fait preuve d'une connaissance très précise du cadre d'évaluation et des spécificités de son ancrage politique. S'il reconnait l'importance de planifier le devis méthodologique au sein des paramètres prescrits par les clients dans la DDP (politique évaluative fédérale, plan d'évaluation quinquennal de l'agence commanditaire) et les directives de

la politique évaluative (cadre d'évaluation de référence), pour lui l'équipe évaluative externe est responsable de la définition du contenu du cahier des charges. Il explique que le caractère générique des éléments prescrits nécessite comme pour tout projet d'évaluation l'entreprise d'une phase de description et d'évaluation de l'évaluabilité du programme dont l'extrant est la matrice évaluative produite par l'équipe évaluative externe. Enfin, il a une assez bonne compréhension de la dynamique relationnelle entre l'équipe évaluative externe et le comité d'évaluation interne qu'il perçoit comme collaborative. Il souligne le caractère essentiel d'une collaboration étroite avec la division interne de l'évaluation pour une planification, conduite et communication de qualité de l'évaluation (aiguillage méthodologique, facilitation de la sélection et de l'accès aux sources d'information, liaison avec les parties prenantes clés, etc.). S'il perçoit le comité d'évaluation interne comme un groupe de contrôle qualité, veillant notamment à la représentativité des constatations évaluatives, il demeure conscient de la nécessité pour l'équipe évaluative externe d'être vigilant face à la potentielle manipulation sur le plan de la communication des résultats et constats évaluatifs. Il n'est pas toujours clair qui, entre les parties prenantes clés du comité d'évaluation interne et la division d'évaluation interne, détient un pouvoir décisionnel versus un pouvoir davantage consultatif.

À mesure de l'enrichissement de l'expérience professionnelle des juniors exécutants, on observe une amélioration des connaissances générales propres à l'évaluation, et spécifiques à l'environnement professionnel et l'évaluation en situation réelle (éléments du bagage de savoirs théoriques/procéduraux [Le Boterf, 2002]), mais celles-ci demeurent lacunaires. Le travail en silo maintenant un angle mort qui les empêche de saisir pleinement la mesure des responsabilités méthodologiques de l'équipe évaluative externe sur le plan de la conception du plan d'évaluation, et les particularités de la dynamique entre cette dernière et les différents membres du comité d'évaluation interne. Il s'agit ici des apprentissages relatifs à la dimension socioculturelle et politique de l'activité d'évaluation (Chouinard et *coll.*, 2017). L'expérience et les descriptions fournies par 'J3' et 'J1' nous montre que l'acquisition de telles connaissances repose sur l'opportunité pour les évaluateurs subalternes de tenir au sein de l'équipe évaluative consultante le rôle de junior collaborateur, soit de participer à l'ensemble des étapes du projet d'évaluation, de la rencontre de travail initiale avec les clients à la soumission de la première ébauche du rapport d'évaluation.

‘*J1’ (22 - 28 mois d’expérience)* en tant que junior collaborateur a participé à l’ensemble des réunions de travail, des discussions dans le cadre de la conception du plan d’évaluation (incluant les interactions avec les clients), la révision du contenu de l’instrumentation, la coordination et la réalisation des activités de collecte et d’analyse de données (incluant la préparation et la supervision des évaluateurs exécutants, la collaboration étroite avec l’équipe évaluative interne), la synthèse des résultats et à la rédaction du rapport final (responsable de la première ébauche, incluant les outils d’implantation, de monitorage de l’approche évaluée - principaux extrants du projet d’évaluation). La chance d’être le collaborateur de l’évaluateur sénior chargé du projet d’évaluation à l’externe représente une expérience d’apprentissage très riche ainsi qu’un moment charnière dans le développement professionnel du jeune évaluateur. Outre la poursuite du renforcement des capacités abordées jusque-là, il s’agit ici de l’acquisition d’un bagage de savoirs procéduraux, opérationnels et expérientiels relativement aux tâches de conception du devis d’évaluation. Il s’agit ici de savoirs liés aux volets *analyse situationnelle* et *enquête systématique* du socle des compétences essentielles de l’évaluateur de programme (SCÉ, 2018 ; Stevahn, King, Ghere et Minnema, 2005). Ils portent sur les tâches de description du programme, de l’analyse des besoins des utilisateurs ciblés, du contexte politique et environnemental, et de l’évaluabilité du programme, du cadrage de l’évaluation et de la conception du cadre méthodologique (définition et sélection des sources d’information, conception de l’instrumentation, préparation des activités de passations).

‘J1’ offre une analyse de pratique intéressante à l’appui de la comparaison de son expérience de junior exécutant à celle de junior collaborateur dont les caractéristiques sont décrites dans l’Encadré 12 (p. 270). Il explique que la pratique des évaluateurs subalternes est rythmée par un travail en silo, l’évaluateur junior exécutant parachuté d’un projet à l’autre (ou sur plusieurs projets à la fois) au moment de la phase de recueil des informations (souvent à la dernière minute), leur point d’entrée dans le projet étant généralement la rencontre de préparation de la collecte des données. Dans le cadre de ladite rencontre, le gestionnaire de l’équipe évaluative externe (un évaluateur sénior, généralement partenaire du cabinet-conseil) réalise une synthèse des informations sur le devis ainsi que les données/résultats d’évaluation obtenus jusque-là, et parcourt le contenu de l’instrumentation, les points clés sur lesquels il est indispensable de recueillir des informations, de faire preuve de profondeur. Dans le cadre de cette préparation,

les évaluateurs exécutants s'appuient également sur différentes sources documentaires telles que la demande de propositions, le rapport méthodologique (contenant le mandat, matrice évaluative, la description des sources d'information, l'instrumentation, etc.), et les matrices de données. Notons que cela expliquerait, entre autres, pourquoi les évaluateurs novices dans le cadre de leurs pratiques recommandées s'attardent tout particulièrement sur l'effort d'explicitation de la méthodologie évaluative, sur l'importance de la rigueur et de la transparence dans la description du devis d'évaluation (matrice évaluative, échantillonnage, instrumentation, protocole d'analyse de données/synthèse des résultats, etc.).

Comme nous l'avons vu un peu plus haut, 'J1' reconnait la pertinence et le caractère formateur de l'expérience professionnelle en tant qu'exécutant, celle-ci permettant le développement de compétences pratiques et méthodologiques qu'il juge essentielles, dans un contexte de travail très dynamique et contraignant. Pour lui, il est particulièrement question du renforcement de la capacité de traitement et de rétention de l'information dans un laps de temps très court visant la maitrise du contenu du programme, de l'évaluation ainsi que de l'instrumentation afin d'être en mesure de collecter des informations utiles et justes. Fort de son expérience de junior collaborateur où il a eu la chance de contribuer à la conception du devis (rencontre avec les clients, réunions d'équipe, révision du contenu de l'instrumentation, préparation et coordination de la collecte/analyse des données), il remarque s'être senti plus en contrôle lors de la collecte de l'information, plus à même de faire preuve d'efficacité et de profondeur dans le cadre de la passation des entrevues, etc. De même, pouvoir contribuer aux propositions méthodologiques et communiquer directement avec les clients, plus précisément la division interne de l'évaluation, génère chez ce dernier le sentiment d'être responsabilisé, et d'être maitre des tâches méthodologiques qu'il entreprend. Il cite l'exemple de la tâche de conception du plan d'implantation de l'approche évaluée (extrant de l'évaluation) dont il était responsable. Où il a pu non seulement suggérer l'utilisation de différents gabarits à l'évaluateur sénior gestionnaire du projet, mais aussi solliciter directement les clients pour obtenir des conseils/rétroactions afin de s'assurer de la pertinence et de l'utilité du format et du contenu de l'outil qu'il proposait.

Extrait de verbatim

"I was always around the table so that was one great thing that I really liked and that helped my learning a lot as well, to see like the start-up meeting with the clients that they always have and then the various internal meetings discussing about the design. I was invited to each one. I felt that I had a voice around the table about how to approach the project, the different methods we were planning on using, I revised the instruments, was able to suggest templates for the implementation plan we were producing as a result of the evaluation. [...]. I was able to contribute much more meaningfully, more like an actual evaluator. One of the partners here was the project manager and the main liaison with the clients, the evaluation committee, and internally within our own team, I was able to kind of take control. I managed the data collection and analysis, supervised the more junior staff. I worked on the evidence matrix with the senior partner, he let me take a first stab at writing the report. I had the opportunity to develop the template and develop this implementation tool. [...]. When it came to collect the data, I knew what was happening, I was involved in the start-up meeting with clients, I revised the interview guides, was able to ask questions, so I really understood what we were asking. I was involved in the training with the instruments. It was easier, better compared to when I was asked to jump the last minute, midway through a project like it usually is for junior staff. I felt more equipped, confident that I would be able to go as deep as needed during the interviews that I would not miss anything important. That's important for our growth and confidence actually. [...]. Because I had met with the clients early on when it came time to develop the toolbox, I had a good grasp on what they needed to make meaningful suggestions and I had their contacts, so I was able to interact directly with them to get more info and like their feedback. And I just think a lot of the times you get thrown midway through a project where they say they need you to do this literature and doc review and you didn't meet with the client. You don't really know the contacts or even have access to them so you're just kind of digging around for whatever you have, it feels like it's out of your control and you have to do the best you can with what you've got. You don't feel like you are as useful. It's essential to feel you have a meaningful contribution, to have more agency, to feel more in control in the different task you are involved in. [...]. Everything I did in a lead capacity I was, of course, going back and force with the senior partner to get confirmation you know because it wasn't necessarily my project. It was a collaborative effort but also opportunities to take the lead as well. Anything that goes out, here the partner or the project manager would do a review and make suggestions, so there's always an opportunity at the end to go through and say you know this is great, but I would suggest that, or I know this client, and this is their preference so let's do it this way. So, it's kind of great because you're not being micromanaged but there's a lot of back and forth with feedback. [...]. I think there needs to be more opportunities like that for young evaluators, and sooner rather than later. To be part of the project from the beginning, to have the opportunity to take the lead, the opportunity for growth, the right way. To ensure that we are equipped, and well prepared to lead evaluation project." (J1).

Cette transition d'exécutant à collaborateur, d'une pratique en silo à une pratique contextualisée et collaborative, où l'évaluateur junior se sent responsabilisé, tout en étant encadré dans sa pratique par un évaluateur expert, constitue un tournant majeur dans le développement professionnel de l'évaluateur. Plus particulièrement la prise de conscience de soi en tant qu'évaluateur de programme, un élément du volet de compétence *pratique réflexive* (SCÉ, 2018 ; Stevahn, King, Ghere et Minnema, 2005). En effet, l'évaluateur junior collaborateur fait part d'une augmentation de la confiance en ses capacités d'évaluateur, tout en faisant état des besoins des évaluateurs subalternes en termes d'occasions d'apprentissage adéquates afin qu'ils soient préparés au mieux à tenir, à leur tour, le rôle de gestionnaire. Pour 'J1', il s'agit d'offrir aux évaluateurs novices « *opportunities for growth, the right way* », à savoir de multiplier les chances pour ces derniers de participer à des projets d'évaluation du début à la fin, et ce, le plus tôt possible

(particulièrement dans le cas de petits projets d'évaluation avec un devis méthodologique assez simple, comme celui auquel il a pris part). Un développement professionnel selon nous ancré dans la réalité sur le terrain, la pratique de l'évaluation en situation réelle (Bamberger et Mabry, 2020); une pratique dans le monde immédiat de la politique, des clients, des ressources, de l'ambiguïté des rôles et de l'évolution des conditions sur le terrain, et l'intérêt pratique de l'accomplissement d'un travail bien fait (Smith et Brandon, 2008 ; Association Américaine en Évaluation, 2004). Enfin, malgré le fait que l'évaluateur junior collaborateur fait preuve d'une grande maitrise des facteurs environnementaux et des enjeux méthodologiques associés, qu'il est en mesure de décrire avec précision la démarche et le processus décisionnel engagé à chaque étape de la démarche évaluative (incluant la communication de l'évaluation), et qu'il semble bien cerner la dynamique entre l'équipe évaluative externe et les clients, certaines zones d'ombres perdurent (Encadré 12, p. 270). En effet, s'il participe à la rencontre de travail initial avec les clients ainsi qu'aux différentes réunions d'équipe dans le cadre de la conception du devis, et qu'il se voit octroyé certaines responsabilités, à l'interne (c.-à-d. révision de l'instrumentation, coordination collecte/traitement des données, rédaction de la première ébauche du rapport final), il y a tout un pan d'activités entreprises par l'évaluateur sénior responsable du projet d'évaluation à l'externe qui semble demeurer inconnu et (ou) méconnu de l'évaluateur junior collaborateur (particulièrement en ce qui a trait au volet *gestion de projet*). Il s'agit notamment : du traitement de la demande de propositions, de la budgétisation du devis d'évaluation et planification de la répartition des ressources, de la construction à proprement parler du devis d'évaluation (conception/finalisation de la matrice évaluative, du cahier des charges) et de l'instrumentation, du processus de consultation/validation du devis d'évaluation auprès des commanditaires ainsi que du rapport final. Aussi, il est question ici de nombreux temps d'interaction entre l'évaluateur sénior gestionnaire du projet à l'externe et les clients, soit le comité d'évaluation interne, auquel l'évaluateur junior collaborateur n'assiste pas nécessairement. Par exemple, des temps de consultation avec la division interne de l'évaluation afin d'assurer la pertinence et la faisabilité du devis d'évaluation, des temps de rétroaction des gestionnaires de programme dans le cadre de la finalisation du devis d'évaluation, des temps de validation du contenu du devis d'évaluation et du rapport final d'évaluation auprès des clients. En revanche, si l'expérience de collaborateur amène l'évaluateur junior à percevoir le caractère essentiel de la collaboration étroite avec le comité d'évaluation interne, et la richesse de la contribution des utilisateurs ciblés qui le constitue, à savoir la division

interne de l'évaluation et les gestionnaires de programme, ce dernier ne semble pas être en mesure de percevoir clairement qui sont les détenteurs d'un pouvoir décisionnel, et qui sont les détenteurs d'un pouvoir davantage consultatif à l'interne. 'J1' reconnait faire l'amalgame entre les clients, c'est-à-dire les gestionnaires de programme membre du comité d'évaluation interne, et la division interne de l'évaluation. En effet, il semble tantôt percevoir la division interne de l'évaluation comme le client, tantôt comme une ressource méthodologique interne, un agent de liaison, un conseiller auprès des clients. Des clients, dont les attentes et les préférences jouent un rôle déterminant dans la méthodologie évaluative, et la façon dont l'évaluation et ses résultats sont communiqués dans le rapport d'évaluation. 'J3', qui a davantage d'expérience dans le présent contexte, fait preuve davantage de précision et de clarté sur ce point. Il reconnait le comité d'évaluation comme un groupe d'assurance qualité, qui examinent les produits livrables (rapport méthodologique, rapports techniques et rapport final), et s'assure que les résultats communiqués sont représentatifs du programme évalué. Il nous explique que le point de contact clé de ce comité est la division interne de l'évaluation, le collaborateur essentiel de l'équipe évaluative externe qui aiguille et facilite le processus de consultation interne, l'accès aux sources d'information, etc. En revanche, 'J3' semble reconnaitre la division interne de l'évaluation comme l'instance directrice, l'autorité décisionnelle, mais ce n'est pas toujours très clair (perpétuelle ambiguïté sur le plan de l'utilisation du terme *client*).

Encadré 9. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J2'*

'J2' **Junior exécutant**	A travaillé principalement sur le profil du programme, son historique, ses activités, etc. - Examen de la documentation, des données de programme (codage, analyse-synthèse des informations), rédaction d'éléments sur le profil du programme en collaboration avec l'évaluateur sénior chargé du projet à l'externe (à des fins d'intégration dans le rapport d'évaluation final). - Recherche d'écrits scientifiques. - Ébauche de résumé des résultats d'enquête.
Firme [A] Formation [PhD SHS] Expérience pro. ÉP [-1 an/Act.1] [16 mois/Act.2]	

- Il a une assez bonne connaissance globale du cadre d'évaluation.
- Il perçoit le processus décisionnel engagé dans la conception du plan d'évaluation comme par défaut – application du modèle évaluatif fédéral standard. Il fait allusion aux directives de l'IGF, mais il ne semble pas avoir connaissance du contenu/de la démarche d'application de ce cadre de référence, puisque celui-ci l'omet totalement lorsqu'il émet des hypothèses quant à la conception des questions d'évaluation. Par ailleurs, il ne fait aucune mention de la demande de propositions (DDP).
- Il n'est pas en mesure de décrire, ni la démarche entreprise pour définir les éléments clés du plan d'évaluation, ni la relation et la dynamique de pouvoir (décisionnel) entre le comité d'évaluation interne et l'équipe évaluative externe. D'ailleurs, lorsqu'il aborde les décisions prises, il parle toujours de « ils », mais nous ne sommes pas en mesure d'identifier qui prend les décisions.
- La discussion autour des enjeux méthodologiques face à la complexité du programme, et la nécessité de tenir compte des contraintes de temps, d'argent, et d'accès aux sources d'information, nous montre que s'il réitère à plusieurs reprises ne pas être en mesure de décrire les choix, que le choix des méthodes est standard, etc., il semble avoir plus de connaissances concernant le raisonnement méthodologique qu'il ne laisse initialement paraître, il ne fait juste pas de lien - raisonnement très séquentiel.
  - Impossibilité de recourir devis expérimental ou quasi-expérimental ;
  - Recours à une approche multiméthodes pour accéder de différentes manières à la même information ;
  - Recours à l'enquête (permet d'interroger un large éventail d'individu à un moindre coût, de palier le non-recours à un devis expérimental/quasi-expérimental) ;
  - Miser sur l'examen de la documentation, plus précisément le recueil de données secondaires, de données de programme pour l'évaluation de l'efficacité (gestion de certaines limites méthodologiques - non-recours à un devis expérimental/quasi-expérimental, faible taux de réponse à l'enquête).
- Il décrit de manière assez claire et précise les méthodes de collectes de données auxquelles il a pris part, et semble maitriser l'utilisation de l'instrumentation de collecte, et de codage de données qualitatives qu'il a utilisé.
- Ses connaissances et habiletés en recherche expérimentale semblent lui avoir permis d'émettre des suppositions pertinentes quant à certaines limites méthodologiques (non-recours à un devis expérimental/quasi-expérimental, faible taux de réponse à l'enquête, etc.).
- Pour lui, l'assurance de la fiabilité de l'évaluation est inscrite dans la notion fondamentale véhiculée selon laquelle pour ce type d'évaluation dans le contexte du gouvernement fédéral, la triangulation des sources et des méthodes permet de contrôler les potentiels biais ou représentations erronées des informations, des données.
  - Notons que son expérience (enrichie entre les deux passations) fait, qu'il a une meilleure connaissance (certes générale) de la culture évaluative du milieu fédéral, des enjeux politiques, etc. - prise de recul vis-à-vis de certains de ses « idéaux », tels que la notion que le manque de description de la méthodologie signifie obligatoirement un manque de rigueur.
  - Acquisition de connaissances initiales en évaluation de programme inscrite exclusivement dans le milieu du gouvernement fédéral.

Encadré 10. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J4'*

'J4' **Junior lead exécutant**	A participé à l'ensemble des collectes de données qualitatives, à l'analyse de données qualitatives ainsi qu'à la rédaction de rapports techniques (matrices de résultats/réponses aux questions d'évaluation). - Conduite de l'analyse documentaire – gestion et organisation de la documentation, codage des informations (incluant la supervision des assistants juniors). - Passation d'entrevues. - Analyse de données qualitatives (analyse documentaire, entrevues). - Rédaction de rapports techniques – matrice de résultats de l'analyse documentaire et des entrevues d'informateurs clés, rapport de cas à partir d'un gabarit.
Firme [B] Formation [M.P.P.A] [DES É.P] Expérience pro. ÉP [2 ans - 2 ans 1/2]	

- Il a une bonne connaissance globale du cadre d'évaluation et de son ancrage politique. En revanche, il fait preuve d'une certaine nonchalance face à l'évaluation de programmes fédéraux, et ce, du fait qu'il perçoit l'équipe évaluative externe comme exécutante du comité d'évaluation interne (parties prenantes clés et division interne de l'évaluation), plus particulièrement des paramètres de l'évaluation qu'ils prescrivent dans la DDP (objets/questions d'évaluation, méthodes, sources d'information) ; avec le contenu de la DDP lui-même déterminé par la politique évaluative fédérale et le plan d'évaluation quinquennale de l'agence commanditaire.
  - Il fait preuve de critique face à la forte influence politique sur l'activité d'évaluation, et ses conséquences – évaluation mécanique pour « cocher une case » (fournir une preuve d'activité évaluative), recours à une démarche méthodologique systématique prescrite et non innovante (climat d'obligation/d'exécution).
- La perception d'obligation semble occulter chez l'évaluateur junior tout le travail de conception du plan méthodologique par l'équipe évaluative externe, plus précisément de la matrice évaluative avec la définition des indicateurs, l'identification des sources d'information disponibles/indispensables, l'organisation des méthodes, et ce, pour chaque question d'évaluation.
  - Pourtant, lorsqu'on aborde le recours à de multiples sources de données, en dépit du choix prescrit des méthodes, il explique que l'une des pratiques standards de sa firme est de miser sur la complémentarité des méthodes qualitatives et quantitatives pour répondre à chaque question d'évaluation. Il perçoit la valeur de cette approche multiméthodes pour évaluer des programmes complexes, de grandes envergures, où les données quantitatives ne peuvent pas donner à elles seules une image représentative de l'objet d'évaluation (on a les contours de l'image, mais on n'a pas les détails des couleurs, des nuances, etc.). Pour lui, le devis cible d'abord la collecte de données quantitatives, qu'il vise ensuite à enrichir avec des données qualitatives qui racontent la « vraie » histoire.
  - D'autre part, il déclare qu'une matrice évaluative a bien entendu été conçue, avec les indicateurs et les méthodes, mais il ne donne pas davantage de précision. Aussi, il semble davantage reconnaitre son rôle essentiel en tant qu'outil de conception de l'instrumentation, de collecte et d'analyse de données, et de synthèse des résultats - là où pour lui, l'équipe évaluative externe devient maître de l'évaluation.
  - Notons qu'il mentionne initialement la réutilisation du devis de la précédente évaluation du programme, mais il ne revient pas sur cet élément durant la discussion sur les éléments clés du devis d'évaluation - hormis dans le choix des cas à étudier, et la réutilisation de la recension des écrits. Il ne fait pas référence à la matrice évaluative, ou encore à l'instrumentation (raisonnement séquentiel).
- Il décrit de manière claire et précise les méthodes de collecte de données auxquelles il a pris part. Il a une bonne connaissance globale des procédures de conception et des modalités de passation des instruments de collectes de données qualitatives, incluant les raisons quant à certains choix méthodologiques (p.ex. partage du guide d'entrevue avec l'interviewé en amont, non-enregistrement des entrevues, utilisation d'un chiffrier électronique pour la retranscription et l'analyse des informations, etc.). Il détient des connaissances tacites et savoirs opérationnels sur le plan de l'utilisation de l'instrumentation, et des procédures d'analyses de données qualitatives et de synthèse/interprétation des résultats.
- Du fait de son expérience de participation à la conception de matrices de réponses aux questions d'évaluation, plus particulièrement dans la rédaction de synthèse de cas, il fait état de préoccupations quant à la représentativité des résultats, la possibilité d'émettre des constats évaluatifs généraux face à l'envergure, la complexité et la diversité des modalités d'implantation de programmes pancanadiens ayant une gouvernance à la fois au niveau national et provincial.
- Pour la fiabilité et la validité, repose sur la transparence et la rigueur dans la description de la méthodologie évaluative (rapport méthodologique, rapports techniques), le recours au système de matrice (matrice évaluative, matrices de données, matrices de résultats) et la triangulation des sources de données.

Encadré 11. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J3'*

'J3' **Junior lead exécutant** *(avec quelques expériences en tant junior collaborateur dans d'autres projets)* Firme [A] Formation [M.Écon.] Expérience pro. ÉP [4 ans — 4 ans1/2]	A participé à l'ensemble des collectes de données qualitatives, à l'analyse de données qualitatives ainsi qu'à la rédaction de rapports techniques (matrices de résultats/réponses aux questions d'évaluation). - Conduite de la recension des écrits — gestion et mise en œuvre du processus de recherche/sélection/codage des écrits (incluant la supervision des assistants de rech.). - Conduite de l'analyse de données financières. - Collecte d'informations dans le cadre de l'analyse documentaire. - Passation d'entrevues. - Analyse de données qualitatives (recension des écrits, analyse documentaire). - Rédaction de rapports techniques (matrice de réponse aux questions d'évaluation) — recension des écrits, entrevues d'informateurs clés, rapports financiers.

- Il a une connaissance très précise du cadre d'évaluation, et des spécificités de son ancrage politique. Il a une assez bonne compréhension de la dynamique relationnelle entre le comité d'évaluation interne et l'évaluateur sénior chargé de l'évaluation à l'externe, et de ce que cela implique dans la pratique et les décisions opérées dans le cadre de l'évaluation.
  - Il perçoit le comité d'évaluation interne comme un groupe d'assurance qualité, qui examinent les produits livrables (rapport méthodologique, rapports techniques et rapport final), s'assure que les résultats communiqués sont représentatifs du programme évalué. Le point de contact clé du CMI est la division interne de l'évaluation, collaborateur essentiel de l'équipe évaluative externe qui aiguille et facilite le processus de consultation interne, l'accès aux sources d'information, etc.
  - Il n'est en revanche pas toujours clair concernant les détenteurs d'un pouvoir décisionnel versus davantage consultatif, plus particulièrement entre les clients et la division interne de l'évaluation.
- Il ne perçoit pas l'équipe évaluative externe comme exécutante de l'évaluation planifiée par les clients. Pour lui, si le devis d'évaluation doit être en effet arrimé aux paramètres d'évaluation définis dans la demande de propositions ainsi qu'aux directives de l'IGF (cadre d'évaluation de référence de la politique évaluative), en tant que tels les éléments prescrits demeurent génériques et doivent être opérationnalisés. De fait, il insiste sur le fait que la conception du plan d'évaluation est du ressort de l'équipe évaluative externe, plus précisément de l'évaluateur sénior responsable du projet.
- S'il n'était pas impliqué, du fait de ses quelques expériences en tant que junior collaborateur, il a connaissance de la démarche entreprise pour définir le plan d'évaluation dans le respect des paramètres prescrits, à savoir la phase de description du programme et d'évaluation de l'évaluabilité (examen préliminaire de la documentation et données administratives disponibles) dont les extrants sont les questions d'évaluation et la matrice évaluative.
- À conscience du fait que plus le programme gouvernemental est complexe (de grande envergure avec de multiples partenaires organisationnelles, un grand nombre de parties prenantes clés et d'activités, etc.), plus le devis méthodologique devient conséquent, la phase d'évaluation de l'évaluabilité cruciale et la collaboration étroite avec la division interne de l'évaluation indispensable.
  - Répartition/optimisation des ressources, envergure des données collectées et profondeur des analyses — pour 'J3' le recours aujourd'hui à des logiciels de gestion de référence et traitement de données qualitatives représente un tournant important dans l'effort de rigueur et de profondeur dans le recueil, l'organisation et l'analyse des données qualitatives (rappelons que dans son cas plus de 3000 documents ont été examinés et une soixantaine d'entrevues ont été réalisées).
- Il décrit avec clarté et profondeur les méthodes de collecte de données auxquelles il a pris part. Il démontre des savoirs opérationnels et expérientiels (connaissances tacites) importants en ce qui a trait au processus de sélection de sources documentaires, de gestion, de codage et d'analyse d'information qualitative, à l'utilisation de l'instrumentation et de logiciels de gestion de référence et de traitement de données qualitatives, des procédures de synthèse et interprétation des résultats.
- Pour lui, la fiabilité et la validité reposent sur la réalisation des tâches méthodologiques « critiques » (incluant la communication de l'évaluation et de ses résultats), par l'évaluateur sénior chargé de l'évaluation externe avec l'appui du comité d'évaluation interne, plus précisément de la division de l'évaluation interne.
  - Il reconnait l'aspect paradoxal du rôle de garant de la qualité du comité d'évaluation interne, puisqu'il peut être une source potentielle de biais, surtout dans la synthèse des résultats et la formulation des conclusions — ici, pour 'J3', l'équipe évaluative externe joue à son tour le rôle de « contrôleur ».

Encadré 12. *Les caractéristiques de la pratique réelle de l'évaluateur novice 'J1'*

'J1' **Lead junior** **collaborateur principal** (*1re expérience dans ce rôle*)	A participé à l'ensemble des étapes du projet d'évaluation. - Participation à l'ensemble des rencontres de travail de conception du devis d'éval. - Révision de l'instrumentation. - Coordination des activités de collecte et d'analyse de données (incluant liaison avec l'équipe éval. interne, la supervision des éval. juniors exécutants). - Participation à l'ensemble des activités de collecte de données - analyse documentaire, entrevue. - Analyse de données qualitatives (préparation d'une matrice de résultat par méthode). - Synthèse des résultats et formulation des constatations évaluatives - accord interévaluateurs avec l'évaluateur sénior chargé du projet d'évaluation à l'externe. - Rédaction de la 1re version du rapport d'éval. final et conception du plan d'implantation et de monitorage de l'approche (extrants du projet).
Firme [B] Formation [PhD SHS] [DES ÉP] Expérience pro. ÉP [18 mois/Act.1] [28 mois/Act.2]	

- Fait preuve d'une grande maitrise des facteurs environnementaux (contraintes politiques, ressources disponibles, particularités du programme et du projet d'évaluation, etc.) et des enjeux méthodologiques associés à ceux-ci.
- Décrit avec clarté et précision la démarche et le processus décisionnel engagé dans le cadre de la conception du devis d'évaluation, phase à laquelle il a la chance de contribuer.
  - En toile de fond, on trouve un point de référence (de comparaison), à savoir les pratiques évaluatives standard du gouvernement fédéral. Mais, au cœur des descriptions des décisions se trouve les caractéristiques de l'*evaluand*, les préoccupations relatives à la représentativité, à l'utilité.
  - Fait preuve d'un esprit critique vis-à-vis du contexte souvent rigide de l'évaluation dans le milieu du gouvernement fédéral, un cadre très prescriptif.
  - Notons, qu'ultimement la construction du devis d'évaluation, de l'instrumentation est du ressort de l'évaluateur sénior chargé du projet d'évaluation à l'externe.
- A une bonne compréhension du rôle des parties prenantes clés qui composent le CMI (gestionnaires de programme, division interne de l'évaluation), et de la dynamique entre le CMI et l'équipe évaluative externe.
  - Fait état du caractère indispensable de l'implication des gestionnaires de programme dans la démarche de cadrage de l'évaluation, l'identification des sources d'information.
  - Perçoit la division interne de l'évaluation comme un allié crucial – agent de liaison entre l'équipe évaluative externe et les gestionnaires de programme, aiguilleur méthodologique, facilitateur d'accès aux données.
  - En revanche, il reconnait faire l'amalgame entre les clients, c'est-à-dire les gestionnaires de programme, et la division interne de l'éval. Il semble tantôt percevoir la division interne de l'éval. comme le client, tantôt comme une ressource méthodologique interne, un agent de liaison, un conseiller auprès des clients, ce qui freine la possibilité d'identifier clairement où repose le pouvoir décisionnel, qui est l'entité dirigeante du projet à l'interne.
- Participer à la conception du devis et être force de propositions méthodologiques, coordonner la collecte et la gestion des données, interagir librement avec les clients, rédiger l'ébauche du rapport final, et concevoir les outils extrants du projet, représente une expérience d'apprentissage riche, un moment charnière dans le développement professionnel de l'évaluateur junior.
  - Armer d'une connaissance et compréhension profonde du devis et des besoins des utilisateurs ciblés il se sent plus en contrôle des tâches de collecte et d'analyse de données, plus en maitrise de l'instrumentation, plus à même d'aller en profondeur et de recueillir des informations utiles et justes.
  - À la différence de son expérience d'exécutant, il se perçoit comme étant un réel contributeur au projet d'évaluation et s'est sentit responsabilisé – « prise de conscience de soi » en tant qu'évaluateur de programme (identité professionnelle).
  - Rend compte d'une augmentation de la confiance en ses capacités d'évaluateur, tout en faisant état de ses besoins en termes d'apprentissage et de l'importance d'offrir très tôt, et souvent, la chance aux évaluateurs juniors de participer à des projets d'évaluation du début à la fin.
- Les mesures d'assurance de la fiabilité et de la validité reposent pour lui sur les efforts de rigueur et de transparence dans la communication de la démarche évaluative, sur la reproductibilité du devis, la démonstration de l'alignement « question éval. – indicateurs - sources de données – données – résultats ». Elles font échos aux pratiques recommandées de l'évaluateur junior (principal outil – matrice évaluative, matrice de données/résultats).
  - Il a conscience que cela ne transparait pas dans le contenu du rapport final dont il a rédigé la première ébauche. Pour lui, le contenu du rapport d'évaluation doit répondre aux exigences/attentes des clients, et faciliter l'utilisation des résultats. Il souligne l'importance de produire un rapport méthodologique et des rapports techniques complets.

4.2.2 Le cas des évaluateurs séniors (experts)

L'activité d'explicitation portait pour chacun des évaluateurs séniors, sur une évaluation d'un programme du gouvernement fédéral prescrite par la politique évaluative fédérale de 2009, soit l'obligation d'évaluation des programmes tous les cinq ans afin de rendre compte de leur pertinence et de leur rendement (efficacité, efficience et économie). Les évaluateurs séniors participants eurent la charge de leur projet d'évaluation respectif, ce qui a permis le recueil de détails riches en ce qui a trait à leurs choix et pratiques méthodologiques réels.

À l'instar de l'activité d'analyse critique, une grande partie de l'activité d'explicitation fut consacrée au développement du cadre d'évaluation, plus particulièrement les processus décisionnels et méthodologiques engagés sur le plan du contenu des éléments clés du cahier des charges (définition des questions d'évaluation et des indicateurs de mesure/observation, choix des sources de données, identification des sources d'information). Les explicitations fournies par les évaluateurs séniors nous ont permis d'explorer davantage les dynamiques de pouvoir décisionnelles mises en exergues durant la conception et la mise en œuvre du devis d'évaluation, où le rôle de l'évaluateur sénior oscille entre exécutant, collaborateur et décisionnaire.

Le Tableau 30 (p. 272) offre une synthèse générale de nos observations en ce qui a trait au contenu du devis d'évaluation présenté au sein des différents rapports faisant l'objet de l'activité d'explicitation, et que les évaluateurs séniors participants ont rédigé, avec la rétroaction du chef de l'évaluation (l'évaluateur interne).

Tableau 30. *Synthèse des observations dans le cadre de l'analyse du contenu des différents devis d'évaluation rapportés par les évaluateurs séniors*

Critères principaux	Critères détaillés	Respect des critères			Observations
		Non	Part.	Oui	
Présentation du modèle logique				S1 S2 S3 S4	• Inclus en annexe. Dans les cas de 'S1' et 'S4', le modèle logique était accompagné d'une description détaillée de la théorie du programme et de ces différentes composantes.
Description des objectifs de l'évaluation	Clarté des objectifs		S2 S3	S1 S4	• Le mandat de l'évaluation était généralement décrit succinctement. S'il est clair que les différentes évaluations répondaient à l'obligation d'évaluation des programmes tous les cinq sous la Politique évaluative fédérale, c'était parfois la seule justification offerte. Dans le meilleur des cas, l'évaluateur précisait qu'il s'agissait de rendre compte de la pertinence et du rendement du programme comme prescrit par les directives de la politique évaluative fédérale et du plan d'évaluation stratégique de l'agence gouvernementale commanditaire. • Dans le cas de 'S4', l'auteur apporte des précisions supplémentaires en ce qui a trait à sa décision d'inclure un axe d'évaluation de la conception du programme, celui-ci ayant fait l'objet de changements majeurs. • Si le plan d'évaluation semble généralement fidèle au type d'évaluation choisi, l'absence des matrices évaluatives dans la plupart des cas ainsi que le manque de précision sur le plan de la description du contenu du cadre méthodologique souligne un certain manque de transparence.
	Précision des objectifs	S2	S4 (+) S1 S3		
Justification du type d'évaluation	Pertinence du type d'évaluation choisi pour répondre aux objectifs			S1 S2 S3 S4	
	Justification du choix de l'évaluateur		S1 S2 S3	S2	
	Fidélité du plan d'évaluation au type d'évaluation choisi			S1 S2 S3 S4	
Présentation des questions d'évaluation	Pertinence des questions au regard des objectifs évaluatifs			S1 S2 S3 S4	• 'S4' est le seul évaluateur à avoir inclus les questions dans la section « objet et portée de l'évaluation » du devis d'évaluation. 'S1', 'S2' et 'S3' les présentent dans la partie « résultats » du rapport d'évaluation. 'S1' et 'S3' les ont également inclus en annexe. • Si les questions d'évaluation étaient généralement très claires, elles n'étaient en revanche pas toujours précises. • Dans le cas de 'S2', le contenu des questions était très général et offrait peu d'indications en termes d'observables. 'S4' offre des questions généralement précises, cependant certaines nécessitaient un découpage en sous-questions. 'S1' et 'S3' présente un éventail de questions et sous-questions d'évaluation très précis.
	Clarté des questions			S1 S2 S3 S4	
	Précision des questions	S2	S4 (+)	S1 S3	
Présentation des matrices d'évaluation		S2 S3 S4		S1	• Le manque de précision observé sur le plan des questions d'évaluation pour certains évaluateurs séniors aurait pu être résolu par l'inclusion des matrices dans le rapport.
Description de la méthodologie	Description des sources d'information et échantillonnages		S2 S3 (-)	S1 S4	• Espace accordé aux procédures de collectes et d'analyse de données très restreint (1 à 3 pages) sachant que chaque devis comprenait au moins quatre sources de données. Le contenu rapporté dans le cas de 'S3' est le plus lacunaire (1 page), avec peu voir aucun effort de description. • Le choix des multiples méthodes de collectes de données n'est pas justifié, mais dans le cas de 'S1' nous avons les matrices évaluatives. Pour chaque source de données, les évaluateurs offrent une brève description des échantillons, une indication quant aux objectifs de la collecte d'information, et quelques éléments procéduraux relatifs au recueil des données. On remarque un plus gros effort de description des procédures d'échantillonnage dans le cas de 'S1' et 'S4', des procédures de collectes de données dans les cas de 'S2'. • Seuls les cas de 'S1' et 'S2' font brièvement mention d'éléments relatifs à la conception de l'instrumentation (p.ex. prétest des questionnaires d'enquête, modèle d'examen de dossier conçu à partir des matrices ou socle d'indicateurs de mesure de rendement). • Les procédures d'analyse et de l'agrégation des données ne sont pas présentées dans les cas de 'S3' et 'S4', et si elles ne sont pas nécessairement décrites par 'S1' et 'S2', elles sont tout même brièvement mentionnées. • Chaque rapport offre une description détaillée des limites méthodologiques et les efforts entrepris pour les contrecarrer.
	Description des procédures de sélections des sources de données	S2 S3 S4	S1 (-)		
	Description des procédures de collectes de données		S1 S4 S3(-)	S2	
	Description des procédures de conception de l'instrumentation	S3 S4	S1 S2 (-)		
	Description des procédures d'analyses/agrégation des données	S3 S4	S1 S2 (-)		
	Description des limites méthodologiques			S1 S4 (+) S2 S3	

4.2.2.1 Conception du cadre d'évaluation : explicitations du processus d'opérationnalisation des paramètres de l'évaluation prescrits dans la demande de propositions

Au regard des explicitations fournies par les évaluateurs séniors, le mandat et les paramètres de l'évaluation qu'on trouve dans la demande de propositions semblent être prescrits par le plan d'évaluation stratégique quinquennal des programmes financés par l'agence gouvernementale (commanditaire de l'évaluation) qui est lui-même conçu selon les directives de la politique évaluative fédérale de 2009. À ce niveau de l'évaluation, le prestataire de service, tout comme les parties prenantes clés du programme, est un exécutant, le pouvoir décisionnel appartenant à la division interne de l'évaluation (et au gestionnaire de l'évaluation/évaluateur interne). Il est question ici d'évaluations sommatives centrées sur la pertinence et le rendement (efficacité, efficience et économie) du programme sur une période donnée (cycle de cinq ans de financement).

En outre, les évaluateurs séniors nous expliquent qui si la demande de proposition prescrit tous les éléments clés du cadre d'évaluation, son contenu demeure générique, à l'image du cadre de référence de l'IGF, particulièrement en ce qui a trait aux questions et à la matrice d'évaluation dont le contenu doit être affiné et opérationnalisé. Pour ce faire, le gestionnaire de l'évaluation externe entreprend une phase itérative de « *scoping-design* » à l'intérieur des paramètres d'évaluation prescrits (portée, intrants disponibles, sources d'information, sources de données, etc.), en collaboration étroite avec l'évaluateur interne, et en consultation avec les parties prenantes clés du programme et de l'agence gouvernementale.

Comme nous l'explique les évaluateurs séniors, cette étape du projet d'évaluation débute toujours par une rencontre avec le comité d'évaluation interne afin de cerner les attentes des parties prenantes clés, de discuter du contenu du cahier des charges envisagé, d'obtenir leur rétroaction, leurs recommandations quant aux informateurs clés et autres sources d'information essentielles, et d'accéder à toutes les informations complémentaires accessibles sur le programme. Par la suite, les évaluateurs séniors mobilisent différentes méthodes pour soutenir le processus de définition et d'opérationnalisation du contenu du cadre d'évaluation (c.-à-d. définir pour chaque question d'évaluation : son contenu spécifique, ses indicateurs de mesure/d'observation, et les sources d'information et de données qui permettent de les documenter). 'S4' mentionne réaliser une

analyse du programme et de l'évaluabilité ancrée dans l'examen approfondi de la documentation du programme et la collaboration étroite avec le gestionnaire de l'évaluation à l'interne. 'S2' et 'S3' expliquent recourir à des entrevues préliminaires (« scoping interviews ») avec les gestionnaires de programme, et la consultation du client (évaluateur interne). Dans tous les cas, les évaluateurs séniors s'appuient sur des outils ou ressources documentaires clés comme : la matrice évaluative ('S2' et 'S3' mentionnent une matrice standard de l'agence commanditaire incluse dans la demande de propositions), le modèle logique, la stratégie ou le cadre de référence de mesure de rendement du programme, grille de rapport de rendement (socle d'indicateurs pour évaluer l'efficacité – atteinte des résultats), etc. Enfin, pour 'S1' la situation est quelque peu différente puisque cette phase avait eu lieu une première fois dans le cadre d'un précédent contrat, lorsqu'il était chargé de la conception d'un plan d'évaluation détaillé pour chacun des services de l'agence gouvernementale commanditaire (en réponse aux exigences de la politique évaluative fédérale), dont celui faisant l'objet de l'évaluation discutée. Aussi, au départ de ce projet d'évaluation, il était en possession d'une matrice évaluative complète dont le contenu avait été conçu à partir de groupes de travail avec les gestionnaires de programme, d'entrevues de « *scoping* », et d'une analyse documentaire. Il s'agissait donc davantage de valider le contenu du plan d'évaluation développé deux ans auparavant à la lumière du contexte évaluatif et de la « réalité du programme ».

Remarquons que le non-recours à des entrevues préliminaires auprès des gestionnaires de programme ne signifie pas nécessairement que ces derniers n'ont pas été sollicités afin d'obtenir des précisions, des renseignements complémentaires. En effet, les évaluateurs séniors communiquent très régulièrement avec le chef de l'évaluation à l'interne, non seulement pour obtenir sa rétroaction et ses recommandations, mais aussi pour qu'il consulte les parties prenantes clés internes (ou facilite la consultation avec ces derniers), afin d'obtenir des précisions, des renseignements complémentaires, de la documentation supplémentaire.

Extraits de verbatim

"Scoping interviews with the key manager of the program to ask questions like – what are you interested in? What do you want to find out with the evaluation? Just to make sure the evaluation responds to senior management interest. Also, you get a little extra of information about the program, where to find the data, relevant source of information, identifying the main issues, etc. [...]. It's all part of the RFP issued by the department. What they want to evaluate, the evaluation questions they want you to answer, the line of evidence. And with this department, they kind of have this standard evaluation matrix that they kind of apply systematically to every evaluation. But then for instance during the scoping interview phase you can play a little bit with the questions, I can take out indicators if it's not relevant here, or if I can't get information, data for it. [...]. For example, the last question about effectiveness was not part of the standard template matrix but it's something that came in the scoping interviews, details regarding initiative goals, some work being done in a specific area, and they wanted us the see if it was achieving its goal. [...]. So yeah, by looking at the documentation like the performance reporting, the logic model, etc. and using the scoping interviews with program managers, and feedback for the evaluation steering committee, we developed a list of outcomes used to assess performance more precisely the efficacy so the achievement of intended outcomes." (S3).

« Pour moi la conception des questions d'évaluation c'est une tâche beaucoup bâtie sur notre expérience. C'est de l'art plus qu'une science. C'est pour ça que c'est toujours un évaluateur sénior qui est en charge. Il faut se bâtir rapidement une tête sur le fonctionnement du programme et ses rouages, absorber des exemples de projets, cerner les enjeux, etc. Évidemment, consulter les évaluations antérieures. Le modèle logique pour les stratégies de mesure de rendement. Mais on ne l'a pas pris à la lettre, car il était beaucoup axé sur les résultats, et on voulait regarder l'implantation aussi, il fallait que je comprenne ce que fait le gars le lundi matin quand il arrive au bureau. [...]. Et puis après je m'assois et m'attelle aux questions, aux matrices. Quand j'ai une bonne ébauche, je l'envoie au client, c'est-à-dire l'évaluateur interne. On échange, revoie le contenu, il me conseille, et puis quand on est satisfait je mets tout ça dans le rapport méthodologique. Il consulte, sollicite les gens du programme à l'interne si on a besoin de précisions, de renseignements supplémentaires. Parfois je suis présent. [...]. Par exemple, on a observé que la base de données ne générait pas systématiquement des renseignements sur tous les aspects des projets financés. Ça pose des problèmes ici, on touche aux indicateurs de rendement du programme... On veut pouvoir produire des renseignements globaux sur l'ensemble des projets, les détails des contrats, les extrants et les résultats atteints, les dépenses, etc. Du coup, on a travaillé en étroite collaboration avec les gestionnaires pour déterminer la nature, et puis ben la portée des informations, des données disponibles aux fins de l'évaluation. [...]. C'est un gros travail, car il faut s'assurer qu'on peut répondre aux questions, qu'on va être en mesure de trouver l'info. Donc on réfléchit où trouver l'information, les bases de données, etc. On décide quand les entrevues sont pertinentes pour accéder aux infos manquantes, aller plus en profondeur. Et puis ce n'est pas fini, parce qu'on a tant de semaines pour le faire, tant de budgets, ici on était plus limité que la norme pour les entrevues, alors on a décidé de les faire en groupe pour voir plus de monde, avoir la meilleure couverture possible de la population du programme. Après, si je n'ai rien pour documenter ou que c'est trop compliqué en termes d'accès, de ressources, ben l'indicateur saute. » (S4).

À la lumière des propos tenus par les évaluateurs séniors vis-à-vis de leurs expériences d'évaluateur responsables de l'évaluation à l'externe, le pouvoir décisionnel apparait s'équilibrer entre les gestionnaires de l'évaluation (interne et externe) durant le cadrage de l'évaluation. En effet, s'ils exécutent l'évaluation planifiée par leur client, et se voient prescrire le contenant de l'évaluation, ils sont de leur côté maitres du contenu, des détails procéduraux (opérationnalisation du cahier des charges). Nous revenons plus en détail sur les décisions engagées dans la planification de la méthodologie évaluative dans l'introduction de la section portant sur le cadre méthodologique. L'évaluateur interne quant à lui facilite le processus de consultation en interne,

aiguille quant à la faisabilité des procédures méthodologiques, donne sa rétroaction et relais celle du comité d'évaluation interne. Ultimement valide le contenu du cahier des charges, la division de l'évaluation étant l'entité décisionnaire. En outre, 'S4' fait explicitement mention d'une équipe évaluative composée de l'évaluateur interne et des membres de l'équipe évaluative externe. 'S4' et 'S2' parlent d'une collaboration étroite entre le chef de l'évaluation (l'évaluateur interne) et l'évaluateur sénior responsable du projet d'évaluation à l'externe. 'S1' et 'S3' ne font pas nécessairement référence au chef d'évaluation ou d'évaluateur interne, ils citent leur interlocuteur de référence clé, source principale de soutien et de rétroaction dans le cadre du projet d'évaluation, soit le représentant de la division interne de l'évaluation (S1), ou encore le gestionnaire du comité d'évaluation interne/représentant de l'autorité interagence (la division de l'évaluation interne) (S3).

Enfin, tout au long de l'évaluation, les parties prenantes clés du programme et de l'agence gouvernementale membre du comité d'évaluation interne dirigé par l'évaluateur interne exercent quant à eux un pouvoir consultatif.

4.2.2.2 Communication du devis d'évaluation : explicitations des carences observées sur le plan du contenu rapporté

Dans le cadre de l'analyse du contenu des différents rapports d'évaluation faisant l'objet de l'activité d'explicitation, nous avons constaté que l'espace octroyé au devis d'évaluation, et plus particulièrement au cadre méthodologique, demeurait très restreint. En effet, comme nous avons pu le voir dans la synthèse de nos observations (Tableau 30, p. 272), son contenu oscille entre une page et demie, et quatre pages, ce que nous avions trouvé surprenant sachant que chaque devis d'évaluation comprenait au moins quatre sources de données. Aussi, on nous décrit de manière plus ou moins succincte, pour chaque source de données, ses objectifs, les sources d'information consultées et la taille des échantillons, et le déroulement du recueil de données. Parmi les carences observées, on pouvait trouver : 1) l'absence de justification des choix de méthodes de collectes de données ; 2) le manque d'explicitation des procédures de sélection de certaines sources d'information (échantillonnages) ; 3) le manque de description des procédures de collectes de données et de conception de l'instrumentation ; et (ou) 4) le manque, voir l'absence de description des procédures d'analyse et d'agrégation des données.

Face à ce constat, les évaluateurs séniors nous ont expliqué que chaque agence gouvernementale a ses propres habitudes de communication de l'évaluation, et que la plupart des hautes instances de gestion exigent des rapports finaux beaucoup plus succincts, et axés sur les résultats d'évaluation. Aussi, les détails d'ordre méthodologique ayant fait l'objet de différents rapports techniques, ces derniers sont les premiers écartés du corps du rapport final (parfois inclus en annexe). Le Tableau 31 à la page suivante illustre les explications fournies par les participants en réponse à nos différentes observations.

Tableau 31. *Explicitations des évaluateurs séniors quant aux carences observées sur le plan du contenu du devis d'évaluation présenté dans leur rapport d'évaluation respectif*

Critères principaux	Critères détaillés	Respect des critères			Explicitations
		Non	Part.	Oui	
Présentation du modèle logique				S1 S2 S3 S4	« *Ce que je trouve gênant ici c'est que pour moi le modèle logique c'est le fondement de l'évaluation - pour la compréhension du programme, pour la conception des matrices, de l'instrumentation, etc., et maintenant c'est mis en annexe, voire pas du tout inclus dans le rapport. Et ça j'ai vraiment du mal.* » (S4)
Présentation des questions d'évaluation	Dans le devis d'évaluation	S1 S2 S3		S4	"*It's not like the questions are missing from the report, it's structured around the matrix, and they are they are the headings structuring the results section. Like having the logic model in the appendix, it's a way to lighten the core content of the report. To focus more on reporting detailed findings.*" (S1)
	Dans la section résultats			S1 S2 S3 S4	
	En annexe	S4		S1 S2 S3	
Présentation des matrices d'évaluation		S2 S3 S4		S1	« *Franchement je ne trouve pas ça nécessaire d'avoir les matrices. Au fur et à mesure qu'on connait mieux le programme elles dérivent beaucoup, on change les indicateurs, on en enlève au moment du développement des instruments si on se rend compte qu'il y en a trop.* [Chercheur - ça permettrait de saisir pleinement les choix méthodologiques, les procédures, surtout quand on n'a pas les questions évaluatives dans le devis] - *Ça c'est sûr, mais encore une fois c'est eux qui décident du contenu.* » (S2) "*So often, the way of the report, the wording, the content, what's in it, we have no control. The content of the report varies between agencies. Here, at X Canada, it's something they do, having everything in the appendix - the logic model, the evaluation matrix, the instruments. I know that not all agencies have this amount of detail. That's why I like to work for them. It's more rigorous in term of describing the methodology, how we went about answering the evaluation questions, which indicators we looked at and what lines of evidence we used to look at it, etc.*" (S1)
Description de la méthodologie	Description des sources d'information et échantillonnages		S2 S3 (-)	S1 S4	« *Quand je rédige le rapport d'évaluation, je me fie au guide de rédaction de l'IGF, aux pratiques recommandées, ce qu'il faut présenter, les questions d'évaluation, le modèle logique, etc. J'aime beaucoup ce format là et j'y répond, mais après ça c'est le modèle de X Canada, c'est eux qui décident du contenu final. Les ministères d'importance, ils ont chacun leur format d'évaluation. La norme ici... c'est un rapport de max 30 pages, la métho on diminue beaucoup, met en annexe, c'est une victime de cet effort-là.* [...]. *Pour les instruments, c'est parce qu'ils sont dans le rapport méthodologique.* » (S2) "*Looking at other reports from X Canada... This agency really thrives for shorter report like 30 pages maximum. The details about the design like the evaluation matrix, the methodology procedures are in the methodology technical report. [...]. They basically have a template for the final report content. At this point all reports look the same. If there is something special that should be added they might be open to it, but usually its very prescriptive, centered on the findings. [...]. I would like to add that the technical report is really transparent in terms of detailing the procedures, the methods, the instruments, etc.*" (S3) « *Au fil des années les évaluations fédérales ont été beaucoup critiquée pour la lourdeur des rapports. Les utilisateurs ciblés ce ne sont pas des chercheurs, eux ils veulent voir les résultats. 15 pages de métho ça les intéresse pas du tout. D'où les rapports techniques. Oui, ils veulent s'assurer que c'est bien fait, et là c'est le rôle du technicien interne. À leur niveau, ils ne veulent plus voir ça. 10 pages pour présenter le programme, le devis c'est déjà le max. Pour eux la qualité y est du fait du processus. Et puis, il y a le coût de la traduction. [...]. On en revient aux targeted audience. Mon rapport est pour les dirigeants, ils l'ont vu eux le rapport métho, les décisionnaires auront accès à tout, même s'ils ne lisent probablement rien. [...]. Mais c'est vrai quand on présente les résultats ils nous relancent sur la métho, ils veulent la comprendre.* » (S4)
	Description des indicateurs de mesure			S1 S2 S3 S4 (-)	
	Description des procédures de sélections des sources de données	S2 S3 S4	S1 (-)		
	Description des procédures de collectes de données		S1 S4 S3 (-)	S2	
	Description des procédures de conception de l'instrumentation	S3 S4	S1 S2 (-)		
	Description des procédures d'analyses/agrégation des données	S3 S4	S1 S2 (-)		
	Description des limites méthodologiques			S1 S4 (+) S2 S3	

Ce qu'il est important de retenir ici, c'est que le contenu méthodologique des rapports discutés ne répond pas nécessairement aux exigences des évaluateurs séniors participants, qui étaient chargés de la rédaction des versions initiales. En effet, ces derniers étaient parfaitement conscients des carences sur le plan de la description du devis d'évaluation, déplorant comme nous l'absence de certains éléments essentiels comme les matrices évaluatives, ou encore l'impossibilité pour le lecteur de saisir pleinement les aspects d'ordre procédural en ce qui a trait à la conception et la mise en œuvre du cadre méthodologique. En revanche, les évaluateurs séniors ont tenu à expliquer que cela ne reflète en aucun cas la qualité (ou plutôt le manque de qualité) de l'évaluation. Chaque étape de conception du devis d'évaluation est conciliée dans un rapport méthodologique, le principal outil de collaboration des gestionnaires de l'évaluation (interne et externe), et de rétroaction du comité d'évaluation interne. Son contenu est approuvé par le chef de l'évaluation, l'évaluateur interne, représentant de l'autorité interagence (la division de l'évaluation interne de l'agence gouvernementale commanditaire de l'évaluation). Ce rapport technique, dont le contenu est extensif, contient les détails de la phase de planification de l'évaluation : les questions et les matrices évaluatives, la description des différentes procédures de collecte de l'information, l'instrumentation, les matrices de données (ou preuves) sur lesquels sont basées les procédures d'analyse et d'interprétation des résultats, etc. Aussi, selon 'S4', la confiance des utilisateurs ciblés de l'évaluation envers la qualité de ce processus explique leur choix de centrer le rapport sur la présentation des résultats, même s'il reconnait être souvent questionné sur la méthodologie lorsque ces derniers se penchent sur les résultats présentés. L'effort de transparence et de rigueur dans la description du processus de conception et de mise en œuvre de l'évaluation dans le cadre d'un rapport méthodologique (dont le contenu est validé par le chef de l'évaluation interne), et de rapports techniques sur l'analyse des données et l'interprétation des résultats pour chaque source de données est le premier moyen cité par les évaluateurs séniors pour assurer la validité, la fiabilité et la crédibilité des procédures et des résultats d'évaluation. Cela inclut les temps de présentation orale du devis d'évaluation une fois qu'il est finalisé, et des résultats de l'évaluation avant la rédaction/soumission de la première ébauche du rapport final auprès du comité d'évaluation interne.

4.2.2.3 Conception du cadre méthodologique

On compte parmi les détails méthodologiques fournis dans la demande de propositions, des indications quant aux sources de données attendues pour répondre aux objectifs évaluatifs, et les sources d'informations disponibles. Les évaluateurs séniors expliquent qu'il n'y a rien de novateur et qu'il s'agit des méthodes qu'on retrouve toujours dans le cadre d'évaluations de programmes du gouvernement fédéral. Ils mentionnent également que la demande de propositions rend compte des bases de données existantes, des groupes d'informateurs clés ciblés, et donne également au consultant un ordre d'idée quant au nombre de personnes à interviewer et de cas à étudier.

Les décisions engagées par les gestionnaires de l'évaluation à l'externe (nos évaluateurs séniors) sur le plan de la définition de la méthodologie de collecte de l'information reposent davantage sur les aspects d'ordre procéduraux en ce qui a trait à la complémentarité des méthodes et l'importance de la triangulation ainsi que l'appréciation de la pertinence des méthodes envisagées au regard de la faisabilité (intrants disponibles, accessibilité des données/informations, contexte évaluatif, etc.), et de la fiabilité et représentativité des informations qui vont être collectées. Ainsi les évaluateurs séniors nous ont fait part notamment des préoccupations suivantes dans la définition des procédures de collectes de l'information (dont les objectifs sont résumés dans le Tableau 32, p. 286) :

Le cas de 'S4'

- Contraintes budgétaires et temporelles : 'S4' insiste sur la nécessité de demeurer réaliste dans le choix final des sources de données, et de faire preuve d'ingéniosité pour contourner les contraintes susceptibles d'affaiblir la représentativité des résultats.

  → 1) Exclusion de la recension des écrits, une source de données généralement faible dans le cadre de ce type d'évaluation ; 2) optimisation du processus de consultation des informateurs clés malgré un budget « entrevue d'informateurs clés » limité en optant pour un format de groupe, lorsqu'approprié, et en utilisant les études de cas, ces dernières comprenant la passation d'entrevues (19 entrevues, n=31 informateurs clés). Remarquons que ces stratégies d'optimisation sont observées chez tous les évaluateurs séniors participants.

- Contraintes d'accès aux données - le risque d'épuisement des informateurs clés : 'S4' nous explique qu'on se trouve dans un contexte où très souvent, plusieurs évaluations se déroulent sur une même période donnée, et certains groupes de répondants sont sursollicités. Il nous

explique que le sondage, le moyen de consultation idéale au regard de l'envergure de la population ciblée des évaluations fédérales, est aujourd'hui utilisé avec parcimonie. En effet, le faible taux de réponse traditionnellement obtenu dans le cadre de ce type d'évaluation, pousse les évaluateurs à émettre des réserves quant à son utilisation, même s'il demeure un choix privilégié des clients.

→ 1) Exclusion de l'enquête ('S4' n'étant pas convaincu de son utilité), et utilisation des résultats d'une enquête auprès des industries financées réalisée durant une autre évaluation ; 2) augmentation du nombre d'informateurs clés consultés grâce au recours à des entrevues de groupe. Si 'S4' décrit le choix explicite d'écarter l'enquête au profit de l'utilisation de données secondaires, on retrouve parmi la documentation utilisée par les différents évaluateurs séniors, des enquêtes auprès des membres du personnel, des clients, etc., qui ont été réalisées à l'interne ou dans le cadre d'autres projets d'évaluation.

- Contraintes d'accès aux données - base de données du système d'information de gestion incomplète : 'S4' aborde ici la disparité observée sur le plan des données de gestion recueillies pour les différents projets financés par le programme et examinés dans le cadre de l'évaluation. En effet, cela posait des difficultés dans la production de renseignements globaux sur l'ensemble des projets, un élément essentiel pour rendre compte de l'atteinte des résultats du programme.

  → 1) Collaboration étroite avec les gestionnaires de programme pour déterminer la nature et la portée des renseignements et des données disponibles pour l'examen des projets financés (complémentarité analyse documentaire – analyse de données administratives) ; 2) utilisation des entrevues d'informateurs clés et des études de cas pour obtenir des renseignements plus détaillés sur les projets et les activités financés dans le cadre du programme.

Le cas de 'S3'

- Complexité du programme et intrants disponibles pour l'évaluation : le programme évalué par 'S3' est une initiative vaste et complexe qui implique plusieurs agences gouvernementales et un nombre important de parties prenantes clés dans la gestion et la mise en œuvre de ses cinq volets. Aussi, 'S3' indique qu'il était impossible d'offrir une rétroaction détaillée sur tous les aspects de l'initiative, et qu'ils n'ont pas eu d'autre choix que de centrer l'évaluation sur les activités clés des différents volets de l'initiative. Ils ont ensuite redoublé d'efforts pour assurer la représentativité des résultats d'évaluation.

→ Optimisation du nombre d'entrevues budgétées pour couvrir l'ensemble des composantes du programme (avec consultation équilibrée d'informateurs clés internes et externes), en ayant recours à des passations en groupe (47 entretiens, 61 informateurs clés).

- Malgré tout, 'S3' nous explique qu'ils obtenaient un nombre limité de perspectives internes et externes sur chaque volet du programme (et leurs activités clés) ; et qu'ils traitaient certains aspects fondamentaux du programme à un niveau plutôt global, générant ainsi un potentiel manque de représentativité et de profondeur en ce qui a trait à son déroulement et son efficacité.

  → À la lumière des informations obtenues (analyse documentaire et entrevues), sélection des cas à étudier pour combler les lacunes méthodologiques. 'S3' remarque que, généralement illustrative et peu utile, l'étude de cas permet ici de recueillir des informations plus détaillées sur les initiatives clés des différents volets du programme, avec pour chaque cas, entre trois et cinq entrevues d'informateurs clés supplémentaires. Malgré tout, 'S3' reconnait qu'en fin de compte l'envergure du programme est telle que celui-ci possède des angles morts, des éléments qui sont passés au travers des mailles du filet, non seulement de l'évaluation, mais aussi du monitorage interne du programme.

Le cas de 'S1'

- Utilité des sources de données : 'S1' nous explique que le cadre méthodologique est le même pour chaque plan d'évaluation des différentes branches de l'agence commanditaire, celles-ci ayant des caractéristiques similaires sur le plan des pratiques professionnelles dans le cadre du traitement des dossiers/affaires. Aussi, cela a permis à 'S1' (en collaboration avec la division interne de l'évaluation) de revoir le choix des méthodes à la lumière notamment de la pertinence et de la richesse des informations collectées à la suite de l'évaluation d'une branche.

  → Exclusion des groupes de discussion, ces derniers ne semblant pas enrichir les informations obtenues grâce à d'autres sources de données, particulièrement l'examen de dossiers/affaires.

- Nécessité de trianguler et contraintes éthiques/déontologiques : la branche évaluée et ses employés travaillent sur des dossiers/affaires soumis au secret professionnel, ce qui comme nous l'explique 'S1' compliquait le processus de triangulation, l'équipe évaluative externe ne pouvant pas accéder aux traces écrites susceptibles de corroborer les pratiques décrites par les informateurs clés, et d'étayer les informations recueillies dans le cadre de l'examen compréhensif des données administratives.

→ Inclusion de l'examen de dossiers pour soutenir l'évaluation du rendement de la branche - échantillonnage et extraction des données réalisées par l'évaluateur interne et son équipe ainsi que les professionnels ayant eux-mêmes travaillé sur certaines affaires auxquelles la division interne de l'évaluation ne peut accéder. Pour 'S1', les caractéristiques professionnelles des praticiens du milieu, notamment leurs capacités d'analyse critique au regard de leurs pratiques, et leurs préoccupations en ce qui a trait à la qualité et la rigueur des services offerts à leurs clients, éliminent toutes inquiétudes face à de potentiels biais. Par ailleurs, pour lui, la notion de biais a tendance à disparaitre lorsqu'on rend compte des pratiques idéales, et qu'on ne cherche pas à faire ressortir des erreurs ou difficultés dans les dossiers traités.

Le cas de 'S2'

- Contraintes d'accès aux données et rentabilité : 'S2' explique qu'il avait accès à des séries de données de rendement sur les projets financés par le programme basées sur de très bons indicateurs quantitatifs, très clairs et bien délimités, et ce, pour chaque résultat direct et intermédiaire avec la cible et le résultat déclaré ; alors que très souvent ils ont affaire à des descriptions très qualitatives. En outre, dans le cas de 'S2', le problème reposait sur le fait que les données n'ont pas été transmises sous forme de base de données, et qu'il s'agissait en réalité d'une succession de rapports d'impacts des projets financés par le programme rassemblés dans un imposant classeur de feuilles de calcul électronique. En l'état, l'évaluateur sénior nous explique qu'il n'était donc pas en mesure de rapporter directement les résultats ou de mener facilement des analyses afin de rendre compte du rendement à l'échelle du programme.

  → Reconfiguration complète de la « base de données » par le technicien spécialiste de bases de données de l'équipe évaluative externe, et ce, en dépit des coûts supplémentaires importants (temps, budget) et par extension de la faible rentabilité financière du projet d'évaluation pour le cabinet-conseil contractant. Puis, réalisation des analyses statistiques descriptives à partir de différents tableaux de données par 'S2' afin de montrer les résultats du programme pour chaque indicateur. L'évaluateur sénior réitère à plusieurs reprises que ça lui tenait à cœur de bien faire les choses, particulièrement face à la qualité et l'ampleur des données de mesure de rendement disponibles et la possibilité de réaliser des analyses significatives de l'efficacité du programme. Ce qu'il n'est pas toujours possible de réaliser dans le cadre d'évaluations de programmes gouvernementaux. Il remarque tout de même que malgré l'emploi d'un spécialiste informatique, la situation n'est pas toujours propice pour une telle entreprise, son caractère chronophage nécessitant

l'adhésion et le soutien du gestionnaire des données à l'interne, ce qui n'est pas toujours acquis (particulièrement quand le système de gestion des données est remis en question). Ici, le gestionnaire de la base de données de l'équipe évaluative externe avait déjà collaboré à plusieurs reprises avec le responsable du système de gestion des données de rendement interne, et ce, avec succès, ce qui a conforté 'S2' dans ses choix méthodologiques. Enfin, 'S2' précise que malgré tout, la base de données contenait des données manquantes pour 20% des projets financés, et que ces derniers ont dû être exclus des analyses de rendement comparant les résultats ciblés aux résultats atteints.

Enfin, il est important de garder à l'esprit que le processus de définition des éléments clés des procédures de collectes de données (sources d'informations, indicateurs de mesure, sources de données/instrumentation, etc.) débute au moment de la conception du cadre d'évaluation, plus précisément de la matrice d'évaluation. En effet, nous l'avons vu un peu plus tôt dans le cadre de l'explicitation de la phase de « *scoping-design* », les choix méthodologiques des évaluateurs séniors dépendent de la possibilité de recueillir, pour chaque indicateurs ou groupe d'indicateurs de mesure, des informations provenant de multiples sources de données. Par ailleurs, comme les participants nous le mentionnent à plusieurs reprises, le contenu de la matrice évaluative et du cadre méthodologique évolue tout au long du projet d'évaluation. En effet, outre la collecte d'information pour répondre aux questions d'évaluation, la mise en œuvre des différentes sources de données peut permettre l'identification de nouvelles sources d'informations ou spécifier où se trouvent l'information (réexamen des banques de données et fichiers associés tels que les rapports finaux des projets financés, par exemple), la définition/modification/suppression d'indicateurs, la facilitation de la conception de l'instrumentation, la définition/modification de certaines procédures de collectes de données, celles-ci étant mises en œuvre de manière asynchrone, etc.

a. Les méthodes de collecte de données sélectionnées

Nous venons de voir que le choix des sources de données est généralement prescrit par le client dans la demande de propositions, et que les décisions méthodologiques des évaluateurs séniors reposent sur la définition des objectifs de chacune des méthodes de collectes de données, et l'identification des différentes méthodes pour chaque indicateur ou groupes d'indicateurs de mesure. Ultimement, il s'agit d'assurer la triangulation des sources de données, et de miser sur la complémentarité des méthodes pour garantir la fiabilité et la représentativité des informations

collectées (Tableau 32, p. 286). Ainsi, au regard des explications fournies par les évaluateurs séniors, une hiérarchie apparait dans l'organisation des méthodes standards utilisées, particulièrement en ce qui a trait à l'évaluation du rendement du programme, ce qui nous parait tout à fait logique et pertinent. En effet, comme les évaluateurs séniors l'ont souligné durant l'activité d'analyse critique, il est essentiel de recueillir des données probantes sur les extrants et l'atteinte des résultats du programme, et les principales sources de données sont l'examen de données administratives (exploitation des données des systèmes de monitorage et de gestion de l'information), et la revue de documents (rapports de progrès, d'activités, de rendement, résultats d'enquête auprès des employés, des clients, etc.). 'S2' s'appuie également fortement sur l'enquête pour compléter le contenu de la base de données de rendement générée par le système d'information de gestion du programme. Dans les cas de 'S1' et 'S4', on observe le recours à l'examen de dossiers/projets qui s'inscrit dans la lignée de l'examen des bases de données. Le but de l'examen de dossiers selon les évaluateurs séniors est de rendre compte du cycle de vie d'un échantillon de projets financés dans le cadre du programme (S4) ou de dossiers traités par la branche (S1), à l'appui des indicateurs de processus et de rendement. Pour 'S1' il s'agit également d'un moyen de corroborer les informations fournies par les professionnels spécialisés de la branche qui ont été interviewés. Le recours aux entrevues et enquêtes auprès d'informateurs clés vise à approfondir, expliquer ou compléter les données sur l'efficacité du programme (extrants et atteintes des résultats). À travers ces sources de données incontournables, il s'agit de permettre aux gestionnaires, aux financeurs, aux employés, aux clients, au membre de la population ciblée du programme, etc., de donner leur rétroaction, de partager leurs expériences, leurs points de vue, etc. sur le fonctionnement, les activités et le rendement du programme. Enfin les études de cas ont une visée davantage illustrative afin de donner une idée au lecteur sur ce que le programme fait, les défis, les réussites, etc. Sauf pour 'S3', où elle permet de recueillir des informations plus détaillées sur les initiatives clés du programme avec pour chaque cas, entre trois et cinq entrevues d'informateurs clés supplémentaires.

Dans le cas de 'S3', il nous parait important de mentionner que contrairement aux autres évaluateurs séniors, le cadre méthodologique rapporté ne donnait aucune indication quant aux objectifs des différentes collectes de données, un manquement renforcé par l'absence des matrices évaluatives.

Tableau 32. *Objectifs des méthodes de collecte de données utilisées par les évaluateurs séniors*

'S3'	'S2'	'S4'	'S1'
Analyse documentaire • Examiner les processus du programme et recueillir des informations sur ses différents volets et leurs activités respectives. • Compiler des données permettant de rendre compte de l'efficacité du programme. Entrevues d'informateurs clés • Obtenir la rétroaction des parties prenantes clés (internes et externes) sur le fonctionnement de chaque volet du programme, le déroulement des activités, les résultats, les défis, etc. • Étayer les données extraites de la documentation sur le rendement. Enquête auprès des membres des groupes de travail externes • Recueillir la perspective des participants sur l'efficience et l'efficacité des consultations de parties prenantes Études de cas • Explorer en profondeur chaque volet du programme et renforcer la compréhension de certains aspects importants. • Expliquer ou compléter les informations obtenues durant les autres collectes de données → intrant très important concernant l'efficacité des activités clés du programme.	Analyse documentaire • Recueillir des renseignements sur les paramètres de financement, la gouvernance, l'architecture du programme et l'alignement avec les priorités ministérielles, résultats attendus, rendement, etc. • Fournir des réponses sur la pertinence et le rendement du programme. Examen des données du Système d'information de gestion • Évaluer l'atteinte des résultats du programme, l'efficience et l'économie. • Réaliser une analyse du rendement. Entrevues d'informateurs clés • Permettre aux intervenants internes et externes d'exposer leur point de vue sur l'ensemble des questions d'évaluation. • Recueillir des informations en ce qui a trait aux points forts/faibles de la gouvernance du programme. Enquêtes auprès des candidats du programme *Enquête 1 - bénéficiaires du financement* • Recueillir le point de vue des clients sur l'efficacité des services du programme • Obtenir des informations sur les effets des différents projets financés. *Enquête 2 - non-bénéficiaires du financement* • Obtenir un aperçu de l'impact différentiel du programme en comparant les projets financés et les projets non financés. Études de cas/Examen de dossiers • Illustrer les types de projets soutenus par le programme. • Mieux comprendre la dynamique et les processus à l'échelle du projet (objectifs, intrants, extrant, atteinte des résultats, défis, leviers, etc.).	Examen de documents • Recueillir des renseignements liés aux activités et services fournis par les groupes ciblés (ou sous-programmes), aux accords de financement avec les compagnies privées subventionnées par les groupes ciblés, etc. Examen des données du Système de suivi du rendement • Recueillir des renseignements quantitatifs liés aux activités menées par les groupes ciblés du programme. • Recueillir des renseignements primaires sur les projets financés par groupes ciblés. Examen des données de projets • Offrir un aperçu de la variété des projets financés. • Examiner les processus de l'appel d'offre jusqu'au rapport soumis pour les projets sélectionnés et recueillir des données sur les extrants et les résultats atteints, afin de prouver l'efficacité et le rendement du programme. Entrevues d'informateurs clés • Acquérir une compréhension approfondie du contexte des groupes ciblés du programme, des résultats obtenus par ces groupes et les défis auxquels ils font face. • Corroborer, expliquer ou préciser les conclusions émises à la suite des résultats des autres sources de données → intrant très important concernant l'atteinte des résultats du programme. Études de cas • Illustrer les réussites, les défis et les pratiques exemplaires de projets financés par les groupes ciblés du programme.	Examen de documents • Aider à la conception de l'instrumentation. • Fournir des réponses à l'ensemble des questions d'évaluation. • Identifier les composantes du programme, les indicateurs nécessitants de plus d'informations. Examen des données du Système de gestion de l'information • Aider à la conception de l'instrumentation • Recueillir des données sur l'efficacité de la branche (observation des indicateurs de rendement). Examen de dossiers • Offrir un aperçu du large éventail de dossiers/affaires traités chaque année et mettre en exergue les cas les plus complexes, les plus sensibles. • Acquérir une compréhension plus approfondie du cycle de vie dossier au regard des indicateurs de rendement de la branche. • Corroborer les informations obtenues dans le cadre des entrevues d'informateurs clés. Entrevues d'informateurs clés • Acquérir une compréhension approfondie de l'expérience des personnes clés, et obtenir des informations sur le besoin continu de la branche et l'efficacité de ses activités. Études de cas • Illustrer des pratiques exemplaires mises en œuvre dans le cadre de dossiers/d'affaires spécifiques. Enquête anonyme auprès d'informateurs clés • Obtenir la rétroaction des professionnels qui travaillent sur les différents dossiers de la branche.

b. La sélection des sources d'information

Selon les explications fournies par les évaluateurs séniors, la sélection des sources d'information se fait en collaboration étroite avec la division interne de l'évaluation et les parties prenantes clés du programme membres du comité d'évaluation interne. Aussi, le gestionnaire de l'évaluation externe, en collaboration avec l'évaluateur interne, détermine les caractéristiques descriptives des sources d'information, des échantillons, et l'évaluateur interne et les parties prenantes clés identifient les sources d'information. Il est question d'échantillon de convenance pour chaque collecte de données, et c'est pour cette raison que 'S2' mentionne préférer parler de sélection des sources d'information plutôt que d'échantillonnage. En outre, les caractéristiques descriptives des sources d'information et les critères de sélection sont déterminés à partir du contenu de la matrice évaluative défini par le gestionnaire de l'évaluation externe (en collaboration avec l'évaluateur interne), et des intrants disponibles pour chaque méthode de collectes de données (c.-à.-d. budget, échéancier, mains-d'œuvre, etc.). Il est également important de préciser que les extrants de certaines sources de données facilitent l'identification/le processus de sélections des sources d'information pour d'autres procédures de collecte de données.

Documentation examinée : le processus d'identification et de sélection des documents se déroule en interne. Les gestionnaires du programme et de ses différents volets, à l'appui des matrices évaluatives, identifient et font parvenir à l'équipe évaluative interne les documents qu'ils estiment pertinents pour montrer la pertinence, l'efficacité et l'efficience du programme (sources importantes de données secondaires sur le rendement du programme). Il est généralement question ici des documents suivants (documents générés par le programme, le département, l'agence gouvernementale) : plans stratégiques, documents de planification et d'opération, descriptions du programme (objectifs, exigences, intrants, extrants, cibles, etc.), rapports d'activités, rapports de rendement, accords de financement, rapports financiers, enquêtes auprès du personnel ou des clients, rapports d'évaluations antérieures, présentations et documents d'information, etc. Aussi, comme nous l'expliquent les évaluateurs séniors, l'évaluateur interne et son équipe compilent une liste des documents reçus qu'ils évaluent en fonction de leur contribution aux différentes questions d'évaluation (et leurs indicateurs respectifs), avant de faire parvenir le corpus au gestionnaire de l'évaluation externe. Au fil de l'analyse documentaire, l'équipe évaluative externe relance les

gestionnaires du programme si un complément d'information est nécessaire (toujours par l'intermédiaire gestionnaire de l'évaluation interne).

Extraits de verbatim

"We don't select the documents, it's given to us, we just review what's given to us. It's what happened when you work with an effective evaluation division like the one from this branch. The staff really knows what kinds of documents we will need. Sometimes we must help then but typically they get there. They are very proactive in making sure that we get the documentation that's exactly needed for the evaluation. They are given the task to gather information and to send it to us. Then during the review process, we will identify the documentation we still need and notify them, and they will in turn solicit the people from the program." (S1)

"Based on the evaluation questions and the indicators, each area of the program sent documents to the internal evaluation manager, the key person from X Canada, who then compiles a list of the key documentation and reviews the documents in terms of their contribution to the evaluation questions and their indicators. Some programs try to direct us by replicating the questions and giving the reference where we would find the data. Sometimes they would do it directly in the matrix." (S3)

Données administratives examinées : les évaluateurs externes sont avisés dès la demande de propositions de l'existence de bases de données provenant des divers systèmes de suivi - des systèmes d'information de gestion informatisés de dossiers/affaires traités par la branche (S1), de projets financés par le programme ('S2' et 'S4'), sur la période évaluée. Ces fichiers électroniques (classeurs de feuille de calcul) fournissent des renseignements descriptifs, des données sur le rendement du programme (extrants et résultats atteints, résultats ciblés versus résultats réels), des données opérationnelles sur l'affectation des ressources humaines, des données financières, etc. En outre, les évaluateurs séniors expliquent que le processus d'obtention des données administratives est entrepris très tôt par le gestionnaire de l'évaluation externe (immédiatement après l'obtention du contrat d'évaluation), du fait notamment du caractère chronophage et laborieux de l'obtention de l'ensemble des données nécessaires, et des enjeux méthodologiques importants pour la planification des autres procédures de collectes de données (échantillonnages, conception de l'instrumentation, passation, etc.) et les efforts de triangulation. Comme nous l'explique 'S2' cela implique la collaboration, ou au moins la coopération des gestionnaires des données du programme, mais aussi la patience et la compréhension des clients lorsqu'il s'avère primordial, pour le bien de l'évaluation, d'attendre la réception des données.

Extrait de verbatim

« *On nous avait dit dans l'appel d'offres qu'il y avait un système d'information de gestion, et donc qu'il y avait une base de données de performance. Et ce qu'on a reçu c'était effectivement des données de performance, mais ce n'était pas sous forme de base de données. C'était essentiellement une suite de rapports soumis par les gens qui avaient obtenu un financement pour leur projet dans le cadre du programme, et qui étaient colligés dans un méga fichier Excel. Donc onglet 1-projet 1 voici les impacts, onglet 2-projet 2 voici les impacts, etc. Pas du tout organisé comme une base de données. On ne pouvait pas juste faire un simple examen du fichier et extraire les données sur le rendement du programme afin de rapporter les résultats à l'échelle du programme. Donc heureusement, ici, chez nous on a un très bon technicien en bases de données qui lui a réussi à réorganiser l'information pour la transformé en véritable base de données. [...]. Ah, puis c'est important de préciser qu'ici, on a vraiment été chanceux, car notre gestionnaire de la base de données, notre informaticien, il connaissait personnellement le gestionnaire de données interne et ça nous a beaucoup aidé, ça a grandement facilité notre accès aux données, l'obtention de définitions, etc. Parfois on fait face à beaucoup de résistance à l'interne pour accéder à toutes sortes d'informations. Dans le pire des cas, on a les informations trop tard. Ça m'est déjà arrivé. Ça fait beaucoup de dommages si on dépend de ces infos pour d'autre méthode comme le sondage. On a besoin des coordonnées, s'ils nous envoient les coordos très tard dans le processus évaluatif, du coup on n'a pas de temps pour faire le sondage. Et puis ben parfois l'information ne se rend pas jusqu'à nous. [...]. Les données administratives, les données financières, il faut les demander dès le début, car on sait qu'il va y avoir beaucoup de va-et-vient entre. Parce qu'on va les recevoir, puis il va manquer des données, on ne saura pas ce que veut dire telle affaire, puis là il faut qu'on pose des questions. Ça peut durer des mois, des mois. Moi je fais beaucoup d'effort pour obtenir les données administratives. J'essaie de m'y prendre le plus tôt possible, et puis j'essaie de convaincre mes clients d'attendre les données si on ne les a pas, que c'est pour le bien de l'évaluation. »* (S2).

Projets ou dossiers examinés : Le processus de définition des critères de sélection des projets financés par le programme (S4), ou dossiers traités par la branche (S1), est facilité par l'examen des données administratives. Dans le cas de 'S4', l'équipe évaluative a collaboré avec les gestionnaires du programme pour sélectionner les projets à examiner. Dans le cas de 'S1', les dossiers traités par la branche sont protégés par le secret professionnel, aussi il a tenté de guider au mieux le processus de sélection des dossiers à partir de critères précis, de données de rendement et autres renseignements non confidentiels rapportés dans les matrices de résultats, la sélection étant réalisée par la division interne de l'évaluation (en collaboration avec le comité d'évaluation interne), et avec la participation des praticiens spécialisés de la branche. 'S1' précise que la définition des critères d'échantillonnage ne fut pas aisée compte tenu du fait qu'il n'avait aucune idée du contenu même des dossiers.

→ Le cas de 'S1' - échantillon intentionnel, non représentatif : l'évaluateur sénior nous annonce dès le départ que la quantité de dossiers traités chaque année par la branche est telle que l'échantillon sélectionné est beaucoup plus illustratif que représentatif. Aussi, il s'agissait de rendre compte au mieux du large éventail de dossiers traités par la branche, et d'offrir un aperçu des affaires les plus complexes ou sensibles.

→ Le cas de 'S4' - échantillon intentionnel, représentatif : l'évaluateur sénior nous explique que le nombre de projets sélectionnés a été déterminé selon la qualité, l'accessibilité et l'ampleur des renseignements disponibles sur les extrants et les résultats atteints par les projets, lesquels prouvent l'efficacité du programme et son rendement. L'échantillon représente une fourchette de variation dans les caractéristiques descriptives (contrats et accords de financement, taille du projet, institution/organisme commanditaire du projet, type de projet, etc.).

Extrait de verbatim

"The branch treats thousands of cases each year so there is no way to be representative here. The files were not chosen randomly, and the sample is not large. The goal is to give a flavour of the broad spectrum of the kind of work they do and particularly highlight the more high-profile cases, the ones that representant more risk, the difficult ones, sensitive ones. The file review is intended to illustrate the branch approach to its work. We asked them to give us files that represent best their work. We did not have much control with this particular one. We tried to... We did get performance data, etc. and make a general request with criteria but, because of the [confidentiality] we can't know the files or any details, we were a bit blind. And they are areas where even the evaluation branch can't look at the files, so we had to rely on the branch specialized practitioners themselves to choose the files to be examined." (S1)

Individus interviewés : les évaluateurs séniors débutent leurs propos en précisant que le nombre d'entrevues réalisées est déterminé par le budget accordé pour cette méthode de collecte de données, un budget qui reflète le nombre d'individus à interviewer anticipé/recommandé par le client dans la demande de propositions. Ils expliquent ensuite que parmi les groupes d'informateurs ciblés se trouvent les incontournables, ceux que nos participants appellent « *les gens du programme* », soit les informateurs clés internes (haut gestionnaires et gestionnaires de programme, financeurs, employés, etc.), puis les informateurs clés externes, les individus qui ont un lien de près ou de loin avec le programme (clients, population ciblée, communauté, organismes gouvernementaux ou non gouvernementaux, etc.). Aussi, au cœur des critères de sélection des individus pour chaque groupe ou catégorie de répondants clés, on trouve des préoccupations relatives à la qualité de la couverture des composantes/activités du programme, à l'équilibre entre les perspectives internes et externes, et à l'optimisation du processus de consultation des informateurs clés (nombre d'entrevues budgété versus nombre de personnes interviewées, utilisation des études de cas qui comprennent également la passation d'entrevues d'informateurs clés). À la lumière de tout cela, les évaluateurs séniors sollicitent les parties prenantes clés membres du comité d'évaluation interne pour qu'ils fournissent une liste d'informateurs clés à partir de laquelle l'évaluateur externe, en collaboration avec l'évaluateur interne, va finaliser le processus d'échantillonnage.

Extraits de verbatim

"It's really a collaborative process in terms of identifying the specific individuals. We give a list of criteria for each targeted groups, and the internal evaluation working group with the program managers gives a list of informants. Here the main goal is to roughly reflect the branch, you know, the number of people within each group of respondents that roughly reflected the composition of each division. I wouldn't say there was some sort of scientific sampling that was being done, and it was more of an idea of we want to have a fair weight and range of types of key informants to provide us with a number of perspectives." (S1)

« Il y a d'abord les des gens inévitables, les gens aux différents niveaux du programme, les parties prenantes internes, et puis ensuite il faut couvrir les groupes d'individus qui ont un lien avec le programme de près ou de loin, les représentants de projets financés par le programme, les industries concernées par les projets, les experts, etc. On veut garder une proportion appropriée de parties prenantes internes et externes. Un certain équilibre sur le plan des perspectives. Dans ce cas ici on voit l'impact du budget, ce n'était vraiment pas beaucoup 19 entrevues. Après en faisant des passations en groupe quand c'était approprié, selon le profil des individus, des groupes. On veut que le gens se sentent libres de parler ouvertement. On a pu tout de même interviewer 31 personnes. Après il y avait aussi les études de cas qui nous ont permis de réaliser une dizaine d'entrevues supplémentaires, et de combler un peu les manques. » (S4).

Cas étudiés : le nombre de cas à étudier semble généralement déterminé dans la demande de propositions, ou au moment de la répartition du budget alloué à la collecte de l'information**.** Le processus d'identification et de sélection des cas est un effort collaboratif entre les gestionnaires de l'évaluation (externe et interne) et les parties prenantes clés du programme membre du comité d'évaluation interne. Ce processus est généralement facilité par les informations obtenues dans le cadre des autres collectes de données et le contenu du modèle logique. Il peut s'agir de projets financés par le programme ('S2' 'S4'), de dossiers d'affaires traitées par les employés de la branche (S1), ou encore des initiatives ou activités mises en œuvre dans le cadre des différents volets du programme (S3). L'objectif ici est de sélectionner des cas qui permettent de rendre compte du fonctionnement du programme, des défis rencontrés, des pratiques idéales, etc. L'ensemble des évaluateurs séniors s'accordent sur le fait qu'à la lumière du nombre limité de cas, généralement compris entre trois et cinq, il n'est aucunement question de représentativité. Il s'agit plutôt d'illustrer, ou dans le cas de 'S3' d'étayer des données collectées sur certaines dimensions clés du programme dans le cadre des entrevues.

Extrait de verbatim

"Here, with a massive program, with so many different players, and with five components and their respective activities, once we allocated the 47 interviews to 61 key informants, we had maybe two people per activities, so once down to one component you didn't have a lot of views to sort of understand what was going on regarding certain aspects. We had a quality coverage but limited in terms of quantity. We were dealing with things at a more aggregated level. If something important emerged, we did a case study. Generally useless, here it is an example where case studies contributed a lot. We waited for the results of other lines of evidence before sampling case studies, to cover gaps in the methodology. With four case studies, we were able to do a deep dive in the program different components, we talked to an additional three to five key informants per program's selected initiatives." (S3)

Individus sondés : il semblerait qu'aucun processus d'échantillonnage n'ait été mis en œuvre. En effet, les caractéristiques descriptives et la disponibilité de différentes listes de contact à l'interne, font que les évaluateurs ont la plupart du temps directement accès à leurs groupes de répondants ciblés. 'S2' mentionne ainsi qu'il est désormais très rare de faire face au besoin d'échantillonner dans le cadre du recours à l'enquête, un extrant des cycles d'évaluation des programmes gouvernementaux, et des différents systèmes d'information de gestion interne. Ici, parmi les individus sondés, on retrouve l'ensemble des informateurs clés internes et (ou) externes, et ce, généralement en vue de renforcer la représentativité des données collectées en ce qui a trait à l'efficacité du programme dans le cadre des entrevues. Il s'agit par exemple des individus ayant obtenu une subvention auprès du programme (S2), des membres de groupes de consultation communautaire (S3), ou encore d'employés spécialisés qui traitent des dossiers/affaires gérer par la branche évaluée (S4).

c. Les procédures de collectes de données

Les différentes procédures de collecte de données sont déterminées par le gestionnaire de l'évaluation externe et documentées dans le rapport méthodologique, principal outil de communication et de rétroaction avec le gestionnaire de l'évaluation interne, et de consultation pour les parties prenantes clés du programme membres du comité d'évaluation interne. Ainsi, les évaluateurs séniors participants furent responsables de la conception des différents instruments de collectes de données, et de la préparation/coordination des passations dans le cadre de leur évaluation respective. Veuillez noter que les choix et procédures méthodologiques observés sont globalement les mêmes pour chaque évaluateur sénior.

Les principaux outils d'aide à la *conception de l'instrumentation* (structure, organisation, formulation des items, etc.) sont la matrice évaluative (questions/sous-questions d'évaluation et indicateurs de mesure) ainsi que les matrices de données et de résultats produites par les différentes méthodes de collectes d'informations (informations à corroborer, compléter, expliquer).

Dans le cadre des différents types d'examen de sources documentaires (revue de documents, de données recueillies par les systèmes d'information de gestion, de projets/dossiers), la matrice évaluative transformée en matrices de données constitue l'instrumentation. 'S1', 'S2' et 'S3' ont eu recours à différents gabarits ou modèle - fichier de feuilles de calcul ou de traitement de texte électronique, où ils retranscrivent les informations pertinentes, produisant ainsi une

matrice de données pour chacune des sources de données utilisées. Dans le cas de ‘S4’, ce dernier explique utiliser un logiciel de traitement de données qualitatives où il importe l’ensemble des sources documentaires (documents internes du programme, base de données et documentation des projets financés) afin de coder et classer directement l’information.

Extraits de verbatim

“For the document analysis, we don’t use any software to collect the data. We use the evidence matrix in Excel. You know just as you are coming through the docs that audit and evaluation sent, you fill in the matrix which is a tool itself. It’s a document review template. Some program even replicates the evaluation matrix and gives the references where we would find the data for the different questions or directly writes the data in the matrix.” (S3).

« *Pour l’examen des documents internes, du programme et des projets financés, mais aussi de la base de données des projets financés, on n’avait pas un gabarit dans Excel ou Word où placer toutes les informations, les types de documents sont généralement trop différents – listings, courriels, memos, pages web, fichiers Excel, PDF, PowerPoint, etc. On avait déjà établi dans notre cadre d’évaluation, dans la matrice, selon les indicateurs, ce que les documents devaient documenter. Donc là ce qu’on faisait, c’est très technique là, je vais essayer d’être succinct là. Nous on part de l’indicateur et si on sait que tel indicateur est documenté dans tel et tel document, on va aller chercher ces docs, et on va les amener dans NVivo et on va coder. Si par exemple notre indicateur c’est - quels sont les types de compagnies qui ont fait une demande ? On va recevoir toute la liste des compagnies, on va y mettre dans NVivo, et puis là on va classer selon les régions, le type de compagnie, le type de technologie, le financement alloué, etc. Et puis ce sera la base de notre analyse documentaire et des données administratives qui incluait aussi une analyse plus approfondie des projets financés par le programme, de l’appel d’offres jusqu’au rapport soumis, le recueil de renseignements sur les extrants et les résultats atteints par les projets, etc. Et on fait ça pour chaque indicateur. Dans NVivo, chaque indicateur va se retrouver dans une famille d’indicateur dans NVivo.* » (J4).

Remarquons que si les bases de données générées par les différents systèmes d’information de gestion des projets financés par le programme ou des dossiers traités par la branche évaluée constituent en soi des sources documentaires, ces dernières s’avèrent également être un outil essentiel de collecte et d’analyse de données. En effet, nous avons vu plus tôt que le processus d’obtention de ces données administratives, plus précisément d’une base de données complète, peut s’avérer être une entreprise laborieuse, car les évaluateurs externes sont dépendants des gestionnaires de données internes et de la qualité du système de suivi du programme. Aussi, dès la réception des banques de données (classeurs de feuilles de calcul électroniques), les évaluateurs externes semblent entreprendre un examen approfondi de leur contenu afin, notamment, de cerner et comprendre les renseignements recueillis, de définir les besoins d’informations complémentaires, de déterminer les données manquantes et identifier si les données existent et sont accessibles, de retirer les projets ou dossiers ayant une série de données incomplètes, de réorganiser les tableaux de données à la lumière des indicateurs, des questions d’évaluation, etc.

Retour sur le cas particulier de 'S2' – Analyse des données du système d'information de gestion à partir d'une base de données de rendement du programme conçu par l'équipe évaluative externe

Nous avons mentionné plus tôt que le programme possédait un système d'information de gestion visant à faire le suivi des mesures de rendement mises en place, un système qui recueillait des données tout au long du cycle des projets financés, provenant de formulaires de demande de financement, d'accords de contribution, de rapports de visite de site, de rapports finaux, etc. Ainsi, le système a généré notamment des séries de données quantitatives sur le rendement de chaque projet financé, qui rendaient possible la comparaison entre la valeur cible et la valeur réelle pour chaque indicateur de mesure de rendement du programme. Cependant, les séries de données n'ont pas été colligées sous la forme d'une réelle base de données, et ont été présentées sous forme de rapport d'impact pour chaque projet dans le classeur de feuilles de calcul électronique transmis par le gestionnaire de données interne. De fait, l'équipe évaluative externe n'avait pas en sa possession un rapport de rendement à l'échelle du programme à partir duquel elle pouvait extraire les données de mesure pertinentes (résultat du programme pour chaque indicateur de rendement), ou encore réaliser aisément les analyses permettant de rendre compte de l'atteinte des résultats du programme.

'S2' explique que le gestionnaire de bases de données de l'équipe évaluative externe a entrepris la reconfiguration de la base de données transmise par le gestionnaire de données interne afin de lui permettre de réaliser les analyses statistiques descriptives nécessaires pour répondre aux questions d'évaluation de rendement. Il a pu dresser des tableaux de données détaillés pour chaque indicateur clé (résultats directs et intermédiaires du programme) grâce à l'utilisation d'un modèle d'examen de données basé sur la stratégie de mesure du rendement du programme. Ce modèle permettait de déterminer les analyses à effectuer pour répondre aux questions d'évaluation, principalement sur les résultats du programme, mais aussi sur son efficience.

Retour sur le cas particulier de 'S1' - l'examen de dossier d'affaires protégées par le secret professionnel

Afin de corroborer les données collectées dans le cadre des entrevues d'informateurs clés, en ce qui a trait au rendement de la branche (efficacité, efficience), le cadre méthodologique incluait l'examen de dossiers d'affaires traitées par la branche. Cependant, en raison du secret professionnel, l'équipe évaluative externe n'avait pas accès aux renseignements recueillis par le

système d'information de gestion ou à la version papier du dossier propre à chaque affaire traitée. Ainsi, cette collecte de données a été réalisée en interne par la division de l'évaluation, et les professionnels spécialisés de la branche qui ont travaillé sur les différentes affaires sélectionnées. En outre, 'S1' nous explique qu'un modèle standard de révision de dossier a été utilisé pour s'assurer que les informations collectées sur les dossiers étaient comparables. Il se concentre sur la collecte d'informations factuelles sur la manière dont la branche procède lorsqu'elle travaille sur un dossier, sur le cycle d'un dossier. Son contenu a été développé à partir des indicateurs de processus et de rendement inscrit dans la matrice évaluative, définis en collaboration avec les parties prenantes clés de la branche dans le cadre d'un précédent contrat (conception d'un plan d'évaluation en réponse à la politique évaluative incluant le modèle logique, les matrices, grille d'indicateurs de rendement). Une phase dont il souligne le caractère essentiel et « salvateur » compte tenu des limites d'accès à la documentation protégée par le secret professionnel et ce que cela aurait impliqué dans le cadre de la conception de l'instrumentation. Enfin, ce formulaire d'examen de dossier était majoritairement constitué de questions fermées avec quelques questions à réponses courtes visant principalement à recueillir des données quantitatives (p. ex. le nombre d'employés par catégorie et le nombre d'heures effectuées par ces derniers sur le traitement du dossier), mais aussi de brève description de la nature des dossiers sans briser le secret professionnel.

Les évaluateurs séniors ayant eu recours à l'enquête ne se sont pas attardés sur le choix du type d'item utilisé. 'S1' mentionne l'utilisation d'items de questionnaire typiques et de l'ajout de quelques questions ouvertes, et il dit également avoir travaillé à partir d'un questionnaire qu'il avait utilisé pour une évaluation similaire. Dans le cadre de la discussion du traitement des données d'enquêtes, 'S2' et 'S3' mentionnent le recours à l'analyse de fréquence, car les items sont de types Likert, et 'S2' explique avoir également inclus des questions à réponse objective afin de recueillir des données complémentaires à celles de la base de données de rendement. Il s'agissait, plus précisément, d'items visant la déclaration du nombre d'heures de bénévolat et du nombre d'heures par bénévole enregistré dans le cadre des projets communautaires des répondants afin de pouvoir utiliser cette information dans l'analyse de l'efficience. En effet, 'S2' explique que les projets communautaires comme ceux financés par le programme évalué reposent sur un nombre considérable d'heures de bénévolat qui contribuent non seulement à l'efficacité du programme, mais aussi aux économies réalisées par ce dernier. Pour lui il était important de pouvoir attribuer

une valeur monétaire à la contribution de la communauté (application d'une théorie économique visant à traduire le nombre d'heures de bénévolat en coût financier). Il précise également que « *la politique évaluative voulait qu'on donne une idée du rendement et des bénéfices dans le cadre de l'analyse de l'efficience, coûts-efficacité, donc j'avais vraiment fait l'effort pour cette évaluation d'essayer de quantifier la valeur monétaire de la contribution bénévole.* ».

Enfin, les guides d'entrevues semi-dirigés semblent également tenir le rôle de gabarit de prises de notes, les entrevues n'étant pas enregistrées ('S1', 'S2' et 'S3'), ou retranscrites ('S4'). Dans le cas de 'S4', l'évaluateur sénior nous mentionne que les entrevues étant enregistrées, l'interviewer avait la possibilité de réécouter ses entrevues et d'avoir des notes plus complètes pour certains éléments clés abordés. Enfin, 'S2' et 'S3' mentionnent que les notes d'entrevues sont retranscrites dans une matrice de donnée où l'information est organisée par question d'évaluation, indicateur (ou question d'entrevue), et catégorie d'informateurs clés, et ce, en vue de préparer les données d'entrevues pour l'analyse.

Extraits de verbatim

"To design the survey, like for the guides, I sat down with the evaluation matrix, I looked at the questions and the indicators, and then I came up with questions that will be sure to address each question and indicator you know in a regular survey type of question with some open-ended questions. With the firm we did prior evaluation of X services where we used a survey so I was also working with that survey instrument that we had already designed and used, well because we could just tinker with that. Some of the questions were going to be the same, and they are good ones. The content of the survey was validated by the evaluation division which opted to not pilot it." (S1).

« *On a conçu un guide d'entrevue semi-structuré adapté pour chaque catégorie d'informateurs clés interviewés. On utilise le guide de l'interviewer qui contient plus de détails sur les indicateurs, etc. pour prendre des notes, certains essais de prendre les notes directement sur l'ordinateur, dans la matrice de données vu que c'est par téléphone. Donc, à la suite des entrevues, les notes sont saisies dans la matrice, et classées en fonction de la question d'entrevue comme indicateur, ou de la question d'évaluation pour qu'il soit possible de classer l'information selon la catégorie d'informateurs clés, et selon la question d'évaluation, à des fins d'analyse.* » (S4).

Au regard des explications fournies par les évaluateurs séniors, *différentes procédures semblent être mobilisées pour assurer la fiabilité et la représentativité des informations collectées*. Par exemple : l'adaptation d'un questionnaire utilisé auparavant, la validation du contenu auprès de la division de l'évaluation interne, le prétest des questionnaires auprès d'un échantillon de la population de répondants ciblés ('S2' et 'S3'), la traduction des questionnaires en français par un traducteur professionnel à la suite du prétest en anglais, la « formation » des interviewers à l'utilisation des guides d'entrevue, ou encore la révision du contenu des guides d'entrevues à la lumière des premières passations (au besoin). De surcroit, 'S2' et 'S4' mentionnent également partager le guide d'entrevue (ou une version abrégée) avec l'interviewé au moment de la prise de

rendez-vous afin que ce dernier puisse se préparer. Pour eux, cela ne constitue pas une potentielle source de biais, mais plutôt un moyen de renforcer la qualité des informations collectées, et de limiter le risque que l'interviewé ne soit pas en mesure de répondre, ou qu'il procure des informations incomplètes voir même erronées. Il s'agit donc, selon eux, d'un moyen de s'assurer que l'interviewer sera en mesure de recueillir des informations factuelles, et d'engager une conversation plus approfondie sur divers points clés tant sur le programme que sur leur expérience, sur des éléments qui ne sont pas dans le guide d'entrevue. Enfin, 'S2' rapporte qu'en faisant l'effort de reconfigurer la base de données de rendement transmise par le gestionnaire du système d'information de gestion du programme, ils ont été en mesure de repérer et de retirer 20 % des projets financés dont la série de données de rendement était incomplète. Il précise qu'en collaboration avec le gestionnaire de données interne, l'équipe évaluative a bien entendu tenté de récupérer les données manquantes avant d'écarter lesdits projets de l'échantillon final (examen de la documentation des projets, prise de contact avec les chefs de projets, etc.).

Extraits de verbatim

« Je communique toujours le guide. Mes évaluations sont beaucoup fondées sur la transparence. J'ai des collègues qui ne le font pas, d'autres qui partagent les grands thèmes du guide. Moi, je mets tout. Ça me permet le jour de l'entrevue de mettre vraiment l'accent sur le factuel. Je veux vraiment avoir la possibilité de creuser, voir les choses en détail, je veux vraiment savoir le pourquoi. Pour moi ce n'est pas un biais, au contraire. Il y a plus de risque si tu interview à froid, et que les gens ne pensent pas à une réponse. 'Je n'y ai pas pensé' est pour moi le plus gros biais quand il est question d'aller en profondeur dans les entrevues. » (S2).

"*We pre-tested the instruments. Well, we did an actual pretest for the survey with a sample of people from the targeted groups. For the interviews, we use the few first to readjust the content of the guides. Plus, the instrumentation is always included in the methodology reports for feedback and validation by the evaluation branch. The evaluation committee also has the opportunity to review the instruments and give feedback to the internal evaluator in charge. I don't know if they did it here."* (S3).

« *Les gens se sont préparés à l'avance, on leur a envoyé le guide, certains viennent même avec des notes détaillées. Une fois qu'on commence, on ne va pas formuler les questions à la lettre, on verbalise différemment, on relance, etc. Il y a beaucoup d'interactions qui ne sont pas dans le guide. C'est une vraie conversation. Et c'est là que l'expérience des intérieurs est importante. Les experts passent au travers des guides avec les plus juniors. C'est à moi, en tant que manager, de former l'équipe à l'utilisation des guides. Je prépare les interviewers avec les questions les plus importantes, celles où il est important de mettre l'accent. Je les préviens aussi des points les plus sensibles, et ceux sur lesquels j'anticipe une potentielle résistance. Et puis il y a un feedback loop au fil des entrevues.* » (S4).

Par ailleurs, lorsque nous avons questionné 'S1' sur le risque de biais de réponse dans les cas où la revue d'un dossier est réalisée par le professionnel qui était chargé de l'affaire, celui-ci nous a répondu qu'étant donné la nature de sa profession, ce dernier se doit d'être honnête. De plus, si les informations ne sont pas enregistrées dans le système d'information de gestion, elles le sont dans la copie papier du dossier. Il y a donc une certaine traçabilité et il n'est donc pas dans

leur intérêt de fournir des informations erronées, le département pouvant décider de vérifier ultérieurement certaines données recueillies pendant l'évaluation.

d. Les procédures d'analyses de données et de synthèse des résultats

Selon les évaluateurs séniors, les procédures d'analyses des données collectées, et de synthèse des résultats dans le cadre de leur évaluation respective sont très standards, et assez simples. Aussi, cela pourrait expliquer pourquoi ces derniers ne se sont pas attardés sur la description desdites procédures. Tout comme les procédures de conception de l'instrumentation et de passation, les procédures d'analyse semblent faire l'objet de schèmes automatiques bien huilés. En revanche, il semblerait que les plans d'analyse de données ne fassent pas l'objet d'un protocole méthodologique à soumettre auprès des clients pour approbation. Nous verrons que le comité d'évaluation interne se voit tout de même présenter les résultats de manière préliminaire, à des fins de rétroaction avant la rédaction du rapport final, et que les évaluateurs séniors remettent au gestionnaire de l'évaluation interne différents types de rapport technique (matrice de résultats ou de réponse aux questions d'évaluation), selon les attentes et les préférences de l'agence fédérale commanditaire.

Extrait de verbatim

« *There is no approval of this process, we don't put forward our data analysis plan as a separate deliverable document that need to be approved. We just go from the data collection to the data analysis to here are the preliminary findings. So, there is no oversight of that process. They will have technical reports for the different line of evidence, and the preliminary presentation of the findings, it's a sort of validation process. [...]. Also, we draft the final report, and it goes through a couple of cycles of revisions in light of comments or suggestions for the evaluation committee internally.* » (S2).

Dans le cadre du traitement des données qualitatives, les évaluateurs séniors ont recours à l'analyse de contenu thématique. Aussi, au regard de leurs explications, chaque corpus (documentation et données administratives, notes d'entrevues, projets financés/dossiers traités, etc.) semble d'abord faire l'objet d'une phase de codage thématique au cours de laquelle la matrice évaluative est transformée en une matrice de données organisée par question d'évaluation/famille d'indicateurs/type de source d'information ou catégorie de répondant, où selon l'outil utilisé, les informations pertinentes sont retranscrites (gabarit) ou codées (logiciel d'analyse de données qualitatives ou QDA). Par la suite, pour chaque thème/question d'évaluation, les informations sont analysées, et les résultats sont rapportés dans une matrice de résultats.

Sur le plan de l'analyse des données d'entrevues d'informateurs clés, les procédures mises en œuvre sont différentes selon nos évaluateurs séniors. Dans les cas de 'S2' et 'S3', chaque

interviewer retranscrit les notes de ses entrevues dans la matrice de données (gabarit). Par la suite, soit les questions d'évaluation sont réparties entre les différents interviewers ('S3', qui a participé à l'analyse), soit chaque interviewer analyse les données des entrevues qu'il a réalisées, ce qui semble davantage correspondre à une répartition par catégorie de répondants ('S2', qui a réalisé la majorité des analyses). En ce qui concerne 'S1' et 'S4', le corpus de notes d'entrevues est réparti entre deux évaluateurs responsables de l'analyse qui ont recours à un logiciel QDA. Dans le cas de 'S1, il s'agissait de lui-même, à titre de gestionnaire de l'évaluation, et de son collaborateur principal (ceux qui rédigent le rapport final). 'S4' quant à lui n'était pas impliqué dans l'analyse des données d'entrevues, mais il explique en revanche qu'à titre de gestionnaire, responsable de la synthèse et de l'interprétation des résultats d'évaluation (rédaction du rapport final), il a la possibilité de recoder ou de contacter les individus responsables pour qu'ils revoient leurs analyses.

Extrait de verbatim

« *On entre les notes d'entrevues dans la matrice de données d'entrevues, qui contient nos indicateurs, nos questions, et on fait une analyse qualitative proprement dite, c'est-à-dire qu'on va identifier les catégories de réponses, les thèmes. On ne va pas quantifier les catégories de réponses, mais on va retenir celles qui sont le plus utiles pour l'évaluation. Les réponses inutiles, on les rejette tout simplement. Puis on va retenir les explications, pourquoi les gens ont tels types de réponse, sans quantifier. On va donner des proportions. C'est pour ça qu'on a un peu une approche hybride au fédéral. On va à la fois faire une analyse de typologie de réponses, et puis pas pour tout, mais on va donner une idée sur la proportion des gens qui ont eu tels types de réponse. Moi là je ne donnerais pas de pourcentage sauf si on m'y oblige, et ça m'est déjà arrivé plus d'une fois, mais je n'aime vraiment pas ça. Et puis quand ça arrive, j'essaie de leur expliquer que si 80 % des individus disent une chose, ça ne veut pas dire que les 20 % restant pensent le contraire. Et c'est toujours ça le danger ! Si 80 % des gens disent que le programme est bon, car il implique la communauté, ça ne veut pas dire que les 20 % qui restent ne pensent pas la même chose, ils ne l'ont peut-être juste pas dit. Aussi, c'est pour ça que je vise la tendance majoritaire le plus possible. Et puis, ça aide vraiment le lecteur à comprendre les résultats majeurs. Sauf si bien entendu c'est pertinent pour la question de montrer les potentielles divergences, ou différences.* » (S2).

Pour la formulation des résultats obtenus dans le cadre des entrevues d'informateurs clés, 'S1' et 'S2' expliquent donner un ordre d'idée en termes de proportion d'individus ayant donné tel ou tel type de réponse selon un thème donné (c.-à-d. utilisation de quantificateurs : la majorité, la plupart, presque tous, quelques, etc.). Cependant, pour certains clients, cela ne s'avère pas toujours suffisant, ces derniers désirant un dénombrement exact – « 80 % des individus pensent que », « trois informateurs clés sur quinze rapportent que », etc. Or, selon les évaluateurs séniors, si certaines questions, en fonction du type d'informateurs clés, peuvent se prêter à une telle forme de présentation des résultats d'entrevues, cela constitue un risque important de mérepréSentation des résultats. En effet, 'S2' exprime le danger face à des formulations comme « 80 % des gens pensent que », car cela sous-entend que les 20 % restants pensent le contraire (ou ne le pense pas

du tout), alors qu'ils ne l'ont peut-être tout simplement pas dit. C'est pour cette raison qu'il choisit de rapporter, dans la mesure du possible, les tendances majoritaires, mettant ainsi en exergue les résultats majeurs. 'S1' insiste sur le fait que l'entrevue d'informateurs clés n'est en rien une enquête, et que ce désir de « comptage » apparent chez certains clients est potentiellement contreproductif. Il s'attarde ici plus particulièrement sur les situations où seul un nombre restreint d'informateurs clés est en mesure de fournir des informations sur un élément précis, ces derniers étant les seuls à avoir l'expérience (ou la perspective) requise. S'il recommande la formulation de ce type de résultats comme suit – « parmi les informateurs clés en mesure de discuter de la problématique suivante, la majorité explique que », il explique que certains départements ne sont pas satisfaits de ce genre de formulation. En effet, préoccupés par le nombre restreint d'individus concernés comparativement au nombre total d'informateurs clés, ces derniers souhaitent que ce soit reflété dans la formulation des résultats en indiquant que « seulement quelques informateurs pensent que » ou en précisant le nombre exact d'individus concernés. 'S1' explique qu'on minimise alors des résultats qui sont importants du fait justement de la provenance de l'information. Enfin, si ce ne fut pas le cas dans le cadre des évaluations discutées, les évaluateurs 'S1' et 'S2' expliquent que dans le cadre d'autres projets, malgré leurs efforts, ils n'ont pas été en mesure de convaincre leurs clients et n'ont pas eu d'autre choix que de répondre à leurs exigences. Face à cela, 'S1' exprime le souhait que certains clients fassent davantage confiance au jugement de l'évaluateur.

Extrait de verbatim	*"In my experience, there is this kind of movement with some departments where they really want the number of key informants who said something, like percentages. And if it's a few of them, there is this desire to just count it. And sometimes that's fair enough. It depends on the question you are answering, on the key informants. But I wish they would be a bit more open to kind of allowing us to bring our professional judgment there, on the information we are getting. Because it really isn't a survey. It's about getting a profound understanding about specific issues. They should focus more on the relevance, the quality of the information, the way it was obtained. [...]. I feel they are missing the point of these interviews, because sometimes only of a few informants can actually address a particular element, because only a few of them will have that experience and perspective. [...]. Every so often we have big conversations on how to report and use the results. Like we use a scale here, we don't give numbers. And some departments are fine with the scale as you see it here which gives an idea of proportion – a few, most, almost all, etc. But others want percentages attached to it, numbers. [...]. Also, for a lot of the questions, you can use this scale pretty effectively, but in some instances, you have to try to negotiate with clients, explain that you think this finding was important based on who is providing you with this information and even though it was just a small number of the overall number of key informants, they have a unique perspective that should be reflected in the report and not minimized. [...]. For me if only a few key informants could say something, then 90% of the key informants did. I would just say something like - key informants who could speak to this issue X, most of them think... But some departments will be like - how many people is that? If that's just a few, it's not as... It might not be removed but it might just say - only few said this or specify the number of people. So, from my perspective it would mean that it's been given less weight than it deserved. But from their perspective clearly, they think that's the better way to go. Sometimes you just have to agree to disagree with your client and do what they are comfortable with at the end of the day. But here again this department is very good with that, they allow you to report on key informant the way where I feel like you are really giving them the best information that you possibly can from what the key informants told you." (S1).*

Dans le cadre du traitement des données quantitatives collectées au cours d'enquêtes, les évaluateurs séniors concernés ('S1', 'S2' et 'S3') mentionnent l'utilisation d'un logiciel d'analyse statistique en vue de réaliser des analyses descriptives, plus précisément des analyses de fréquence, la majorité des items étant de type Likert. Les évaluateurs séniors 'S2' et 'S3' ont rapporté les résultats dans un rapport technique propre à l'enquête, tandis que 'S1' les a inscrits dans la matrice de résultats organisée par source de données.

Extrait de verbatim	« *For the survey data analyze it's as pretty straightforward with simple descriptive statistics looking at frequencies of responses using [a quantitative data analysis software]. And then like for every line of evidence, the survey had his own technical report where for each evaluation question we would give the relevant survey results and also a little bit of interpretation. [...]. I could be who write this report, or it could be another team member who was involved with the survey. But if I remember correctly that time here, it was me.* » (S2).

Rappelons que dans le cadre de l'enquête de 'S2' auprès des bénéficiaires du programme de financement, des données déclaratives sur le nombre d'heures de bénévolat (et nombre d'heures par bénévole) rapporté par chaque projet financé ont été recueillies. Aussi, celles-ci ont été retranscrites dans la base de données de rendement, qui contenait des données financières, afin de procéder à l'analyse de l'efficience (analyse coûts-efficacité). Ainsi, 'S2' nous explique qu'à l'appui d'une théorie économique sur l'estimation du coût du bénévolat, il a converti le nombre

d'heures de bénévolat en valeur monétaire. Pour lui, cet exercice a permis de mettre en exergue des résultats intéressants, notamment lors du calcul du ratio financier visant à rendre compte de la capacité du programme à se financer par d'autres sources de financement (*calcul du rapport 1 CAD dépensé par le programme/montant provenant de financements autres que le gouvernement fédéral*). En effet, en attribuant un coût de main-d'œuvre aux heures de bénévolat, il a pu mettre en lumière la forte contribution financière de l'engagement communautaire, celles-ci représentant presque 50 % des sources de financement autres, le reste provenant de contributions monétaires. 'S2' mentionne qu'il est satisfait d'avoir fait cet exercice, celui-ci ayant permis de rendre compte du coût réel des activités entreprises, de la production des extrants, et des économies réalisées par le programme.

Extrait de verbatim

« *Parce que l'évaluation s'est intéressée aussi, entre autres, à quelle est la valeur de ce qu'ils font ? Et puis je l'ai quantifié, entre autres, en regardant notamment combien d'heures de bénévolat sont allées dans chaque projet financé par le programme, et puis en utilisant une théorie économique qui est de... quel est le montant que les gens sont prêts à payer pour un service. Ça, c'est une théorie économique. Tu veux savoir combien coûte une thèse de doctorat, combien tu es prêt à payer pour une thèse de doctorat. C'est ça que ça vaut. Donc ce que j'ai fait... parce que je me suis rendu compte qu'il y a beaucoup de projets communautaires où il y a beaucoup de bénévoles qui participent donc j'ai compté le nombre de bénévoles, puis le nombre d'heures, et j'ai quantifié la contribution de la communauté et c'est comme ça que j'ai donné une valeur monétaire à ce qu'ils faisaient. [...]. Ils sont prêts à donner combien pour avoir des arbres dans leurs cours d'école. Ça donne des résultats forts intéressants d'ailleurs. Et puis, je suis content d'avoir fait cet effort-là, cet exercice-là qui demeure très rare pour ce genre d'évaluation. La politique évaluative voulait qu'on donne une idée du rendement et des bénéfices dans le cadre de l'analyse de l'efficience, coûts-efficacité. [...]. On veut voir si les activités entreprises et les produits livrés dans le cadre du programme le sont-ils au plus bas coût possible. Alors on regarde notamment les leviers financiers, avec ici par exemple la capacité du programme à se financer par d'autres sources de financement. Alors en regardant les données, en attribuant un coût de main-d'œuvre au nombre d'heures considérables de bénévolat, on s'aperçoit que les autres sources de financement sont constituées de presque 50 % de contribution en nature, et un peu plus de 50 % de contributions monétaires autres que le gouvernement fédéral. Alors oui, si on regarde le ratio financier montant dépensé par le programme/montant de contributions financières autres on a déjà un peu plus de 1,50 CAD provenant de contributions monétaires autres pour 1 CAD dépensé par le programme. Mais en incluant la valeur monétaire de la contribution bénévole, on passe à presque 3 CAD provenant de sources de financement autres pour 1 CAD dépensé par le programme. Et ça, ce n'est pas rien ! C'est important de rendre compte du coût réel de ce qu'ils font, et puis ben de donner une estimation de la valeur monétaire des économies réalisées par le programme grâce à la contribution de la communauté.* » (S2).

Toujours dans le cas de 'S2', nous avons vu que son équipe avait dû créer la base de données de rendement à partir des données recueillies sur les projets financés par le système d'information de gestion du programme, et ce, en vue de réaliser les analyses nécessaires pour générer un rapport de rendement à l'échelle du programme, de rendre compte de l'atteinte des résultats à l'échelle du programme. Les procédures d'analyse entreprises ont été déterminées par

la stratégie de mesure du rendement du programme. ‘S2’ ne précise pas s’il a recouru à un logiciel d’analyse de données quantitatives ou s’il a effectué les analyses statistiques descriptives directement dans le fichier de la base de données (feuilles de calcul électronique). Il explique qu’il s’agissait de simples calculs de sommes, et de différences afin de calculer le pourcentage d’atteinte d’objectif pour chaque indicateur de mesure de rendement du programme.

Extrait de verbatim

« *[t]out est basé sur le modèle, la stratégie de mesure de rendement du programme. On avait les séries de données pour chaque projet, il fallait qu’on les collige, qu’on fasse les analyses pour rendre compte du rendement à l’échelle du programme, du résultat atteint pour chaque indicateur de mesure du rendement. [...]. C’est tout simple, pour chaque indicateur, on a fait la somme des valeurs cibles, la valeur cible c’est le résultat visé de chaque projet, la somme des valeurs réelles, et puis on a fait une comparaison entre le rendement visé et le rendement actuel et donne le résultat en pourcentage d’atteinte de l’objectif.* » (S2).

Dans le cadre de la discussion sur l’examen des données provenant du système d’information de gestion des dossiers traités par la branche évaluée, ‘S1’ mentionne que son équipe était composée d’un statisticien qui est très compétent pour manipuler les données dans les fichiers électroniques de base de données (feuilles de calcul). Si on ne sait pas dans quelle mesure les données ont été remaniées ou les analyses statistiques retravaillées, il semblerait qu’il ait au moins réorganisé les tableaux de données afin notamment de générer différents graphiques visant à illustrer la répartition des ressources humaines selon les types de dossiers traités, etc. dans le cadre de la formulation des réponses à des questions sur l’efficience.

Enfin, sur le plan de la synthèse et de la communication des résultats, nous venons de voir que différents types de rapports techniques ou matrices de réponse aux questions d’évaluation peuvent être produits afin de rendre compte des résultats d’analyse des données provenant des différentes méthodes utilisées pour recueillir de l’information, et de faciliter le processus de rédaction de la sélection résultats et interprétation du rapport d’évaluation final. Dans les cas de ‘S2’ et ‘S3’, ces derniers ont produit un rapport technique par source de données, avec par exemple un rapport où on répond aux questions d’évaluation seulement à partir des résultats d’analyse des données d’entrevue, en fonction des différents groupes d’informateurs clés. En ce qui a trait aux études de cas, celles-ci font l’objet d’un rapport narratif à partir des informations, des données recueillies par le biais de différentes méthodes (examen de la documentation et des données administratives, et entrevues), rapport dont le contenu est révisé par le gestionnaire du service ou projet faisant l’objet de l’étude de cas. Aussi, ‘S3’ rapporte que l’ampleur des données collectées dans le cadre des études de cas peut représenter un défi au moment de la synthèse. Il revient sur

une tendance passée de rédiger des rapports de cas, très denses, très détaillés, qui était finalement lu par deux personnes, et retranscrit en deux lignes dans le rapport d'évaluation. Il précise que les rapports s'avéraient plus bénéfiques pour les évaluateurs, pour leur compréhension des mécanismes du programme, de ses activités concrètes sur le terrain. Tels quels, ces rapports décrivaient aux parties prenantes clés, plus particulièrement aux gestionnaires, des éléments clés de leur programme qu'ils connaissaient déjà très bien. 'S3' nous explique donc qu'aujourd'hui il vise la rédaction de rapports de cas courts (trois pages), privilégiant la rédaction des résultats en point de forme, plutôt que la narration, et se concentrant davantage sur les éléments de résultats essentiels qui contribuent aux réponses apportées aux questions d'évaluation. On touche ici à la question du rapport coût-bénéfice par rapport à l'effort fourni dans la production des rapports de cas et l'utilisation réelle des résultats des études de cas dans la formulation des réponses aux questions d'évaluation. En effet, nous avons vu dans le cadre de la discussion sur les méthodes de collecte de données que l'étude de cas a une visée davantage illustrative comparativement aux autres sources de données. D'ailleurs, les évaluateurs séniors ont tous mentionné à un moment ou un autre que les études de cas ne constituent pas une grande contribution à la méthodologie évaluative, certains allant même jusqu'à remettre en question leur utilisation (choix de méthodes standards). Malgré tout, 'S3' revient sur une contribution « inhabituellement » importante des études de cas, du fait de la complexité du programme et de la nécessité d'optimiser aux mieux le budget de la collecte de données, misant ainsi sur la complémentarité des méthodes pour combler les lacunes sur le plan de la collecte des informations essentielles pour veiller à la représentativité des résultats d'évaluation. Il explique qu'en dépit de son désir de produire des rapports de cas plus succincts, ces derniers faisaient entre sept et dix pages, et que malgré le rôle important des études de cas, leur contenu a été réduit à quelques lignes ici et là dans le rapport d'évaluation final. Il précise en revanche que les résultats d'un des cas furent essentiels dans la rédaction de certaines recommandations (en lien avec les difficultés rencontrées au niveau de la délivrance d'un des services du programme).

Dans le cas de 'S1', ce dernier n'a produit qu'une seule matrice de résultats, organisée par sources de données. Une matrice unique permet de mettre en exergue au fur et à mesure de sa complétion la façon dont les preuves provenant de différentes sources de données triangulent. Aussi, si a priori la synthèse des résultats des différentes études de cas peut paraitre compliquée, il explique qu'en raison de leur caractère davantage illustratif, il se concentre sur l'intégration des

résultats qui confirment ou viennent enrichir certains éléments de résultats. Il précise que compte tenu du nombre restreint de cas étudiés, les études de cas ont une utilité limitée et demeurent de faibles sources de données.

Extraits de verbatim

"The case studies are a different kind of product where you review a bit of documentation and do a small number of interviews. And then you put together a little narrative, kind of a story of what you found, and that product is then reviewed by the key point person from the case like a project manager, someone in charge. They review it and make a few comments, changes, and such. [...]. Compiling everything is indeed challenging and we tended to go overboard with case study narratives. They would get these 15 pages document of this really focused, super detailed about some aspect of the program, or projects that they founded. And it's interesting, it's good for the evaluators to understand how the program operates on the ground but at the end of the day you have this beautiful narrative that only two people are going to read and ends up being two lines in the evaluation report. So, you have to balance that out a little bit. [...]. It also depends on the situation. Usually, case studies are more illustrative than anything and don't necessarily contribute a lot to the evaluation. But here they were actually quite important, and the narrative itself was still probably seven or ten pages of which that was fairly distilled into a couple of lines in the reports. Having said that, one of the case studies we did on one area of a specific service that the program was struggling with really informed the recommendations we made to the program. So, it did make an important contribution to the evaluation. But yes... More and more we adopt I find a more focused kind of approached and lighter on the reporting, trying not to sort of put it into a lovely prose kind of format. Just three pages, bullet points, what is the meat of the study." (S3).

"*Again really with the completion of that one technical report, kind of the results matrix. It really is now looking at the three case studies and the information and coming up with that result. Now the results matrix is good just because you see how each line of evidence is addressing each evaluation question. So as you're doing it you're starting to see how the evidence is triangulating. So I it really wasn't that difficult to fit it in because again case studies are usually so illustrative that they are, you know, they are really only use to show if they can confirm certain things or illustrate certain things. They're never going to be able to... They're never going to be a strong line of evidence for showing that the other lines of evidence are wrong because you're only looking at three cases out of all of these cases. So they really have kind of limited use at the end of the day. [...]. It really wasn't that much information for this one because the file review template was pretty short and then it was just 11 interviews across the three case studies, so it wasn't a lot of information for these. So it really is the kind of thing where you have you know some key informants and people in the survey saying that there was good collaboration on high-risk cases between the branch department and the client department, and you have a case study, about one high-risk case that says they handled it by having biweekly to discuss strategy, etc. and so you would just use it as an example"* (S1).

Sur le plan de la synthèse et de l'utilisation des résultats provenant des différentes sources de données, 'S2' explique que chaque méthode à sa force, et que selon la question ou l'objet d'évaluation traité, les résultats d'une source de données prend généralement l'ascendance sur les autres. Il remarque que certains évaluateurs préfèrent ou choisissent de répondre à chacune des questions d'évaluation à l'appui des preuves provenant de toutes les sources de données, mais que selon lui ce n'est pas la meilleure manière de procéder. En outre, s'il n'écarte pas cette possibilité lorsque la situation l'exige, il n'en demeure pas moins qu'une, voire deux sources de données s'avèreront toujours plus forte(s), plus pertinente(s). Il aborde par exemple les questions sur le

rendement du programme qui nécessitent des preuves tangibles provenant majoritairement du recueil de données secondaires, de données administratives ainsi que de données financières (c.-à-d. d'examen de sources documentaires) ainsi que des résultats d'enquête (dans un deuxième temps).

Extrait de verbatim

« *Sur le plan des résultats, des sources de données qu'on va utiliser dans le cadre de la formulation des réponses aux questions d'évaluation, aux différents objets d'évaluation, chaque méthode à ça force, et selon ce dont il est question, une source de données va prendre plus d'importance. Pour chaque question on va d'abord mettre de l'avant les résultats de la source de données qui a le plus de force. Ici par exemple, grâce au système d'information de gestion et de la qualité des indicateurs de mesure de rendement, on a pu obtenir une solide base de données pour répondre aux questions sur l'atteinte des résultats du programme. On avait également l'enquête auprès des bénéficiaires du programme de financement. Alors sur le plan de l'évaluation du rendement-efficacité du programme, on n'avait guère besoin de plus. À l'inverse, l'enquête n'avait aucune utilité pour les questions sur la pertinence. [...]. Alors oui, certains évaluateurs aiment utiliser les résultats de toutes les sources de données pour répondre à chaque question, mais je ne trouve pas que ce soit la meilleure façon de s'y prendre. Bien entendu il se peut que je fasse de même lorsque j'aborde certains points particuliers qui le nécessitent, mais en fin de compte, il y a toujours une ou deux sources qui vont prendre le dessus. Et quand on regarde le rendement du programme, c'est ce qui vient des sources documentaires, les données administratives, les données secondaires sur le rendement extraites de rapports produits par le programme s'il n'y a pas de système de suivi, par exemple, et puis ben les données financières. »* (S2).

En ce qui concerne 'S4', ce dernier n'a quant à lui produit aucun rapport technique. En effet, celui-ci nous explique qu'il synthétise l'ensemble des résultats au moment de la rédaction de la section résultats du rapport d'évaluation final directement depuis le dossier d'analyse dans le logiciel de traitement de traitement de données qualitatives. Il précise que le choix de produire ou non un rapport technique par source de données dépend non seulement des contraintes temporelles et budgétaires, mais aussi des préférences de l'agence fédérale commanditaire. Selon lui ce type de rapport n'est pas d'un grand intérêt, sauf pour les évaluateurs de programme. Aussi, lorsque nous avons souligné le caractère très synthétique sur le plan de la communication (utilisation) des résultats provenant des différentes méthodes comparativement à l'ampleur des données collectées et des résultats obtenus, celui-ci nous indique que dans le contexte fédéral, les clients ne souhaitent pas nécessairement passer à travers tout le processus d'investigation et veulent directement voir l'interprétation, le jugement des évaluateurs.

Extrait de verbatim

« *J'ai utilisé le logiciel [QDA] pour tout, toutes les données – les documents, le fichier de base de données, les notes d'entrevues, etc. L'arbre de codage, c'est la matrice évaluative. Pour chaque indicateur, j'ai les données qui sont organisées par source de données, sources d'information – groupes d'informateurs clés, projets financés, cas étudiés, etc. [...]. On ne fait pas toujours un rapport d'analyse pour chaque méthode, par exemple juste sur les entrevues, où on ne répond aux questions d'évaluation qu'à partir des résultats d'analyse des entrevues. Ou juste sur l'examen des données admin. Ici, on était beaucoup trop restreint en termes de temps, de budget, et puis on n'avait pas tant de données que ça. Et puis, on va se le dire, honnêtement, ces rapports-là, ça n'intéresse que les évaluateurs. Il n'y a pas un gestionnaire qui va mettre le nez là-dedans. C'est d'un ennui ! Mais oui, c'est nécessaire quand même. Et puis ce qui est intéressant c'est quand on fait l'effort de présenter davantage de résultats d'analyse provenant des différentes sources de données dans le rapport final... et je parle vraiment au fédéral là, où les gens nous ont eu dit – on ne veut pas que vous nous preniez par la main à travers tous les résultats, on veut savoir, une fois que vous avez tout ça qu'est-ce que vous comprenez. [...]. Ici, j'ai synthétisé directement l'ensemble de tous les résultats quand j'ai rédigé la section sur les findings. Par exemple, pour répondre à la question sur la pertinence, et ben dans le logiciel [QDA], j'ai regardé les indicateurs correspondants, je suis passé à travers les différentes sources de données, et j'ai écrit. Alors oui, dans la présentation des findings on indique les sources, les preuves, sur quoi s'appuient nos constatations, on s'appuie sur plusieurs sources, mais encore là c'est très synthétique nos résultats. On va plus en profondeur dans les données quand c'est vraiment pertinent. »* (S4).

En outre, sur le plan de la communication des résultats, 'S4' réitère à plusieurs reprises que c'est une question de culture, qu'il y a des clients qui désirent voir plus de résultats d'analyse dans le rapport final, ou qui désirent des rapports techniques, et d'autres qui préfèrent que les ressources allouées soient utilisées pour d'autres choses. Il reconnait que cela peut paraitre un peu « *cow-boy* » comme manière de procéder, mais que la confiance en la crédibilité des résultats d'évaluation repose davantage sur la rigueur du travail de conception, de communication et de validation de la méthodologie évaluative en amont (c.-à-d. matrice évaluative, rapport méthodologique, instrumentation). Pour lui, c'est pour cette raison que c'est un évaluateur sénior, celui chargé du projet, qui rédige le rapport final d'évaluation. Il précise en revanche que dans tous les cas, les clients ont accès aux fichiers d'analyse de données du logiciel QDA, s'ils veulent consulter les données d'entrevues, par exemple. D'ailleurs, l'ensemble des évaluateurs experts rapporte qu'à la fin du projet d'évaluation ils transmettent aux commanditaires l'ensemble des documents de travail anonymisés tels que les notes d'entrevues et les données d'enquêtes. L'équipe évaluative externe fait également une présentation préliminaire des résultats auprès du comité d'évaluation interne, avant la rédaction du rapport final, à des fins notamment de validation.

Extraits de verbatim

« Ils veulent notre jugement, notre opinion, ils ne sont pas intéressés de parcourir le processus de recherche avec nous. Et puis, quoiqu'il arrive, on partage nos fichiers [QDA], ils peuvent aller voir ce qui s'est dit durant les entrevues, par exemple. [...]. C'est vraiment une question de culture, ça dépend vraiment des ministères, certains veulent plus de détails sur les résultats d'analyse, les données, dans le rapport final, certains veulent des rapports techniques, alors que pour d'autres ce n'est pas leur truc, ils préfèrent que les ressources soient allouées ailleurs, et ce, peu importe le budget. J'avoue ça fait un peu cowboy. [...]. Au final, la crédibilité des résultats, des conclusions, mais aussi de notre travail d'évaluateur, c'est tout le processus en amont, la conception de la matrice, la définition des indicateurs, la métho, la rigueur des procédures et leurs descriptions, l'instrumentation. Tout ça, c'est tout le contenu du rapport méthodologique que les parties prenantes clés ont été invitées à consulter, et puis qui a été validé par le chef de l'évaluation. Ça, c'est essentiel, parce que quand on arrive à la fin, il faut qu'ils nous croient, qu'ils aient confiance en notre jugement. Et puis c'est pour cette raison que c'est moi, l'évaluateur sénior responsable du projet qui rédige le rapport final. C'est une tâche essentielle complexe, qui nécessite de manier beaucoup d'informations, et ça demande de l'expertise. » (S4).

"*There are these technical reports that circulate, and the preliminary presentation of the findings. It's a sort of validation process. It enables the people from the program to kind of like say things like we can see why you would say that but one thing you might want to look at is this other thing. It triggers a little bit more in terms of double-checking or something that could be added to our analysis, etc. They might see something we don't, they can have a better understanding on some things, etc. [...]. And then at the end of the evaluation project, we would transfer the working document to the key point of contact, the person in charge of the internal evaluation committee, from the evaluation division. So, we transfer the data file from the survey so if they want to analyze it, they can. Same thing with the record of the interview notes, those are all in spreadsheets, so we strip them of all identifying information and then we send them over. It's important for them because they go through this approval process and six months down the line someone asks a question and you know they can go back to the technical reports, and the working documents, and they have some evidence, where findings come from, etc.*" (S3).

Enfin, ***le rapport final d'évaluation fait l'objet de plusieurs cycles de révision*** entre les gestionnaires de l'évaluation (interne et externe). Comme l'explique 'S2', l'évaluateur interne responsable du projet d'évaluation est chargé notamment de vérifier le contenu du rapport d'évaluation, d'assurer la qualité et la précision-justesse des informations rapportées, etc. Il facilite également la circulation du rapport et le recueil de commentaires et suggestions auprès des membres du comité d'évaluation interne afin de permettre à l'évaluateur sénior responsable de l'équipe évaluative externe d'apporter les modifications nécessaires (et pertinentes) au rapport d'évaluation finale.

Extrait de verbatim

"On rédige les résultats, on répond aux questions d'évaluation, etc. Enfin je rédige donc le rapport final, et puis l'équipe évaluative interne va tout passer au peigne fin, évaluer la qualité de toutes les informations, la précision. Et puis, elle communique le rapport aux membres du comité d'évaluation interne et recueil, compile leurs commentaires, suggestions, etc. et nous les faits parvenir, et je retravaille le contenu du rapport. Il y a plusieurs allers-retours comme ça. C'est un travail collaboratif. Et puis, si je ne suis pas nécessairement d'accord, on en discute, on trouve un compromis ou tout simplement ils me font confiance, ils savent que mon intention c'est d'assurer la qualité de l'évaluation, la représentativité des résultats, et surtout l'utilité de l'évaluation. Alors oui bien sûr je les écoute, c'est leur programme, il y a des nuances qu'ils saisissent beaucoup mieux que nous, etc. Mais, il y a aussi des risques qu'à force de reformuler, de réarranger le contenu on peut perdre un peu de précision, de justesse en cours de route. Et puis, comme je te l'ai déjà expliqué, même après qu'on ait fini de travailler sur le rapport, il continue d'être révisé à l'interne, c'est tout un processus, et puis ben là je ne vois pas la version finale avant que celle-ci soit publiée. "(S2).

Par ailleurs, le cycle de révision et d'approbation du rapport ne s'arrête pas au moment où les évaluateurs séniors voient la dernière version qu'ils ont soumise acceptée par le client. En effet, 'S3' rapporte que le chef de l'évaluation interne prend le relais pour les dernières étapes de finalisation du rapport, un processus qui peut prendre au moins six mois. Il nous explique que durant cette période il arrive qu'ils soient sollicités par l'évaluateur interne à des fins de conseils, etc., mais qu'en revanche ils n'ont aucun droit de regard (ou de veto) sur le contenu de la version finale du rapport final rendu accessible au public (publiée en ligne), le rapport d'évaluation étant par ailleurs un produit de la couronne. L'évaluateur sénior reconnait d'ailleurs se sentir nerveux, les choses n'étant pas très clairs en ce a trait à la responsabilité des évaluateurs contractants vis-à-vis de ce qui est officiellement rapporté au sujet de l'évaluation. Dans le cas présent, il évoque avoir été quelque peu préoccupé par la possibilité que le caractère négatif de certains résultats d'évaluation qu'il a rapporté soit atténué pendant le dernier cycle de révision/finalisation du rapport à l'interne. En outre, il remarque que la dernière version qu'il a soumise était proche du produit final, et tenait compte des différentes rétroactions émises les pendant les premières relectures par le comité d'évaluation interne. Que les potentielles modifications apportées par la suite à l'interne ont trait à des éléments omis ou de nouvelles informations qui ont émergé. Pour 'S3', cette manière de procéder le renvoi à son statut d'évaluateur « *intérim* », et au fait qu'il est seulement attendu de lui qu'il produise un rapport de qualité, dont le processus de finalisation et de préparation pour publication s'avère en dehors de ses prérogatives. Si c'est quelque chose qui le préoccupe, il reconnait ne pas avoir la solution, puisque ce n'est pas non plus viable/rentable pour les évaluateurs contractants de s'éterniser sur un projet pendant plus de six mois après sa complétion. Ultimement, il s'avère essentiel selon lui de croire en l'intégrité de ses pairs, de ses clients.

Extrait de verbatim

"I understood that there had been a couple of additional drafts still circulating internally to committee members, and occasionally, that person who was our key collaborator from the evaluation division, would send me an email word or pick up the phone and say OK so there's a question about this I think the answer is that ... what do you think? So, they sort of use our firm almost like a sounding board kind of moving forward to finalizing the draft on their own. [...]. It's becoming more normal that in contracts what they want from the supplier is a couple of solid draft when the report went through multiple revisions, you responded to their comments. I think they let us go at this stage because the report goes through these powerful processes internally for final revisions and approval that take so long, and no contractor wants to be hanging around for another six months. [...]. Once they have this solid draft, they take it from here, and after that there was no real oversight for us. It kind of enhanced our interim evaluator status. We don't give our blessing for the final version. [...]. It does make me feel nervous in a sense that these reports are being published online are accessible to the public, and where our responsibility ended it's not really that clear. Here, the results of this evaluation were a bit negative so it did concern me a little bit that there could be some softening of the findings along the way... But by the time we let it go we already had two or three rounds of comments, so the report was reasonably close to being finalized, and if there are corrections its factual ones that people didn't notice or new information has come to light. It's a bit of a conundrum, I don't know how to solve it really because like I said, from a supplier point of view to be involved for that length of time would be very difficult, with all the dangling projects. And at the end of the day under the contract what we produce is the product of the crown, so we don't really have the authority per se to sort of veto any changes or be a barrier for moving forward. But I haven't sort of encountered a situation where I would, never saw a report online and was like oh they changed everything or like a lot or something like that. You have to trust and believe in the integrity of your colleagues and your clients ... that the report won't go through some sort of reediting." (S3).

En qui a trait à **la formulation des recommandations**, nous avons remarqué que chacune d'entre elles comprenait une réponse de la direction. Les évaluateurs séniors nous précisent alors que c'est toujours inclus dans le contenu du rapport, mais qu'il s'agit d'un élément rajouté dans le cas du cycle de révision à l'interne, soit une fois que le rapport final soumis par l'équipe évaluative externe est accepté par la division interne de l'évaluation. En outre, 'S1' nous précise que ce processus va bien au-delà du comité d'évaluation interne, et engage de plus hautes sphères administratives. Par ailleurs, cette discussion a conduit l'évaluateur sénior à nous faire part du fait que dans la plupart des cas, l'évaluateur interne est responsable de la rédaction des recommandations ainsi que du recueil et de la formulation des réponses de la direction. Il nous explique que l'équipe évaluative externe est rarement consultée, et cela semble dépendre de l'évaluateur de la division interne avec qui il travaille. Aussi, ils sont susceptibles d'aborder les recommandations lorsqu'ils discutent des résultats d'évaluation ou il arrive que l'évaluateur interne contacte le gestionnaire de l'évaluation externe pour parler des recommandations auxquelles ils songent. Mais, ultimement, c'est la responsabilité de la division de l'évaluation interne, et les évaluateurs contractants ne sont guère consultés.

Extrait de verbatim

"*For the management responses to the recommendations, it's always included in the report. It's the internal evaluator who writes it, but it does not come from just the evaluation working group, it's well above that, or the branch, it's at a higher administrative level. Once our final report is accepted, the evaluation division takes it to a specific committee, and they review it. I don't know the name of the committee. It's a process that I don't really know much about. [...]*. [R – Do you discuss the recommendations or is it out of your hands at this point?]. *It depends on who you are working with within the evaluation branch division. Sometimes they will. Sometimes it's being pretty clear that you are just... that you almost kind of talk about the recommendations as you talk about the findings. So they don't... It's rare that they come to you to show you the wording of the recommendations they are going to use to see if you agree or not. I have had that happened but not all the time and not often really. Usually, they might have a discussion on the phone just to say these are the areas I'm thinking for recommendations, do you think I'm missing anything? That kind of thing. But it's really the evaluation division really owns the recommendations themselves. They don't run a lot by us usually."* (S1).

4.2.2.4 *Interprétation de la question tertiaire 2* – les pratiques réelles des évaluateurs experts de programmes publics

Question de recherche tertiaire (2) : Quelles sont les informations fournies par les évaluateurs, pour expliquer et décrire les choix méthodologiques opérés dans le cadre de leurs évaluations de programme?

S'agissant de l'évaluation de programmes gouvernementaux en réponse aux exigences de la Politique évaluative fédérale de 2009, les évaluateurs experts exercent dans le présent contexte, dans un environnement politique et socio-économique empreint de contraintes importantes. Aussi, en tant que gestionnaire de l'évaluation à l'externe, les pratiques réelles des évaluateurs experts semblent s'inscrire dans l'exécution de l'évaluation planifiée par les clients, plus précisément, par la division interne de l'évaluation de l'agence gouvernementale commanditaire. En effet, lors de la discussion sur le contenu du cadre d'évaluation, ces derniers nous ont précisé d'emblée que la demande de propositions contient l'ensemble des éléments clés du cahier des charges (portée de l'évaluation, questions d'évaluation, méthodes préconisées/attendues, sources d'information à consulter, budget, échéancier, etc.), et qu'en définitive leur pouvoir décisionnel s'avère assez restreint à ce niveau-ci de la démarche évaluative. Pour eux, leurs responsabilités semblent reposées davantage sur la conception du cadre méthodologique, la collecte et l'analyse des données ainsi que la synthèse et la communication des résultats, et la formulation des constatations évaluatives (production du rapport d'évaluation final). Il importe de préciser que si les évaluateurs experts sont chargés de l'évaluation à l'externe, l'évaluation est réalisée sous la direction du chef de l'évaluation interne (évaluateur de la division interne de l'évaluation de l'agence

gouvernementale commanditaire), en consultation avec un comité d'évaluation interne composé des utilisateurs ciblés de l'évaluation, à savoir les hauts responsables et gestionnaires du programme, et de l'agence gouvernementale financière.

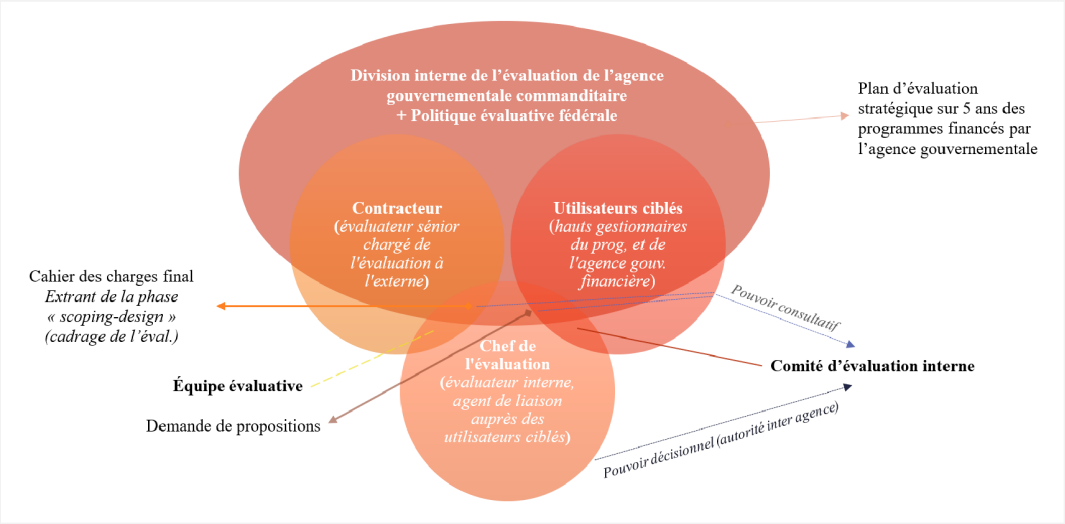


Figure 32. *Dynamique entre les différentes entités gouvernantes de l'évaluation*

Selon nous, en se percevant comme des « exécutants », les évaluateurs experts minimisent, à priori, leur rôle dans le cadre de la conception du plan d'évaluation. En effet, l'exercice d'explicitation des procédures de développement des questions et de la matrice d'évaluation a conduit les évaluateurs experts à reconnaitre le caractère plus ou moins générique des paramètres de l'évaluation prescrits dans la demande de propositions ainsi que la nécessité d'affiner, de compléter ou d'opérationnaliser le cadrage de l'évaluation. En outre, comme pour tout projet d'évaluation de programme, **les évaluateurs experts en tant que gestionnaire de l'évaluation externe entreprennent une phase de « *scoping-design* »**, autrement dit les étapes essentielles d'un processus itératif de description du programme, d'évaluation de l'évaluabilité/faisabilité, et de définition du cadre d'évaluation (McDavid, Huse et Hawthorn, 2019; Patton, 2013 ; Stufflebeam, 2004 ; Darabi, 2002). Le principal extrant de cette étape de conception de l'évaluation est la matrice évaluative, un outil crucial qui rend compte du plan d'évaluation ou cahier des charges (questions d'évaluation, indicateurs, sources d'information, méthodes), sur lequel repose le processus de sélection des sources d'information, la construction des instruments de collecte et d'analyse de données, le procédé de triangulation des sources de données, etc. Le modèle logique constitue également un produit essentiel de la phase de « *scoping-design* » sur lequel l'évaluateur sénior s'appuie pour définir les indicateurs, identifier les sources d'information, construire

l'instrumentation, etc., mais dans le présent contexte ce dernier a été fourni aux évaluateurs experts en amont/dans le cadre de la rencontre de travail initial avec les clients (première activité entreprise durant la phase de cadrage de l'évaluation).

Pour les évaluateurs experts, **la planification de la méthodologie évaluative est une entreprise cruciale tributaire du savoir expérientiel de l'évaluateur** – « *[...]. Pour moi la conception des questions d'évaluation, de la matrice, c'est beaucoup bâtie sur notre expérience. C'est de l'art plus qu'une science, ça exige une capacité d'analyse de la situation pointue. C'est pour ça que c'est toujours un évaluateur sénior qui a la charge du projet, qui est responsable du design. »* (S4) ; *« Quelqu'un de nouveau ne peut pas anticiper très bien selon les choix ce qui va se passer. Je les ai vus ces situations-là. J'ai vu suffisamment de situations pour anticiper, savoir quelle méthode va marcher, que si on s'y prend de telle manière, on n'obtiendra pas les données nécessaires pour répondre correctement à la question d'évaluation. »* (S2). Dans un contexte évaluatif hautement contraignant tel que celui à l'étude, il n'est pas surprenant d'observer que l'une des premières stratégies d'assurance de la fiabilité, la validité et de la crédibilité de l'évaluation citée par nos participants soit l'attribution à l'un des évaluateurs les « plus » de la firme-conseil le rôle de gestionnaire de l'évaluation externe incluant la responsabilité des activités pratiques et méthodologiques cruciales telles que la conception de l'évaluation et sa communication. En effet, la qualité de l'évaluation repose sur la capacité de l'évaluateur de programme à mobiliser des savoirs pratiques en situation réelle, ses valeurs, et son jugement qui lui permettent de prendre les décisions méthodologiques les plus appropriées selon les exigences/besoins des utilisateurs ciblés, l'accessibilité des données, les ressources disponibles et autres contraintes propres à la situation ou au milieu donné (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019 ; McPherson, 2015 ; Fitzpatrick, Christie et Mark, 2008 ; Greene, Lipsey, Schwandt, Smith et Tharp, 2007 ; Greene, 2006 ; King, 2003). Les connaissances et compétences du volet technique à elles seules ne suffisent pas, il s'agit davantage de la mobilisation des savoirs procéduraux et expérimentaux du volet *analyse situationnelle* (SCÉ, 2018 ; Stevahn, King, Ghere et Minnema, 2005). La qualité de l'activité des professionnels de l'évaluation repose ici sur la mobilisation de savoirs tant sur le plan méthodologique, que sur le plan des programmes publics, du système organisationnel fédéral, de la politique évaluative, et des logiques des acteurs publics, de gouvernance de l'évaluation externe de programmes publics, etc. Sur la combinaison de savoirs théoriques et procéduraux, de connaissances spécifiques à l'environnement professionnel, et de savoir-faire cognitifs acquis dans et avec l'expérience,

nécessaire à l'analyse et à la résolution de problèmes, à la conceptualisation et à la réalisation de projet (McPherson, 2015 ; Le Boterf, 2011). L'évaluation est une entreprise à la fois technique et sociopolitique, et les choix de conception de l'évaluation (plan d'évaluation, sélection des sources d'information, instrumentation, etc.) reposent sur le diagnostic approfondi d'une situation évaluative donnée et ses trois dimensions clés : 1) l'état/les caractéristiques de l'objet évalué ou *evaluand*[26] (complexité de la théorie de programme, envergure des objectifs, le caractère facilement observable ou non des résultats, etc.) ; 2) les caractéristiques des stratégies des acteurs autour de l'évaluation (les attentes explicites et implicites, les attitudes envers l'évaluation, et le degré d'exigence méthodologique des utilisateurs ciblés, un élément souvent tacite, émanant d'habitudes, de modèles/présupposés épistémologiques propres à des individus, des lieux et des manières d'évaluer) ; et 3) les moyens méthodologiques employés – ressources disponibles (Tourmen et Doher, 2013 ; Tourmen, 2009, 2007). Une situation réelle que Bamberger et Mabry (2020) organisent en quatre catégories de contraintes méthodologiques (contraintes budgétaires, temporelles, d'accès aux données, politiques et organisationnelles), que l'évaluateur doit bien identifier, anticiper et gérer lors de la conception de l'évaluation ; mais aussi tout au long de sa mise en œuvre, le devis d'évaluation nécessitant souvent des réajustements au fil du déroulement de l'évaluation (Rossi, Freeman et Lipsey, 2019). **La pierre angulaire de l'activité de conception de l'évaluation menée par les évaluateurs experts participants sont les concepts pragmatiques caractéristiquement mobilisés par l'évaluateur expérimenté** pour analyser, interpréter une situation donnée et agir en conséquence. Des concepts-en-actes ou « savoirs tenus pour vrai sur le réel » relatifs à : 1) la *faisabilité* de l'évaluation – le caractère mesurable, observable, évaluable des questions d'évaluation ciblées à la lumière des caractéristiques de l'objet évalué, des moyens méthodologiques employés (pertinence des indicateurs et des sources de données sélectionnées) et des contraintes de temps, d'argent et d'accès aux données ; et 2) la

[26] *"Evaluand*, a generic term coined by Michael Scriven, may apply to any object of an evaluation. It may be a person, program, idea, policy, product, object, performance, or any other entity being evaluated." (Mathison, 2005, p.139-140).

légitimité de l'évaluation - le recours à une démarche méthodologique qui facilitera la production de conclusions utiles, solides, valides, fiables, crédibles (Tourmen, 2009, 2007). Le Tableau 33 illustre la démarche entreprise par les évaluateurs experts durant la conception du plan d'évaluation, qui visait à rendre opérationnel et évaluable/faisable les paramètres prescrits dans la demande de propositions (évaluation sommative de la pertinence et du rendement) :

Tableau 33. *La démarche de cadrage de l'évaluation entreprise par les évaluateurs séniors.*

Démarche de Conception/Affinage/Validation de la matrice éval. (Plan d'évaluation) (questions éval., indicateurs, sources d'information, méthodes)

	'S1'	**'S2'**	**'S3'**	**'S4'**
	•Firme [A] •Formation [M.A. SHS] [Juriste droit civ.] •Évaluateur qualifié [SCÉ] •Expérience pro. ÉP [16 ans]	•Firme [B] •Formation [PhD Sc.Écon] •Évaluateur qualifié [SCÉ] •Expérience pro. ÉP [16 ans] •Enseigne ÉP niveau univ.	•Firme [B] •Formation [M.A. SHS] •Évaluateur qualifié [SCÉ] •Expérience pro. ÉP [20 ans] •Conseil admin. asso. pro. ÉP	•Firme [A] •Formation [M.A. SHS] [Juriste droit civ.] •Évaluateur qualifié [SCÉ] •Expérience pro. ÉP [17 ans] •Ancien président asso. pro. ÉP
	<u>Paramètres de l'évaluation prescrits dans la demande de proposition</u>: portée de l'évaluation, questions d'évaluation, méthodes priviliégiées, sources d'information (identification des bases de données existantes, des informateurs clés internes et externes à interviewer ou sonder, du nombre d'entrevue juger nécessaire, du nombre de cas à étudier), répartition du budget, échéancier. Peut inclure matrice évaluative standard.			
Sources d'information	***Validation du contenu du plan d'évaluation*** • Le comité d'éval. interne (rencontre de travail initial). • La politique éval. fédérale et ses directives. • La documentation sur le projet antérieur de conception du plan d'éval. de la branche (projet mené par S1 - groupes de travail et *scoping interviews* avec gestionnaires de prog., analyse documentaire, etc.). • L'évaluateur interne (informateur clé et collaborateur)	***Analyse documentaire + Série d'entrevues préliminaires*** • Le comité d'éval. interne (rencontre de travail initial). • L'ensemble de la documentation pertinente sur le prog. • Les « *gens du programme* ». • Le client, c'est-à-dire l'évaluateur interne (informateur clé et collaborateur).	***Analyse documentaire + Série d'entrevues préliminaires*** • Le comité d'éval. interne (rencontre de travail initial). • L'ensemble de la documentation pertinente sur le prog. • Les gestionnaires de prog. clés. • L'évaluateur interne (informateur clé et collaborateur).	***Analyse du programme et de l'évaluabilité/faisabilité*** • Les utilisateurs ciblés/parties prenantes clés (rencontre de travail initial avec le comité d'éval. interne). • La politique éval. fédérale et ses directives (cadre de référence). • L'ensemble de la documentation pertinente (plan stratégique, description de prog., rapports éval. antérieures, modèle logique, etc.) • Le chef de l'évaluation interne (informateur clé et collaborateur).
Outils	*Extrant du projet de conception du plan d'éval. du service deux ans plus tôt.* • Modèle logique. • Matrice éval. • Cadre méthodologique générale.	• Modèle logique (déjà existant). • Socle d'indicateurs de rendement (déjà existant). • Matrice éval.	• Modèle logique (déjà existant). • Socle de résultats ciblés visant à éval. le rendement (*extrant phase scoping-design*). • Matrice éval. standard fournit par l'agence commanditaire.	• Modèle logique (déjà existant). • Socle d'indicateurs de rendement basé sur la stratégie de mesure du rendement (extrant *phase scoping-design*). • Matrice éval.
Objectifs	•Définir des questions d'évaluation qui répondent non seulement aux différents axes de questionnement fondamentales de la politique évaluative fédérale, mais aussi aux besoins d'informations des utilisateurs ciblés (hors réponse aux exigences politiques). •Analyser et comprendre en profondeur la théorie du programme, les enjeux de gouvernance, la dynamique entre ses différentes composantes, les liens « objectifs - extrants - résultats ciblés », etc.; identifier la stratégie de mesure de son rendement, identifier/définir les socles d'indicateurs de mesure de processus et de rendement, etc. •Formuler des questions/sous-questions d'évaluation claires et précises, dont le contenu est représentatif de l'objet évalué (lien entre les questions et les différentes composantes du modèle logique, lien entre les questions et les objectifs, les extrants, les résultats visés par le programme). •S'assurer que le devis d'évaluation permet de répondre de manière appropriée à chacune des questions d'évaluation ciblées, de recueillir des informations fiables et représentatives de l'objet d'évaluation. ◦Veiller à la pertinence des indicateurs sélectionnés afin de collecter des informations permettant de répondre aux questions d'évaluation. ▪L'assurance du lien "questions - indicateurs - objet évalué". ▪Le caractère mesurable/observable de chaque indicateur, c'est-à-dire la sélection d'indicateur pour lequel le type et la quantité d'information nécessaire pour le documenté est accessible, facilement obtenable. ◦Veiller à la pertinence des sources de données (alignement « objet évalué – questions d'évaluation – indicateurs – méthodes - sources d'information » ▪L'accessibilité des sources d'information (sources existantes et disponibles, considération éthique), la pertinence et la représentativité des échantillons - miser sur la complémentarité des méthodes pour renforcer/optimiser le processus de consultation des sources d'information clés. ▪Sélectionner les méthodes les plus pertinentes/appropriées pour accéder aux différents types et à la quantité d'information nécessaires pour répondre adéquatement à la question d'évaluation traitée ou l'indicateur ciblé, ce qui inclus automatiquement la triangulation des sources de données, et ce, afin de renforcer la fiabilité, la représentativité et la profondeur des informations recueillies (complémentarité des méthodes, importance de la combinaison de données quantitatives et de données qualitatives). ➧ Vérifier la pertinence et la faisabilité du plan d'évaluation à la lumière des ressources disponibles (échéancier, budget, accessibilité des informations), des contraintes politiques, de l'état de l'evaluand, etc.			

Remarquons que **les stratégies adaptatives d'ordre méthodologique entreprises par les évaluateurs experts (voir listes p. 317-321), en réponse à l'analyse de la situation** dont nous avons synthétisé et représenté les caractéristiques dans la Figure 33 (p. 317), s'inscrivent dans la lignée de celles proposées dans le cadre de l'approche de l'évaluation en situation réelle (Bamberger et Marbry, 2020). **L'une des stratégies globales mobilisées par les évaluateurs experts est l'ancrage de la méthodologie évaluative dans une approche multiméthode**. En effet, nous avons vu que les évaluateurs utilisaient plusieurs méthodes de collecte de données qualitatives et quantitatives, de manière plus ou moins séquentielle, et ce, pour un certain nombre de raisons caractéristiques de l'utilisation des méthodes mixtes en évaluation de programme, plus particulièrement l'évaluation du rendement – efficacité (Bamberger et Mabry, 2020 ; Greene, 2015, 2005c). Parmi lesdites raisons évoquées par les évaluateurs experts participants on trouve : 1) *la triangulation des résultats de l'évaluation* - augmenter la fiabilité et la représentativité des conclusions en comparant les informations provenant de différentes sources de données (s'appuyer sur un moins deux sources de données pour juger de la valeur de chaque indicateur clé) ; 2) *l'élaboration des procédures de collecte de l'information* - utiliser les données et résultats d'une méthode pour aider à élaborer l'échantillon ou les instruments pour une autre ; 3) *la complémentarité des méthodes* - augmenter l'exhaustivité, la précision et la profondeur des résultats de l'évaluation (construire/compléter la base de données sur le rendement du programme, renforcer la qualité de la couverture des composantes du programme, comprendre et décrire les processus menant aux résultats du programme et la manière dont ils sont affectés par le contexte de mis en œuvre, l'environnement du programme, etc.) ; 4) *la combinaison des perspectives* des différents groupes de parties prenantes clés du programme (haut gestionnaires, gestionnaires, employés, clients, bénéficiaires, etc.) vis-à-vis de l'objet évalué.

Les choix méthodologiques opérés par les évaluateurs experts s'appuient sur le recours à diverses stratégies de gestion pour pallier des contraintes situationnelles rencontrées (Figure 33). **Les listes suivantes offrent une synthèse des stratégies adaptatives mobilisées en réponse aux différents types de contraintes rapportées par les évaluateurs experts.** Précisons que beaucoup d'entre elles s'entrecroisent, selon la nature des contraintes traitées et, comme nous l'avons déjà mentionné, au regard d'un certain compromis nécessaire quant aux exigences attendues.

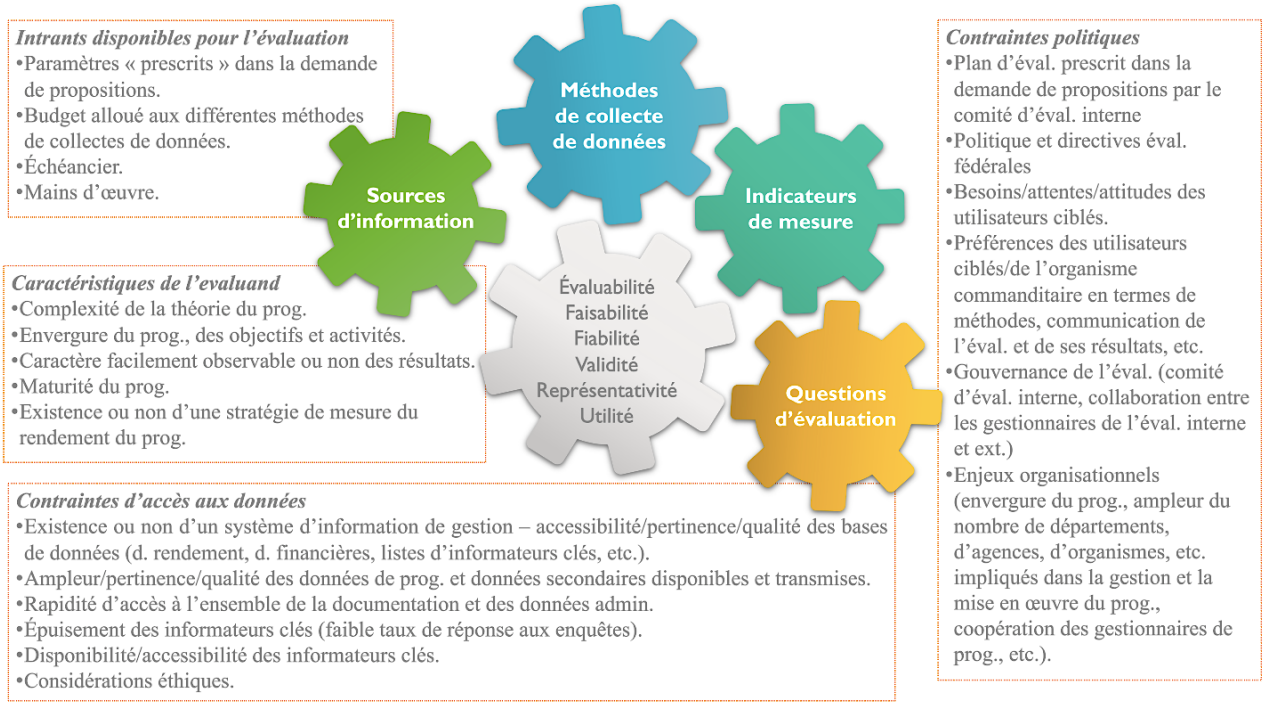


Figure 33. *Représentation graphique des caractéristiques des situations évaluatives rencontrées.*

- **Liste des stratégies de gestion des contraintes budgétaires et temporelles mobilisées par les évaluateurs experts**

Appréciation de l'utilité réelle, de la rentabilité des méthodes de collecte de données sélectionnées

- Miser sur l'utilisation de données existantes fiables — examen de bases de données administratives (données de performance, données financières, données sur la mobilisation des ressources, etc.), de la documentation de programme (rapports de rendement, projets financés, dossiers traités, rapports financiers, rapports d'évaluation, résultats d'enquête auprès d'informateurs clés tels que des clients, des employés), de rapports publics, etc. — 'S1', 'S2', 'S3', et 'S4'.
- Construire la base de données de rendement du programme à partir des séries de données rapportées par le système d'information de gestion pour chaque projet financé – très coûteux en termes de ressources, mais l'ampleur et l'exhaustivité des données basées sur de très bons indicateurs quantitatifs (valeur cible et valeur déclarée pour chaque résultat), permet de réaliser une évaluation de haute qualité de la performance du programme, des analyses significatives de l'efficacité, ce qui

n'est pas toujours possible dans le cadre d'évaluation de programmes gouvernementaux (importance d'assurer la démarche systématique et rigoureuse des analyses de données) – 'S2'.

- Exclure la recension des écrits, une méthode très appréciée des utilisateurs ciblés, mais très coûteuse en termes de ressources, pas toujours utile, et difficile à réaliser en bonne et due forme — 'S4'.
- Vérifier l'utilité des méthodes à la lumière des observations émises à la suite de l'utilisation du devis durant l'évaluation d'une autre branche de services similaires de l'agence commanditaire. Ici, retrait du groupe de discussion sur les résultats, face à la redondance des informations collectées dans le cadre des entrevues et de l'analyse des dossiers traités par la branche évaluée — 'S1'.
- Malgré son caractère idéal dans le contexte d'évaluation de programmes publics de grande envergure, limiter le recours à l'enquête en raison du faible taux de réponse généralement obtenu et privilégier si possible l'utilisation des données secondaires d'enquêtes pertinentes réalisées dans le cadre d'autres projets d'évaluation (beaucoup de programmes d'une agence donnée se recoupent en termes de groupes de répondants ciblés) — 'S4'.

Optimiser la consultation des informateurs clés, le budget « entrevue » étant souvent insuffisant face à la complexité/l'envergure du programme (renforcer la représentativité).

- Recourir à la passation de groupe lorsque les caractéristiques des interviewés le permettent – réserver les entrevues individuelles pour les hauts gestionnaires, et les informateurs clés indisponibles pour la passation de groupe – 'S1', 'S2', 'S3', et 'S4.
- Sur le plan de la sélection des informateurs clés, viser la représentativité en termes de qualité de couverture des composantes clés du programme, d'équilibre « informateurs clés internes-informateurs clés externes » ; et non en termes de quantité — 'S1', 'S2', 'S3', et 'S4.
- Miser sur la complémentarité des méthodes, utiliser les entrevues d'informateurs clés budgétés dans le cadre des études de cas — 'S3' et 'S4'.

- **Liste des stratégies de gestion des contraintes d'accès aux données mobilisées par les évaluateurs experts**

Effort de recueil/compilation de données probantes fiables pour évaluer le rendement du programme

- Existence d'un système de gestion de l'information du programme, des projets financés/dossiers traités, mais :
  - Absence de base de données de rendement à l'échelle du programme — construire la base de données à l'appui de la stratégie de mesure de rendement, et des séries de données rapportées par chaque projet financé (valeur cible et valeur déclarée pour chaque résultat/indicateur). Nécessite l'inclusion d'un technicien spécialiste en base de données au sein de l'équipe évaluative externe, et la collaboration étroite du gestionnaire de données interne — 'S2'.
  - Stratégie de mesure de rendement validée au début de l'évaluation, séries de données incomplètes/inconsistantes d'un projet financé à l'autre, et certains des indicateurs de la stratégie de mesure ne sont pas documentés – compléter la base de données, recueillir les données manquantes à partir de l'examen des projets financés afin d'apprécier le rendement global du programme (complémentarité des méthodes). Nécessite la collaboration avec les gestionnaires de programme pour déterminer la portée et la nature des informations disponibles — 'S3'.

- Existence d'une stratégie de mesure rendement, mais aucun système automatique de suivi, d'information de gestion – construire la base de données de rendement à partir des données extraites de la documentation de programme, des rapports produits par les services, des données secondaires d'enquêtes, rapports publics, etc. — 'S3'.
- Miser sur la complémentarité des méthodes pour compléter/corroborer/étayer les données sur le rendement du programme — 'S1', 'S2', 'S3' et 'S4'.

Considération éthique et déontologique

- Importance d'inclure l'examen des dossiers pour étayer les données administratives, et corroborer les pratiques déclarées par les informateurs clés (entrevues). Cependant, les services offerts sont protégés par une forme de secret professionnel – extraction des données à partir d'un gabarit conçu par l'équipe évaluative externe, réalisée par l'équipe de la division de l'évaluation interne et les professionnels de la branche ayant travaillé sur certains dossiers que la division ne peut pas consulter. Gabarit d'examen de dossier constitué majoritairement d'items fermés, dont les informations prodiguées sont ultimement vérifiables (copie physique des dossiers) — 'S1'.

Risque d'épuisement des informateurs clés

- Garder à l'esprit que dans le contexte de l'évaluation de programmes fédéraux, de nombreux projets d'évaluation ont souvent lieu plus ou moins en même temps, causant la sur sollicitation de certains groupes d'informateurs clés. L'un des phénomènes fréquemment observés est l'obtention de faibles taux de réponse aux enquêtes, une méthode privilégiée des clients, idéale au regard de l'envergure des programmes publics. – Recourir à l'enquête avec parcimonie, s'assurer de son utilité réelle, examiner s'il existe des données secondaires fiables, résultats d'autres enquêtes pertinentes qui pourraient être utilisés à la place – 'S4'.

- **Liste des stratégies de gestion des contraintes politiques et organisationnelles mobilisées pas les évaluateurs experts**

- Répondre aux attentes/exigences de l'organisme commanditaire – planification de la méthodologie évaluative ancrée dans les paramètres de l'évaluation prescrits dans la demande de propositions, et les directives de la politique évaluative fédérale.
  - Proposer des modifications sur le plan du contenu de cadre d'évaluation, lorsque jugé nécessaire, notamment à la lumière de l'analyse du programme, des enjeux organisationnels, des besoins et des préoccupations des utilisateurs ciblés.
    - Ajouter des questions d'évaluation de la conception/implantation du programme si ce dernier a fait l'objet de nombreux changements – 'S4'.
    - S'attarder sur les enjeux de gouvernance et leurs potentiels impacts sur l'efficacité d'implantation face à l'extrême remaniement au niveau de la direction du programme – 'S2'.
    - Élargir la portée de l'axe de questionnement sur les résultats pour renforcer la représentativité des conclusions. Par exemple, lors de l'évaluation de services de réglementation qui ont pour mandat de répondre à des priorités gouvernementales en perpétuelle évolution, ne pas se concentrer uniquement sur les résultats en fonction des jalons finaux (tels que la publication de règlements), rendre compte du travail réglementaire entrepris pour répondre au mandat gouvernemental évolutif, et des enjeux émergents – 'S3'.

- Tenir compte des préférences/attentes pratiques et méthodologiques des parties prenantes clés de l'évaluation – recourir aux méthodes « standards » de l'évaluation de programmes fédéraux, utiliser leurs gabarits/suivre leurs lignes directrices pour la communication de l'évaluation et de ses résultats, etc. Et ce, tout en veillant à la faisabilité méthodologique ainsi qu'à la fiabilité et représentativité des résultats – 'S1', 'S2', 'S3' et 'S4'.
- Anticiper le caractère laborieux et chronophage du processus d'obtention de bases de données complètes et fiables (données de rendement, données financières/ressources mobilisées, listes de projets financés, de contacts d'informateurs clés, etc.) et les difficultés méthodologiques que cela peut représenter (identification de stratégies d'obtention des informations manquantes, conception de l'instrumentation, planification de la triangulation des méthodes, échantillonnages, etc.).
  - Demander l'accès aux données administratives dès l'obtention du contrat d'évaluation afin de débuter leur examen le plus tôt possible, et miser sur la collaboration avec le chef de l'évaluation interne, les gestionnaires de programme et de système de monitorage – 'S1', 'S2' et 'S4'.
  - Demeurer vigilant face au potentiel manque de coopération des gestionnaires de programme, et de données internes, une forme de résistance souvent observée à l'interne – 'S2'.

Un autre élément stratégique essentielle mobilisé afin de mener à bien l'entreprise de conception et de mise en œuvre de l'évaluation dans le présent contexte : **la collaboration étroite, et la communication régulière (et efficace) entre les évaluateurs chargés du projet d'évaluation à l'externe et le gestionnaire de l'évaluation à l'interne (chef de l'évaluation, client, autorité décisionnelle)**. Pour les évaluateurs experts participants, ce travail en duo est une composante essentielle afin d'assurer la faisabilité, l'utilité et la qualité de l'évaluation. En effet, on a pu constater que le chef de l'évaluation constitue pour ces derniers une ressource indispensable pour mener à bien l'évaluation, du cadrage à la communication des résultats et constatations évaluative. Si ultimement le chef de l'évaluation interne vérifie et (ou) valide le contenu des différents livrables (rapport méthodologique [matrice évaluative, protocoles, instrumentation, etc.], rapports techniques [matrices de preuves/résultats], et rapport final [communication des résultats et des constats évaluatifs, articulation des recommandations]), les évaluateurs experts mettent davantage l'accent sur son rôle de personne-ressource clé, d'allié indispensable à l'interne. Dans le cadre de la conception de l'évaluation, ils le perçoivent comme un aiguilleur quant à la faisabilité, la pertinence, et la justesse des propositions méthodologiques (contenu de la matrice évaluative), un agent de liaison afin d'obtenir la rétroaction des parties prenantes clés du comité d'évaluation interne (hauts gestionnaires, utilisateurs ciblés), de recueillir des informations supplémentaires sur le programme, ou encore de faciliter directement l'accès aux gestionnaires de programme/de données, etc. De plus, comme ont pu le souligner 'S1' et 'S2',

l'évaluateur interne est également un garant de la qualité de l'évaluation, qui est notamment là pour s'assurer que le devis d'évaluation permet de répondre de manière appropriée aux questions fondamentales de la politique évaluative fédérale, et prévenir les risques de manipulation ou de détournement de l'évaluation. Nous verrons, un peu plus tard, que le manque d'expérience de l'évaluateur interne et son attitude envers le gestionnaire de l'évaluation externe (souvent 15 ans son sénior selon 'S2') peuvent s'avérer être une potentielle source de mésévaluation. Durant la phase de collecte de l'information, l'évaluateur interne facilite l'accès aux différentes sources d'information, plus particulièrement, il gère le recueil et la gestion de la documentation de programme à l'interne, il facilite le recrutement/l'accès aux informateurs clés internes, et dans certains cas comme celui de 'S1' ou l'équipe évaluative externe n'était pas autorisé à examiner les dossiers des affaires traités par la branche évaluée, l'équipe évaluative interne peut-être amener à conduire elle-même la collecte de l'information (en utilisant l'instrument conçu par l'évaluateur sénior chargé de l'évaluation à l'externe). Enfin, il collabore avec le gestionnaire de l'évaluation à l'externe dans le cycle de révision du contenu du rapport final, plus précisément la présentation et l'interprétation des résultats. Il facilite le recueil de la rétroaction des utilisateurs ciblés. Une fois le rapport accepté, il est responsable du processus de finalisation du rapport à l'interne, incluant la rédaction des recommandations, un exercice dans le cadre de laquelle il est susceptible de solliciter les conseils de l'évaluateur sénior externe, et l'obtention/formulation des réponses gestionnaires auxdites recommandations.

Précisons que si les utilisateurs ciblés membres du comité d'évaluation interne exercent quant à eux un pouvoir consultatif, il constitue une source d'information et de rétroaction essentielle dans la planification de la méthodologie évaluative (rencontre de travail initial, entretiens préliminaires, révision de la matrice évaluative, identification des informateurs clés et des cas à étudier) et présentation/interprétation des résultats (consultation des résultats préliminaires, ébauche du rapport final).

Comme vous pouvez le constater, **la communication émerge donc comme l'une des composantes méthodologiques essentielles**, et ce, à chacune des étapes clés de l'évaluation. Pour les évaluateurs experts participants, la faisabilité, la fiabilité, la validité, et la crédibilité de l'évaluation reposent sur la justesse et la transparence sur le plan de la description/explication, et de la documentation des procédures méthodologiques, de la conception du devis d'évaluation à la formulation des conclusions évaluatives. Il est question pour eux de soutenir et de démontrer

l'effort de rigueur, la pertinence et la qualité de la démarche évaluative. D'assurer la confiance des clients, des utilisateurs ciblés envers la démarche et les résultats d'évaluation, plus précisément le jugement de l'évaluateur. On couvre, ici, des préoccupations liées aux normes *raisonnement évaluatif explicite* et *communication et rapport* du standard précision-justesse (« *Accuracy* »), et documentation de l'évaluation du standard reddition de compte (JCSEE, 2011). L'outil privilégié de communication et de documentation des procédures est le rapport écrit, produit et transmis à des temps clés de la démarche évaluative. Il est question de différents types de rapport avec des fonctions bien précises :

Rapport méthodologique

Contenu entièrement conçu, rédigé par l'évaluateur sénior chargé de l'évaluation à l'externe

- Documente de manière extensive l'ensemble de la méthodologie évaluative – inclus la matrice évaluative, l'instrumentation, les critères de sélection des sources d'information et échantillons, les protocoles de passation, etc.
- Consulté par le comité d'évaluation interne, et validé par le chef de l'évaluation interne – principal outil de communication et de rétroaction entre les différentes entités gouvernantes de l'évaluation.

Rapport(s) technique(s) ou matrice(s) de preuves/résultats*

Contenu rédigé par différents membres de l'équipe évaluative externe sous la supervision de l'évaluateur sénior chargé de l'équipe, ou dans certains cas, rédigé entièrement par ce dernier.

- Extrant(s) des procédures d'analyse des informations provenant des différentes sources de données utilisées – matrice(s) de réponses aux questions d'évaluation dont le contenu est rédigé et (ou) révisé par l'évaluateur sénior charge du projet à l'externe.
- En fonction des habitudes/préférences des évaluateurs experts, des préférences des commanditaires, des contraintes budgétaires et temporelles - production d'un rapport technique par sources de données où les résultats/preuves sont classés par objet d'évaluation, questions et source d'information ; production d'une matrice de réponse aux questions d'évaluation unique dont le contenu est organisé par source de données. Dans certains cas, comme celui de 'S4', le rapport de codage/d'analyse des données exporté du logiciel QDA (incluant les données de rendement) constitue le rapport technique.
- Le recours au système de matrice de résultat,
  - Permet à l'évaluateur sénior comme aux commanditaires, plus particulièrement à l'évaluateur interne, de vérifier la précision-justesse du raisonnement évaluatif (alignement informations-résultats-interprétation-conclusion), que chaque question d'évaluation est bien traitée.
  - Facilite le processus de synthèse et de triangulation des résultats, la formulation des constatations évaluatives lors de la rédaction du rapport final d'évaluation par l'évaluateur sénior.

Rapport d'évaluation final*

Contenu rédigé par l'évaluateur sénior, qui est chargé de la synthèse/interprétation des résultats (formulation des constats évaluatifs) et du processus de révision des premières versions en collaboration avec l'évaluateur interne.

- Majorité du contenu centrée sur les besoins des utilisateurs ciblés – la présentation et l'interprétation des résultats, la formulation des conclusions évaluatives, et la proposition de recommandations.
- L'existence des rapports précédents est la raison pour laquelle le contenu du devis d'évaluation rapporté est lacunaire – la confiance des clients en ce qui a trait aux résultats émerge de la crédibilité des procédures méthodologiques, de rigueur de la démarche évaluative.
- Une fois que la première ébauche du rapport a été soumise, le rapport entre dans un cycle itératif de révision entre l'évaluateur sénior et le chef de l'évaluation interne, ce dernier facilitant l'obtention et la gestion de la rétroaction des utilisateurs ciblés.

* Notons que la circulation de rapports techniques et (ou) la présentation des résultats préliminaires auprès des utilisateurs ciblés sont des moyens mobilisés pour renforcer la fiabilité et la représentativité des résultats et conclusions évaluatives communiquées – « *it triggers a little bit more in term of double checking. The people from the program might see something we don't, can have a better understanding of some key elements, orient us toward other information, something that could be added to our analysis, etc.* » ('S3').

Enfin, remarquons qu'en dépit des efforts entrepris pour gérer aux mieux les contraintes inhérentes au terrain, et assurer la fiabilité, la représentativité et l'utilité de l'évaluation, des écueils méthodologiques perdurent. Aussi, l'analyse de ces écueils permet de mettre en exergue les écarts entre les pratiques recommandées et les pratiques réelles des évaluateurs de programme experts, et d'illustrer l'incidence des facteurs contextuels sur la qualité et l'utilité de l'évaluation ainsi que les risques de mésévaluation. Cela fait l'objet de la section suivante qui porte sur l'interprétation des questions secondaires de recherche.

4.3 *Interprétation des questions de recherche secondaires* – les différences et les ressemblances entre les pratiques réelles et les pratiques recommandées des évaluateurs de programmes publics et l'influence des facteurs contextuels sur lesdites pratiques.

Questions de recherche secondaires :

(1) Quelles sont les ressemblances et les différences entre les pratiques réelles et les pratiques recommandées des évaluateurs externes de programmes publics en ce qui a trait aux choix dits méthodologiques ?

(2) Quelles sont les perceptions des évaluateurs de programme externes quant aux possibles influences des facteurs contextuels sur les choix méthodologiques opérés durant leurs évaluations de programmes publics ?

La pratique des évaluateurs de programme, novices et experts, participants à cette recherche doctorale, est ancrée dans la réalisation d'évaluations externes de programmes publics du gouvernement fédéral, dans un environnement sociopolitique hautement contraignant, habité de contraintes caractéristiques de l'évaluation en situation réelle (Bamberger et Mabry, 2020). En outre, le principal objectif de la comparaison entre les pratiques réelles et les pratiques recommandées des évaluateurs de programmes publics participants à cette recherche doctorale était de mettre en exergue l'influence des facteurs contextuels sur les choix méthodologiques qu'ils opèrent.

Remarquons que les évaluateurs de programme dans le présent contexte exercent au sein d'une équipe évaluative externe où l'évaluateur sénior est chargé du projet d'évaluation, et l'évaluateur novice, selon son statut, peut être junior collaborateur, exécutant avancé, ou simple exécutant. Lorsqu'il exerce dans une capacité d'exécutant, l'évaluateur junior travaille en silo, et il n'est invité à joindre le projet d'évaluation qu'au moment de la collecte et (ou) de l'analyse des données. Lorsqu'il exerce en tant que collaborateur, sa pratique est ancrée en situation, ce dernier participant à l'ensemble des étapes du projet d'évaluation. À titre de collaborateur, l'évaluateur junior à l'occasion de participer aux discussions méthodologiques au moment de la phase de conception de l'évaluation, à la révision du contenu de l'instrumentation, à la coordination des activités de collecte et d'analyse de données, etc. Il peut aussi être amené à rédiger la première version de l'ébauche du rapport final sur les résultats et constatations évaluatives. Cependant, en

fin de compte, quel que soit le degré d'implication des évaluateurs juniors dans le projet d'évaluation, l'ensemble des décisions pratiques et méthodologiques (de la conception à la communication de l'évaluation) reposent sur les épaules de l'évaluateur sénior. **Nous ne sommes donc pas en mesure de répondre à cette question de recherche pour les évaluateurs novices.** Précisons, malgré tout, que les enjeux relatifs au travail en silo et au développement professionnel des jeunes évaluateurs dans un environnement empreint de contraintes sociopolitiques importantes constituent des préoccupations pratiques et méthodologiques importantes. Nous les aborderons dans le cadre l'interprétation de la question principale de recherche qui porte sur la distinction entre les pratiques des évaluateurs de programmes novices et les pratiques des évaluateurs de programmes experts (Chapitre 5, partie 1).

4.3.1 *Interprétation des questions de recherche secondaires* – le cas des évaluateurs de programmes publics « experts »

Qu'il s'agisse des pratiques recommandées ou des pratiques réelles des évaluateurs experts participants à cette recherche doctorale, celles-ci sont ancrées en situation réelle, dans la mobilisation de concepts-en-actes caractéristiques de la pratique des évaluateurs de programme expérimentés, de savoir-faire cognitifs engagés dans l'analyse d'une situation évaluative donnée afin de prendre les décisions méthodologiques les plus appropriées. Les caractéristiques des pratiques réelles mises en exergue dans le cadre de l'analyse des activités d'explicitation exemplifient les pratiques recommandées des évaluateurs experts participants. La priorité, veiller au caractère approprié et faisable du devis d'évaluation à la lumière des caractéristiques ou de l'état de l'*evaluand* (du programme, des différents objets évalués tels que l'atteinte des résultats, etc.) et des contraintes inhérentes au terrain (contraintes temporelles, budgétaires, politico-organisationnelles, d'accès aux données), tout en assurant la fiabilité, la représentativité, la crédibilité et l'utilité de l'évaluation (informations recueillies, résultats, constatations évaluatives).

En revanche, malgré les multiples efforts entrepris par les évaluateurs experts pour gérer le plus efficacement possible les contraintes inhérentes au terrain, il n'en demeure pas moins que des difficultés d'ordre pratiques et méthodologiques perdurent. Ces dernières mettent en exergue les différences entre les pratiques recommandées et les pratiques réelles des évaluateurs experts participants à cette recherche doctorale. Elles représentent également, selon nous, de potentielles

sources de *mésévaluation.* Rappelons que lorsqu'on parle de *mésévaluation*, il n'est pas question de situations dans le cadre desquels de petits écueils méthodologiques ont émergé, mais plutôt de situations où le caractère inadéquat d'une évaluation soulève des préoccupations majeures quant à la fiabilité et la validité du processus d'évaluation et de ses résultats (Alkin et King, 2017). En d'autres termes, il s'agit de situations ou la qualité et l'utilité de l'évaluation sont remises en question.

La plupart des écueils méthodologiques rapportés dans le cadre de la pratique des évaluateurs experts sont liés à : 1) l'accessibilité des données probantes nécessaires pour rendre compte de manière appropriée du rendement du programme, et la capacité des évaluateurs de programme à entreprendre les efforts nécessaires pour assurer la fiabilité et la crédibilité des résultats ; 2) l'obligation de répondre à l'ensemble des questions fondamentales de la politique évaluative fédérale sur le rendement du programme (efficacité, efficience et économie) en dépit des problèmes de faisabilité/évaluabilité ; et 3) la dynamique relationnelle entre les gestionnaires de l'évaluation (duo [évaluateur sénior chargé de l'évaluation à l'externe-chef de l'évaluation à l'interne]).

Les exigences de la politique évaluative fédérale s'inscrivent dans la lignée de la définition de l'évaluation des programmes du Conseil du Trésor du Canada, à savoir : « la collecte et l'analyse systématiques de données probantes sur les résultats des programmes afin d'en évaluer la pertinence et le rendement, et de trouver d'autres modes de prestations ou d'autres façons d'obtenir les mêmes résultats ». Afin de soutenir les efforts de gestion des programmes, et d'évaluation officielle de leur efficacité (périodique – tous les cinq ans), les directives de la politique évaluative fédérale de 2009 (celles applicables aux évaluations de programme qui ont fait l'objet des activités de collecte de données de cette thèse doctorale) ont renforcé les exigences selon lesquelles les gestionnaires de programme doivent recueillir de manière continue des informations sur le rendement de leurs programmes. Le produit direct de ces exigences : la stratégie de mesure de rendement, un outil de gestion axé sur les résultats qui vise notamment à assurer le suivi systématique des extrants et résultats des activités du programme ainsi que la production de séries d'information/données fiables et crédibles qui pourront être utilisées pour l'évaluation du rendement du programme.

Malgré tout, les problèmes bien connus (et reconnus) sur le plan de l'existence, l'accessibilité, la qualité et l'exhaustivité des données de mesure du rendement du programme comptent parmi les plus grosses sources de contraintes et d'écueils méthodologiques dans le cadre des expériences rapportées par les évaluateurs experts participants à cette recherche.

Il s'agit ici de problèmes qui représentent des risques de *mésévaluation*, selon l'ampleur de la situation et la capacité des évaluateurs de programme à la gérer le plus efficacement possible. On les qualifie de « reconnus » puisque les évaluateurs experts participants qualifient ces limites méthodologiques d'habituelles, de normales dans le présent contexte. En effet, malgré l'existence d'une mesure de rendement du programme, voire d'un système d'information de gestion et de sa banque de données, nous avons observé de nombreuses lacunes sur le plan des données de mesure de rendement (évaluations de programme discutées réalisées entre 2013 et 2016). Des difficultés qui sont liées à la nature incomplète, insuffisante ou non fiable des données, de la faiblesse ou du manque de pertinence de certains indicateurs de mesure. Des observations qui ne sont pas surprenantes compte tenu des constatations émises dans le cadre de l'évaluation de la Politique d'évaluation fédérale de 2009 (Secrétariat du Conseil du Trésor du Canada, 2015). En effets, l'évaluation de la Politique évaluative a montré que si la situation sur le plan des données de programme semble s'être améliorée dans une certaine mesure depuis la mise en œuvre de la Politique, parmi les organisations interrogées sur la période de 2012-2013, la moitié rapportait des données de mesure de rendement insuffisantes. De plus, si 90 % des programmes comportaient une stratégie de mesure de rendement, les évaluateurs avaient remis en question l'utilité d'environ la moitié des indicateurs de mesure. De même, les résultats d'enquête d'évaluation des capacités de 2013-2014 des fonctions de résultats du Secrétariat ont montré que : sur les 33 évaluations de programme approuvées entre avril et septembre 2013, 30 ont connu de gros écueils méthodologiques à cause de l'insuffisance des données de mesure de rendement, malgré un taux de 70 % déclaré en ce qui concerne la mise en œuvre de stratégies de mesure du rendement. Un problème par ailleurs signalé par un certain nombre d'agences gouvernementales lors de la soumission annuelle obligatoire par les services d'évaluation d'un rapport sur l'état de la mesure du rendement (ibid.). Par ailleurs, les lacunes sur le plan des données de programme qui doivent permettre de rendre compte du rendement rejoignent également les observations émises dans le rapport de mise au point sur l'évaluation de l'efficacité des programmes fédéraux (Bureau du vérificateur général du Canada, 2013).

Les différents problèmes rencontrés par les évaluateurs experts ont trait à la pertinence et la qualité du cadre de mesure de rendement du programme, plus précisément des indicateurs de mesure d'extrants et de résultats, du respect des exigences de rapports sur les indicateurs, et (ou) de la gestion des informations. Comme on le retrouve dans le rapport Bureau du vérificateur général du Canada (2013), selon la nature et l'ampleur des lacunes sur le plan des données de mesure de rendement, la détermination et (ou) la portée de l'évaluation des résultats du programme peuvent être plus ou moins limitées. Et, dans tous les cas, cela implique généralement la collecte de nouvelles données probantes, ou de miser sur l'obtention de données plus subjectives (questions de rappels, questions déclaratives dans une enquête ou une entrevue auprès d'informateurs clés, par exemple). Stratégies adaptatives de l'évaluation en situation réelle (Bamberger et Mabry, 2020). Notons que la question de la capacité des évaluateurs experts à pouvoir contrecarrer ces contraintes importantes d'accès aux données est tributaire des ressources disponibles pour l'évaluation, sur le plan des intrants (échéancier, budget, main-d'œuvre/technicien spécialiste), des sources d'informations pertinentes et accessibles, de la collaboration des gestionnaires de programme et de données internes.

Parmi les situations les plus potentiellement problématiques rencontrées par les évaluateurs experts, on rapporte celles où les évaluateurs ont dû : 1) créer la base de données de mesure du rendement, car il n'existe aucun système d'information de gestion ; ou 2) compléter une grande partie de la base de données du système d'information de gestion, car celui-ci n'a pas recueilli les données pour certains indicateurs de mesure de rendement (par exemple, lorsque la stratégie de mesure de rendement a été validée au début du projet d'évaluation). En d'autres termes, les séries de données probantes rapportées sont insuffisantes pour rendre compte de manière appropriée de l'avancement et de l'atteinte des résultats du programme.

L'ampleur de la documentation et des informations collectées/générées dans le cadre des activités de monitorage du rendement du programme est une contrainte importante sur le plan de la gestion efficace des données (renseignements administratifs, données sur les indicateurs de mesure de rendement, données secondaires, etc.), particulièrement en l'absence d'un système d'information de gestion. Ce système de suivi recueille des données sur les étapes clés du cycle de vie des projets financés, initiatives entreprises, dossiers/affaires traités dans le cadre des programmes, des services évalués (données administratives, renseignement sur les intrants, extrants produits et résultats atteints, etc.). Il a pour principal extrant une base de données

considérée comme la principale source de données pour générer des résultats globaux à l'échelle du programme, et rendre compte de manière appropriée de son efficacité. En l'absence d'une telle base de données, les évaluateurs de programme n'ont pas d'autres choix que rediriger les ressources disponibles pour la construire eux-mêmes, afin de veiller à ce que l'évaluation des résultats du programme repose sur le plus de données probantes possible. Précisons que l'absence d'un système d'information de gestion ne signifie pas nécessairement que l'ensemble des données sur les indicateurs de mesure de rendement n'existent pas. Particulièrement si le cadre de mesure de rendement comprend la production systématique de rapports d'avancements/de résultats selon les indicateurs, pour chaque initiative du programme, par exemple. Les choses se compliquent davantage lorsqu'il existe que des rapports de rendement global. Mais là encore, c'est possible que les données à l'échelle des différents projets entrepris dans le cadre du programme soient rapportées autrement (rapports publics, productions/publications scientifiques, base de données secondaire, etc.). Autrement dit, selon la nature et la portée des données disponibles et la facilité d'accès aux données probantes, une telle entreprise peut s'avérer plus ou moins chronophage, laborieuse et coûteuse. Et, dans certaines circonstances, les contraintes peuvent être trop importantes pour que les évaluateurs soient en mesure d'obtenir les données probantes nécessaires. Par exemple : les données de mesure de rendement des différents dossiers traités dans le cadre des services offerts par l'organisme évalué existent, mais elles sont transmises dans des rapports physiques (papiers). Selon l'ampleur des dossiers traités sur la période évaluée, récupérer l'ensemble de la documentation papier, recueillir et compiler l'ensemble des données, puis examiner la qualité et l'exhaustivité des données probantes tout en respectant le budget et l'échéancier de l'évaluation relève de l'impossible dans le présent contexte. En effet, les évaluateurs experts nous expliquent qu'ils ont généralement entre deux et trois mois pour collecter et analyser les données. De surcroit, selon l'ampleur des lacunes sur le plan des données, nous avons vu que les évaluateurs peuvent être amenés à devoir combler les données manquantes ou corroborer les informations dont la fiabilité est remise en question. Or, ce n'est pas garanti que les évaluateurs soient en mesure de la faire dans les temps. Malheureusement, même si ces problèmes de faisabilité/d'évaluabilité sont identifiés très tôt (dès la phase de *scoping-design*), et qu'ils nécessiteraient la modification du cadre d'évaluation, les évaluateurs de programme externe n'ont pas d'autre choix que de tout de même traiter l'ensemble des questions d'évaluation de rendement.

Nous revenons davantage sur ce point dans la prochaine section relative aux problèmes pratiques et méthodologiques liés à « l'obligation d'évaluation ».

Lorsque la base de données de mesure de rendement est inexistante ou incomplète, c'est essentiel que les évaluateurs de programme entreprennent le plus tôt possible un diagnostic sur la nature, la portée des renseignements et des données disponibles, pour définir/valider le socle d'indicateurs évaluables (à partir de la stratégie de mesure de rendement). Il se peut dès lors qu'ils décident de retirer certains indicateurs, diminuant plus ou moins la portée de l'évaluation du rendement, tout en limitant le gaspillage des ressources. En outre, si les évaluateurs experts priorisent avant tout le recueil de données probantes dans le cadre de l'examen approfondi de l'ensemble des sources documentaires existantes et disponibles, selon l'ampleur des lacunes des données de mesure de rendement, ils sont souvent amenés à s'appuyer sur des données plus subjectives recueillies dans le cadre de la consultation des parties prenantes clés du programme (gestionnaires, employés, client, bénéficiaires, etc.). Remarquons que l'ampleur et la teneur des données à combler peuvent nécessiter l'ajout d'une procédure de collecte de données telle que l'enquête, non inclus dans la demande de propositions. Or, si les évaluateurs experts rapportent que les commanditaires peuvent être ouverts à la modification du devis méthodologique (à des fins d'amélioration de la qualité de l'évaluation), les paramètres stricts en ce qui a trait aux intrants disponibles représentent une barrière importante. Pour les évaluateurs, il s'agit donc de miser principalement sur des stratégies ancrées dans l'évaluation planifiée (prescrite) dans la demande de propositions. Dans de telles circonstances, si l'ajout d'une méthode de collecte de données n'est pas faisable, les entrevues d'informateurs clés, incluant ceux des études de cas (deux méthodes toujours incluses dans les devis d'évaluation, dans le présent contexte), constituent une source de données cruciales. Mais, là encore, selon l'envergure et la complexité du programme, et les besoins d'information, le budget « entrevues » peut limiter de manière plus ou moins importante la portée des données (selon les besoins méthodologiques, le nombre d'entrevues d'informateurs budgété dans la demande de propositions est souvent insuffisant). De plus, si l'évaluateur opte pour le recours à l'enquête, outre les potentielles contraintes de temps et d'argent, le faible taux de réponse souvent observé dans le cadre l'évaluation de programmes gouvernementaux peut s'avérer problématique (écueils liés au risque d'épuisement des informateurs clés, problèmes sur le plan de l'accessibilité, à l'accès et à la pertinence/exhaustivité des listes de contacts des répondants clés ciblés, au manque de facilitation de la collecte de l'information à l'interne, etc.).

Nous avons fait plusieurs fois référence aux lacunes à combler en ce qui a trait aux données de mesure de rendement du programme, et qui selon leur nature et leur ampleur vont limiter la capacité des évaluateurs à déterminer l'efficacité du programme. Ces lacunes sont liées à la nature, la pertinence et l'exhaustivité des données de mesure de rendement du programme, et elles peuvent limiter ou freiner la possibilité de passer de la mesure de rendement à l'échelle locale (extrants et résultats des projets financés, initiatives entreprises, affaires traitées, etc. dans le cadre du programme/service évalué) à la mesure de rendement à l'échelle globale (extrants et résultats du programme). Elles nécessitent que les évaluateurs de programme collectent des données probantes supplémentaires, ou misent sur des données plus subjectives afin de rendre compte de la manière la plus fiable et la plus crédible possible de l'atteinte des résultats du programme.

Sur le plan des lacunes liées à la qualité et la pertinence des indicateurs de mesure de rendement, à la nature des informations rapportées. Les évaluateurs sont souvent confrontés au caractère trop général et mal circonscrit des indicateurs de mesure de rendement, et la tendance des bénéficiaires de financement à fournir des textes narratifs pour rendre compte des résultats de leur projet. Une autre difficulté à laquelle font face les évaluateurs, que ce soit un problème au niveau de la définition des indicateurs ou de la procédure de collecte/gestion de l'information (données manquantes), vient du fait que les séries de données recueillies ne rapportent pas toujours pour chaque mesure de résultats - une valeur/un résultat cible et une valeur/un résultat réel(le). Dans les deux situations, les évaluateurs experts sont plus ou moins limités dans leur habileté à générer des résultats à l'échelle du rendement local et du rendement global du programme, ce qui nécessite davantage d'effort de leur part pour affiner/compléter/étayer les données. S'ajoute à cela la nécessité, parfois, de corroborer les données dont les évaluateurs remettent en question la fiabilité.

Sur le plan des lacunes liées au manque d'exhaustivité des données, à la nature incomplète des séries de données probantes relatives aux différents projets financés, dossiers traités ou initiatives entreprises dans le cadre du programme (d'un projet à l'autre, d'un cycle de financement à l'autre sur la période évaluée, pour un ou plusieurs indicateurs). Celles-ci résultent souvent de problèmes sur le plan des activités de collecte et (ou) de gestion de l'information. Il peut s'agir :

i. d'un manque d'alignement entre les données recueillies et les indicateurs du cadre de mesure de rendement (p. ex. lorsque le cadre de mesure de rendement a été finalisé au

moment de l'évaluation, et que les procédures de collecte et de gestion de l'information ont besoin d'être mises à jour) ;

ii. du non-respect des exigences de rapports sur les indicateurs de mesure du rendement (rapports manquants ou incomplets, problèmes d'alignement entre les informations transmises et les indicateurs de mesure de résultats, etc.) ;

iii. de l'accumulation de retard dans l'imputation des informations/données dans le système de suivi ;

iv. d'un mauvais système de gestion des informations ; ou encore,

v. d'un problème au moment de la migration des données à la suite de changements apportés au système de suivi informatique.

Selon 1) l'ampleur du programme ou service évalué ainsi que le nombre de projets financés ou dossiers traités avec des séries de données incomplètes ; et 2) la possibilité de récupérer les données manquantes ou non, les évaluateurs n'ont parfois pas d'autres choix que de retirer les projets/dossiers lacunaires au moment de l'analyse des données relatives à un ou plusieurs indicateurs de mesure du rendement. Nous avons vu que cela peut représenter, par exemple, un peu plus de 20 % des projets financés, soit le retrait de deux années de cycle de financement pour plus de la moitié des indicateurs de mesure de résultats, dans le cadre d'une évaluation qui porte sur quatre années complètes de cycle de financement.

Qu'il s'agisse de construire, nettoyer et (ou) compléter la base de données de mesure de rendement d'un programme, il importe que les évaluateurs de programmes soient en mesure d'obtenir et d'entreprendre l'examen le plus tôt possible (dès la phase de conception de l'évaluation), de l'ensemble de la documentation et des données de programme (sources de renseignements administratifs, de données de rendement, de données opérationnelles, de données secondaires, etc.). L'obtention de l'ensemble des données existantes est une entreprise chronophage et plus ou moins laborieuse qui nécessite la collaboration étroite des gestionnaires de programme, dans un environnement organisationnel très compartimenté et souvent empreint de résistance. L'accès aux données peut être compliqué, notamment lorsque l'évaluation met en exergue des problèmes sur le plan de la gestion des données de mesure de rendement du programme (mauvaise qualité du système d'information de gestion, manque de connaissances et de compétences du gestionnaire de la base de données de rendement, problèmes de gestion et de

surveillance de la collecte systématique de rapports/d'information sur les indicateurs de mesure de rendement, etc.). Par ailleurs, rappelons que les procédures d'échantillonnage, de construction et de passation de l'instrumentation d'autres méthodes (examen de dossiers, enquête, entrevues d'informateurs clés) dépendent des résultats de l'examen des données de monitorage du rendement du programme, et des besoins de compléter/corroborer/étayer les données probantes existantes. Aussi, l'ampleur des lacunes à combler, l'incapacité d'obtenir les informations administratives et les données de rendement du programme existantes et nécessaires, en temps opportun, peut entraver considérablement la qualité de la méthodologie évaluative. Par exemple, lorsque les évaluateurs misent sur la passation d'une enquête auprès des bénéficiaires de financement pour renforcer la qualité et l'exhaustivité des données sur l'efficacité du programme, mais que la période de collecte de données est fortement écourtée (voir l'enquête annulée), car ils n'ont pas reçu les listes de diffusion à temps. Un autre phénomène fréquent que nous abordons dans la section suivante, porte sur les difficultés d'obtention des données financières nécessaires pour rendre compte de l'efficience et l'économie du programme (analyse du rapport coût-bénéfice). Des données qui, à l'époque des projets d'évaluations abordés dans le cadre de cette recherche doctorale, étaient très difficiles, voire impossibles à obtenir.

À la lumière des caractéristiques de la pratique des évaluateurs experts, les écueils méthodologiques et le risque de mésévaluation (dans le présent contexte) sont circonstanciels/situationnels. Ils sont ancrés dans un devis d'évaluation prescrit par la demande de propositions, dont le calibrage est inadéquat, voire irréaliste, au regard des ressources disponibles et mobilisables (temps, argent, qualité et exhaustivité des données probantes existantes). On retrouve là les difficultés caractéristiques de l'évaluation de programmes publics, à savoir de livrer une évaluation fiable et fondée sur suffisamment de données probantes, avec un échéancier souvent trop serré et un budget limité face aux contraintes importantes d'accès aux données (Bamberger et Mabry, 2020). Lamari, Poulin-Larivière et Lemieux (2016) parlent d'un phénomène paradoxal qui « entraine un nivellement par le bas des évaluations et limite la collecte de données requises pour répondre aux exigences de rigueur et de validité des évaluations produites. » (p. 25). Les évaluateurs experts font face à des contraintes politico-organisationnelles qui les obligent, en tant que prestataires de services, à tenter de répondre à des questions d'évaluation de rendement qu'ils jugent non évaluables ou mal évaluables à lumière des

caractéristiques de l'*evaluand* et des ressources disponibles. En outre, si la Politique de l'évaluation de 2009 (et les directives qui l'accompagnent) vise à faciliter la mise en œuvre d'évaluations de qualité (démonstration du rendement des programmes à l'appui de données probantes), et soutenir la prise de décision axée sur les résultats et recommandations, elle est également une potentielle source de mésévaluation. En effet, la perspective de reddition de compte associée aux exigences de la politique évaluative semble avoir l'ascendance sur la faisabilité et (ou) l'utilité de l'évaluation, et place l'évaluation/l'évaluateur/les parties prenantes clés en situation de mésusage. Parmi les situations abordées (ci-après), nous verrons que ces contraintes rendent la démarche évaluative vulnérable tant sur le plan de sa fiabilité que de son utilité, ce qui va à l'encontre de leurs principes (pratiques recommandées), mais aussi des standards de l'évaluation du JCSEE (2011).

- *Problèmes d'alignement « question d'évaluation prescrite – état de l'evaluand – besoins méthodologiques – ressources disponibles »*

L'existence d'une stratégie de mesure du rendement du programme voire d'un système d'information de gestion visant à produire l'ensemble des données nécessaires pour l'évaluation de l'efficacité du programme, ne facilite pas nécessairement la tâche aux évaluateurs de programme (externe), comme voulu (ou autant que voulu). Les lacunes observées sur le plan des données de mesure de rendement limitent plus ou moins les évaluateurs dans la détermination de l'atteinte des résultats escomptés du programme. Et les efforts nécessaires pour combler ces lacunes nécessitent souvent de revoir la répartition des intrants, et peuvent être très coûteux selon les circonstances (c.-à-d. complexité/envergure des programmes gouvernementaux, nature et ampleur des données à collecter, etc.). Or, les intrants octroyés à l'évaluation (paramètres stricts de la demande de propositions) ne reflètent pas nécessairement la réalité de la situation et des besoins des évaluateurs pour remédier efficacement aux écueils sur le plan des données probantes, et renforcer la fiabilité et la crédibilité de l'évaluation (construire/reconfigurer/compléter la base de données de mesure de rendement). Dans certaines situations, lorsqu'il n'existe pas de système de monitorage du rendement, que les données probantes sont insuffisantes ou tout simplement inaccessibles (à temps), la force et l'utilité des résultats d'évaluation sont fortement compromises.

Dans la même lignée, l'analyse de l'efficience et de l'économie est une entreprise très coûteuse, qui nécessite le recueil d'une quantité importante de données de rendement et de données financières. C'est une attente jugée souvent irréaliste par les évaluateurs experts, car les données

financières sont la plupart du temps indisponibles ou insuffisantes. Et lorsque les données disponibles sont justes et complètes, la manière dont elles sont organisées ne permet pas d'analyser le rapport coût-efficacité – c'est-à-dire la démonstration du lien « intrants-extrants-résultats » pour les différentes activités du programme. Aussi, la démonstration de l'efficience du programme repose souvent sur de faibles éléments probants. Et si les évaluateurs redoublent d'efforts pour recueillir des données supplémentaires, celles-ci sont souvent d'ordre subjectif, et ce type d'évaluation demande une démonstration empirique. On retrouve ici un constat important émis dans le rapport d'évaluation de la Politique évaluative de 2009 (Secrétariat du Conseil du Trésor du Canada, 2015). Selon Dumaine (2018), il est utopique de penser qu'il est possible de réaliser systématiquement l'analyse de l'efficience et de l'économie lors de l'exercice d'évaluation quinquennale des programmes fédéraux. Ce type d'étude devrait faire l'objet de projets indépendants.

Un autre exemple d'attente irréaliste a trait à la difficulté d'évaluer le rendement de macroprogrammes publics, soit des programmes composés, par exemple, de sous-programmes. C'est un élément pas toujours pris en considération lors de la planification quinquennale de l'évaluation des programmes d'une agence donnée, lors de la conception de la demande de propositions – soit l'obligation d'évaluer simultanément les deux sous-programmes, l'octroi de ressources insuffisantes à la lumière de l'ampleur de la tâche, et des carences sur le plan des données de mesure de rendement (une des situations observées comprenait une base de données de mesure de rendement incomplète). En outre, dans ces circonstances, l'évaluation d'un des sous-programmes à une portée très limitée et dans certains cas ce dernier devient presque un indicateur de mesure de l'autre sous-programme.

Enfin, la portée et la nature des données à recueillir pour mesurer adéquatement les résultats ultimes/l'impact de programmes gouvernementaux de grande envergure (multiplicité des services, diversité des activités entreprises, etc.) nécessite le recours à un devis méthodologique sophistiqué très coûteux et souvent incompatible avec le budget et l'échéancier octroyé à l'évaluation et (ou) les problèmes liés à l'existence, l'accessibilité, la qualité et l'exhaustivité des données de mesure de rendement. Un situation très fréquente dans le cadre de l'évaluation de programmes publics (McDavid, Huse et Hawthorn, 2019). Aussi, on a observé que le devis contient alors une question à laquelle l'évaluation ne répondra pas, ou tentera de répondre en partie, ou de manière quelque

peu détournée. Pour Dumaine (2018), cet exercice devrait être entrepris en dehors du processus d'évaluation quinquennale de programmes du gouvernement fédéral.

- *Problème d'alignement « question d'évaluation prescrite – indicateurs – état de l'objet évalué »*

La nature des initiatives entreprises ou projets financés par le programme évalué peut freiner considérablement la possibilité de rendre compte de manière appropriée de l'efficacité du programme (atteinte des résultats à court, moyen, et long terme) tel qu'exigé par la politique évaluative, et le mandat d'évaluation prescrit dans la demande de propositions. Parmi les situations abordées avec les évaluateurs experts participants, on observe que :

- L'obligation d'évaluation quinquennale, dans certains cas, fait que l'évaluation de l'efficacité du programme, plus précisément des résultats à moyen et (ou) à long terme, arrive trop tôt dans le cycle de vie des projets financés, notamment quand ceux-ci impliquent la création de nouvelles technologies.

- Le manque de maturité et de stabilité du programme limite grandement l'existence et l'accès aux données probantes nécessaires pour rendre compte de son rendement de manière appropriée, plus particulièrement les résultats à moyen et long terme ainsi que le rapport coût-bénéfice des activités du programme.

- L'exigence de rendre compte de l'atteinte des résultats à long terme d'un programme, alors que les initiatives entreprises sont de l'ordre d'un domaine de compétence partagée, et que les évaluateurs ne sont pas en mesure d'isoler l'impact des efforts du gouvernement fédéral dans le cadre du programme évalué (efforts du gouvernement fédéral, des gouvernements provinciaux et territoriaux, et d'autres organisations dans la réduction de l'incidence d'un phénomène donné, par exemple).

- L'évaluation de l'efficacité porte sur l'atteinte de résultats escomptés du programme, de l'atteinte de jalons finaux tels que la publication d'une nouvelle réglementation en réponse à un mandat politique, alors qu'un tel accomplissement n'est pas du ressort du programme. Dans le cas présent, la mesure de l'efficacité ne tient pas compte de la composante externe relative à la phase d'approbation du règlement par un cabinet ministériel, et de la nature changeante des priorités politiques. De plus, outre l'influence du climat politique, le caractère chronophage et laborieux de la rédaction/publication d'une nouvelle réglementation n'est pas toujours compatible avec le

rythme d'évaluation officielle tous les cinq ans. Cela conduit les évaluateurs à dévier du mandat d'évaluation de l'efficacité (prescrit), et de rendre compte des progrès du programme vers l'atteinte des résultats. Un choix rapporté comme une limite de l'évaluation, en dépit d'une attente évaluative inadaptée, et du désir des évaluateurs de renforcer la représentativité des résultats et de tenir compte des préoccupations des utilisateurs ciblés. En effet, selon les circonstances, la non-atteinte du résultat escompté ne reflète pas nécessairement un problème sur le plan de l'efficacité du programme, et l'évaluation n'est pas en mesure de pointer des faiblesses au niveau du programme. Une situation délicate pour l'évaluateur qui n'a pas d'autre choix que de rapporter des résultats négatifs, sans pouvoir fournir aux gestionnaires de programme, utilisateurs ciblés de l'évaluation, des propositions de remédiation (et si certaines recommandations pouvaient être émises au cabinet ministériel concerné, l'évaluateur n'a pas le droit de le faire). Enfin, si 1) les utilisateurs ciblés anticipent les résultats négatifs et comprennent l'obligation de transparence de l'évaluateur et la nécessité de les rapporter ; et 2) que l'évaluateur redouble d'efforts pour rendre compte du travail réglementaire entrepris pour répondre au mandat et des enjeux externes sources de limitations, les évaluateurs demeurent préoccupés face au risque de mésusage des résultats (assouplissement/minimisation des résultats négatifs) au moment de la phase de révision du rapport final à l'interne.

L'ensemble des situations que nous avons décrites font l'objet des limites rapportées par les évaluateurs experts durant les différentes activités de collecte de données de cette thèse doctorale. Précisons qu'à l'instar de leurs pratiques recommandées, les évaluateurs experts ont pris le soin d'expliciter les limites méthodologiques rencontrées et l'influence qu'elles peuvent avoir sur la portée de certains résultats (dans leur rapport d'évaluation respectif qui fait l'objet de l'activité d'explicitation). En effet, ces écueils sont abordés à différents endroits dans le rapport, notamment :

i. Au niveau du cadre méthodologique, où les évaluateurs expliquent non seulement la nature des limites, mais aussi les stratégies mises en œuvre pour y remédier. La principale stratégie mobilisée repose sur la triangulation, la complémentarité des méthodes quantitatives et qualitatives, pour compléter/corroborer/étayer/expliquer les données et résultats obtenus en ce qui a trait aux différents objets d'évaluation, plus particulièrement l'atteinte des résultats escomptés par le programme.

ii. Au niveau de la présentation des résultats et la formulation des constats évaluatifs, où les évaluateurs tiennent compte des écueils méthodologiques dans le poids accordé aux différents résultats. Ils expliquent clairement les raisons pour lesquelles la portée de certains résultats est plus ou moins limitée.

D'après les différentes situations vécues et rapportées par les évaluateurs experts, l'utilité des évaluations produites peut s'avérer plus ou moins limitée. Par exemple, lorsque :

i. La démonstration de l'efficience repose sur une quantité limitée de données empiriques, et repose davantage sur des données subjectives, et ne fournit pas assez de renseignements concrets aux utilisateurs ciblés.
ii. L'évaluation a recueilli très peu de données probantes pour rendre compte des résultats du programme, et que l'évaluateur est limité dans sa capacité à fournir des recommandations pertinentes.
iii. Les résultats de l'évaluation du rendement d'un des sous-programmes du macroprogramme évalué ont une portée très limitée.

Pour terminer, nous avons vu dans le cadre de l'activité d'explicitation qu'un autre allié de la faisabilité, crédibilité et utilité de l'évaluation est la gestion du projet d'évaluation par un évaluateur de la division interne de l'évaluation de l'agence commanditaire, et la création d'un comité interne de l'évaluation composé des utilisateurs ciblés. Outre son rôle de vérificateur/d'approbateur de la pertinence et la qualité du contenu des livrables produits par le gestionnaire de l'évaluation à l'externe (rapport méthodologique, rapport(s) technique(s) – matrice(s) de résultats, rapport final), le chef de l'évaluation interne est une ressource indispensable tout au long de l'évaluation, de la conception de l'évaluation à la communication des résultats. Nous avons vu que pour les évaluateurs experts, il est un collaborateur essentiel, un aiguilleur quant à la faisabilité et la pertinence/justesse des propositions méthodologiques, un agent de liaison/consultation auprès des utilisateurs ciblés, un facilitateur d'accès aux sources d'information, etc. Le chef de l'évaluation interne est là pour s'assurer que le devis d'évaluation permet de répondre de manière appropriée aux questions évaluatives fondamentales de la politique évaluative fédérale, et prévenir les risques de détournement ou de manipulation de l'évaluation par les hauts responsables et les gestionnaires de programme.

Cependant, si la collaboration étroite et la communication régulière (et efficace) entre l'évaluateur sénior gestionnaire de l'évaluation externe et chef de l'évaluation interne sont indispensables pour assurer la qualité et l'utilité de l'évaluation, la dynamique au sein de ce duo peut potentiellement devenir une menace, ou tout du moins être contreproductive.

L'évaluateur expert 'S2' nous a fait part d'un élément quelque peu préoccupant, à savoir que dans le présent contexte, les chefs de l'évaluation interne ont rarement plus de cinq années d'expérience, contre les 15 à 20 années d'expérience des gestionnaires de l'évaluation à l'externe (à savoir nos évaluateurs experts). Comme il nous l'explique, ce manque de compétence perçue chez son principal collaborateur à l'interne, un agent de liaison et de médiation auprès des utilisateurs ciblés, et surtout une autorité décisionnelle en ce qui a trait notamment au contenu du devis d'évaluation, peut s'avérer problématique pour la qualité et le bon déroulement de l'évaluation. Si dans la plupart des cas (selon l'expérience de 'S2'), les évaluateurs internes semblent s'appuyer sur l'expertise du gestionnaire de l'évaluation externe, il n'en demeure pas moins que dans d'autres cas, ces derniers peuvent faire preuve de résistance et adopter une attitude réfractaire face aux propositions méthodologiques de l'évaluateur externe. En effet, comme nous l'explique 'S2', au cours de ses 20 ans d'expérience en évaluation de programme, il a été confronté à suffisamment de situations pour anticiper la résultante et les potentiels écueils des différentes propositions méthodologiques selon les caractéristiques d'une situation donnée, c'est-à-dire l'*evaluand* et les contraintes de terrain associées. Il aborde ici une dimension essentielle de la définition de l'apport de l'expertise (McPherson, 2015, *processus décisionnel des évaluateurs experts*; Tourmen, 2007, *profil de l'évaluateur expérimenté* ; Benner et Tanner, 1987, 1982 ; Dreyfus et Dreyfus, 1980). Cependant, le manque d'expérience chez les évaluateurs internes peut entrainer un manque de compréhension des procédures décisionnelles opérées par l'évaluateur sénior externe et un manque de confiance en ses propositions, ces derniers préférant se fier à leur cadre de référence et expérience dans d'autres domaines plus ou moins connexes. Rappelons que les études sur l'utilisation de l'évaluation ont montré que le manque d'expérience en évaluation de programme des parties prenantes clés et leur manque de compréhension en ce qui a trait au processus évaluatif ainsi que leur manque de confiance en l'évaluateur (faible crédibilité perçue), et l'échec de l'évaluateur sur le plan de de la communication claire et efficace des informations au cours de l'évaluation, sont parmi les premières causes de résistance chez les parties prenantes clés et une influence majeure sur la rigueur de l'évaluation (Marchand, 2020 ; Patton, 2018 ; Stevahn

et King, 2016 ; Alkin, 2005 ; Taut et Alkin, 2003). Ces études ont ainsi mis de l'avant l'importance accordée aux compétences relationnelles et habiletés sociales des évaluateurs de programme. Aussi, si les parties prenantes clés ou plutôt les utilisateurs ciblés dans les présentes études étaient des gestionnaires de programme, ce constat semble pertinent dans le contexte de notre recherche où l'entité perçue comme cliente par nos évaluateurs experts est la division interne de l'évaluation et ses évaluateurs. Dans le même ordre d'idée, parmi les principales causes de mésévaluation et par conséquent de possible mésusage, on retrouve notamment : le manque de qualité de la méthodologie évaluative ainsi que la mauvaise compréhension et appréhension du contexte évaluatif (Alkin, 1990) ; le manque de compétence de l'évaluateur (Cousins, 2004) ; et le recours à des devis méthodologiques inappropriés, ce qui entraine l'utilisation de méthodes de collectes de données incorrectes ou l'absence de résultats probants pour émettre des conclusions valides et fiables (Alkin et King, 2017). Il peut s'agir dans notre contexte de l'évaluateur interne et (ou) de l'évaluateur externe gestionnaire du projet d'évaluation. Le manque de connaissance et d'expérience de l'évaluateur interne et sa potentielle résistance face à l'apport de l'expertise de l'évaluateur sénior externe, et son inhabileté à soutenir le processus évaluatif et veiller à la rigueur ainsi qu'à la pertinence et la précision-justesse des choix et pratiques méthodologiques de l'équipe évaluative externe, peuvent conduire à la remise en question de la validité et la fiabilité du processus et des résultats évaluatifs.

Les résultats montrent que la pertinence du devis d'évaluation repose sur la capacité des évaluateurs à réaliser un diagnostic efficace de la situation évaluative rencontrée afin de revoir et d'opérationnaliser les paramètres de l'évaluation initialement prescrits par les clients, et de s'assurer de la faisabilité et de la légitimité de l'évaluation. Cette entreprise nécessite la mobilisation d'un ensemble de ressources personnelles composé de savoirs méthodologiques, procéduraux et expérientiels, de connaissances spécifiques à l'environnement professionnel, de savoir-faire relationnels, et cognitifs acquis dans et par l'expérience.

L'analyse des pratiques et des perceptions des évaluateurs de programmes publics experts montre selon nous l'incidence du niveau d'expertise de l'évaluateur de programme externe sur sa capacité à anticiper, à identifier, à comprendre et à gérer efficacement les contraintes situationnelles afin de concevoir une évaluation fiable et utile. De tels résultats montrent l'importance d'offrir le plus tôt possible aux évaluateurs novices l'occasion de participer à des projets d'évaluation du début à la fin. D'accroitre les occasions pour ces derniers d'observer puis

de contribuer sur le plan de la conception et de la gestion de l'évaluation dans le cadre de petits projets où le devis est plutôt simple. **Des tâches cruciales qui, dans le présent contexte de pratique, sont réalisées par les évaluateurs experts. Nous revenons en profondeur sur ce point dans le cadre de l'interprétation de la question générale de recherche et la proposition de recommandations (Chapitre 5).**

CHAPITRE 5. CONCLUSION

5.1 Interprétation générale

<u>Question principale de recherche</u> : En quoi les évaluateurs novices et les évaluateurs experts œuvrant dans des cabinets de conseil en évaluation de programmes publics se distinguent-ils dans les choix méthodologiques qu'ils opèrent relativement à la réalisation de leurs évaluations de programme ?

Cette thèse doctorale avait pour principal objectif de décrire les caractéristiques générales des pratiques des évaluateurs de programme et l'influence des facteurs contextuels sur celles-ci. Plus précisément, nous voulions rendre compte des différences et des ressemblances sur le plan des choix méthodologiques opérés par les évaluateurs novices et les évaluateurs experts, et la manière dont ils appréhendent et gèrent les contraintes situationnelles.

Dans le contexte de cette recherche, l'exercice professionnel des évaluateurs de programme s'inscrit dans une activité de services où, en tant que prestataires (ou membres de l'équipe évaluative prestataire), ils répondent à des commanditaires et produisent des connaissances qui permettent l'attribution d'un jugement de valeur sur un programme public fédéral. Le mandat d'évaluation et la conception initiale de l'évaluation (contenu de la demande de propositions) sont ancrés dans les exigences de la Politique d'évaluation du gouvernement fédéral (2009) et les directives qui l'accompagnent (Fonction de l'évaluation, 2009). Les commanditaires sont représentés par deux entités – le chef de l'évaluation de l'agence gouvernementale (pouvoir décisionnel) et le comité d'évaluation interne composée des utilisateurs ciblés (pouvoir consultatif). En tant que prestataire de services, l'équipe évaluative externe est responsable de l'exécution ou plutôt de l'opérationnalisation et la réalisation de l'évaluation planifiée par les commanditaires, plus précisément l'unité d'évaluation interne. L'équipe évaluative externe est dirigée par l'évaluateur expert (la plupart du temps un partenaire de la firme contractante) qui à titre de gestionnaire du projet est chargé des activités cruciales de l'évaluation : liaison avec l'unité d'évaluation interne, conception de l'évaluation, supervision des procédures de collecte et d'analyse de données, vérification et analyse des données, agrégation et communication des résultats, et formulation des constatations évaluatives (rédaction du rapport d'évaluation final).

Aussi, si la combinaison de ces différents éléments (les directives de la politique évaluative fédérale, le duo de gestionnaires de l'évaluation interne-externe et la création d'un comité d'évaluation composé d'utilisateurs ciblés) peut être une source de contraintes méthodologiques plus ou moins importantes, elle constitue une stratégie essentielle pour assurer la faisabilité et la légitimité (c.-à-d. qualité, fiabilité, validité/représentativité, utilité, crédibilité) de l'évaluation. Enfin, les évaluateurs novices quant à eux peuvent avoir différents statuts au sein de l'équipe évaluative externe, des statuts qui leur confèrent différents degrés de responsabilité, de contribution dans le cadre du projet d'évaluation. Ils peuvent exercer en tant que juniors exécutant, exécutant avancé, ou collaborateur. Lorsqu'ils exercent dans une capacité d'exécutant (ou subalterne), les évaluateurs juniors travaillent en silos, ils rejoignent l'équipe évaluative à la phase de mise en œuvre du cadre méthodologique. Lorsqu'ils sont juniors collaborateurs, ils participent à l'ensemble des étapes du projet d'évaluation, de la conception à la communication des résultats d'évaluation. Dans tous les cas, l'ensemble de décisions est du ressort de l'évaluateur sénior chargé du projet à l'externe, limitant considérablement (ou plutôt complètement) notre capacité à comparer les choix méthodologiques opérés par les évaluateurs novices et les évaluateurs experts. En revanche, les résultats obtenus sur le plan des caractéristiques des pratiques recommandées et des pratiques réelles des évaluateurs participants mettent en exergue des profils de pratique. Cela ouvre ainsi une discussion sur développement professionnel des jeunes évaluateurs sur le terrain relativement aux dimensions méthodologique et sociopolitique de l'évaluation de programme. Sur leur préparation à exercer en situation, à gérer de manière appropriée les contraintes de terrain tout en veillant à la qualité et à l'utilité des évaluations dont ils auront la responsabilité, ou partageront la responsabilité (dans le futur).

De manière générale, les résultats obtenus sur le plan des pratiques des évaluateurs de programme dans le présent contexte rejoignent ceux mis en lumière dans l'étude de Tourmen (2007) sur les pratiques des évaluateurs de politiques publiques débutants et expérimentés. L'articulation des pratiques recommandées des évaluateurs novices participants à notre recherche se centre sur une logique technique/méthodologique de production de connaissances sur les objets évalués. En effet, nous avons vu que durant l'analyse critique d'un devis d'évaluation tierce, ces derniers se sont concentrés sur le caractère approprié de la démarche évaluative, sur la démonstration de l'effort de rigueur méthodologique, de la fiabilité et de la représentativité des procédures méthodologiques (alignement « question d'évaluation – indicateurs – sources de

données », caractère systématique des procédures d'échantillonnage, de collecte et d'analyse de données, représentativité des sources d'information, triangulation des sources de données, transparence du raisonnement évaluatif, etc.). Aussi, s'ils adressent certains éléments relatifs au contexte évaluatif, notamment l'influence des contraintes temporelles et budgétaires, la nécessité de respecter les exigences politiques en matière d'évaluation, l'importance de tenir compte de la sensibilité culturelle de l'environnement évaluatif, etc., ces facteurs situationnels semblent être relégués au second plan dans le cadre du raisonnement des évaluateurs novices. Ils sont considérés de manière superficielle comparativement aux évaluateurs de programme experts. En effet, nous avons observé que le raisonnement méthodologique des évaluateurs experts s'inscrit dans une logique de régulation. Centrés sur les résultats, ces derniers apprécient ou adaptent les choix méthodologiques aux caractéristiques des situations évaluatives rencontrées. En somme, il s'agit pour eux de veiller au caractère approprié et faisable du devis d'évaluation à la lumière de l'état de l'*evaluand* (du programme, des différents objets évalués tels que l'atteinte des résultats, etc.) et des contraintes inhérentes au terrain (contraintes temporelles, budgétaires, politico-organisationnelles, d'accès aux données). Et d'assurer la fiabilité, la représentativité, la crédibilité et l'utilité de l'évaluation (informations recueillies, résultats, constatations évaluatives).

Le cheminement de la pratique des évaluateurs de programme novices

De telles constatations ne sont pas surprenantes, puisque la pratique des évaluateurs de programme novices, les premières années, est principalement ancrée dans le volet *technique/méthodologique* du socle de compétences essentielles des évaluateurs de programme (Société canadienne d'évaluation, 2018 ; Stevahn, King, Ghere et Minnema, 2005 ; King, Stevahn, Ghere et Minnema, 2001). Plus particulièrement la maîtrise de l'utilisation pratique de différentes méthodes et procédures, de l'instrumentation, etc. Il s'agit également de développer la capacité analytique des évaluateurs novices, une compétence transversale essentielle à l'entreprise des activités propres au volet *analyse situationnelle* (c.-à-d. description du programme et évaluation de l'évaluabilité) et au volet *technique* (c.-à-d. conception de l'évaluation, collecte/analyse de l'information, interprétation des résultats - *mobilisation du jugement évaluatif*). On observe qu'au gré de leur expérience, les évaluateurs novices subalternes (qui exercent en tant qu'exécutant) contribuent de plus en plus significativement à la collecte des informations, puis à leur analyse, et enfin à la synthèse et l'interprétation des résultats (préparation de rapports techniques sur lequel s'appuie l'évaluateur sénior pour rédiger le rapport d'évaluation final). Notons qu'en dépit de leur

arrivée tardive sur le projet d'évaluation, les évaluateurs juniors exécutants démontrent une bonne connaissance globale du devis d'évaluation et des caractéristiques de la situation évaluative (complexité/envergure du programme, contraintes politiques, ressources disponibles, etc.). En effet, s'ils perçoivent la démarche évaluative et les choix méthodologiques comme standards, presque par défaut du fait de l'ancrage politique fort, ils sont en mesure de les discuter. Ils sont en mesure de cerner, par exemple :

- l'intention derrière le recours à une approche multiméthodes (triangulation, complémentarité des méthodes quantitatives et qualitatives) ;
- le rôle crucial de l'analyse documentaire pour l'évaluation de programme de grande envergure, et plus particulièrement pour le recueil de données de programme/de données secondaires pour rendre compte de l'efficacité ;
- les objectifs des différentes méthodes de collecte de données utilisées ;
- le lien entre les intrants disponibles et les critères d'échantillonnage ; etc.

Outre des temps de discussion informels, les juniors exécutants participent à une réunion de préparation à la collecte d'information. Une réunion durant laquelle l'évaluateur sénior gestionnaire présente à l'équipe les points clés du devis d'évaluation, les données obtenues jusque-là, le contenu de l'instrumentation, et les éléments importants sur lesquels ils doivent investiguer en profondeur. Le travail de préparation implique également l'examen de la demande de propositions et du rapport méthodologique (matrice évaluative, protocole méthodologique, instrumentation, critères d'échantillonnages, etc.). Cependant, la nature décontextualisée de la pratique des évaluateurs juniors exécutants freine leur habileté à saisir pleinement les procédures décisionnelles engagées dans la conception de l'évaluation ainsi que la dynamique de travail entre l'équipe évaluative externe et le comité d'évaluation interne. Le travail en silo constitue un angle mort sur le plan des apprentissages relatifs à la dimension sociopolitique de l'activité d'évaluation. Il est aussi une potentielle source de biais quant à leur perception du poids des contraintes politico-organisationnelles sur les choix méthodologiques, et du pouvoir décisionnel de l'équipe évaluative externe. L'un des éléments saillants repose sur la perception d'un climat d'obligation/d'exécution de la méthodologie évaluative telle que prescrite par la Politique évaluative, la demande de propositions, et les commanditaires/le comité d'évaluation interne. Les évaluateurs experts (gestionnaires de projet d'évaluation à l'externe) ont eux-mêmes initialement déclaré qu'ils réalisaient l'évaluation planifiée par les clients, et que leurs responsabilités reposaient sur la

conception et la mise en œuvre du cadre méthodologique ainsi que la synthèse des résultats/conclusions (production du rapport d'évaluation final). Mais, ces derniers minimisaient leur rôle dans la définition de la démarche d'évaluation. En effet, nous avons vu que les paramètres de l'évaluation prescrits dans la demande de propositions sont généralement génériques et nécessitent l'entreprise d'une phase de *scoping-design* qui vise à cadrer l'évaluation (affinage, opérationnalisation du plan d'évaluation initialement prescrit), dont le principal extrant est la matrice évaluative. Cette phase implique une rencontre de travail initiale avec les utilisateurs ciblés, l'analyse préliminaire de la documentation et des données de programme existantes et rendues disponibles, voire la réalisation d'entrevue préliminaire avec les gestionnaires de programme. Elle nécessite également la collaboration étroite avec le chef de l'évaluation interne (afin d'assurer la faisabilité, la fiabilité, l'utilité), ainsi que l'obtention de la rétroaction des utilisateurs ciblés.

L'entreprise de conception de l'évaluation représente un pan de la démarche d'évaluation auquel les évaluateurs novices ont la possibilité d'assister lorsqu'ils sont intégrés à l'équipe dès le début du projet (en tant que junior collaborateur). Dès lors, la pratique des évaluateurs novices comprend l'acquisition de savoirs liés à la conception de la démarche d'évaluation (les volets *analyse situationnelle* et *technique* du socle de compétences des évaluateurs de programme). Ce type d'expérience leur permet non seulement d'assister, mais aussi de contribuer dans une certaine mesure aux différentes discussions d'ordre méthodologique (rencontre de travail initiale avec les clients, réunions d'équipe, etc.). On se trouve ici à un tournant important sur le plan de la pratique des évaluateurs novices, ces derniers passant d'une pratique décontextualisée à une pratique située, ancrée dans la réalité de terrain. C'est une pratique qui s'inscrit dans un monde habité de contraintes importantes, où l'évaluateur doit veiller à réaliser une évaluation de qualité et recourir à la méthodologie la plus adéquate et rigoureuse possible (Bamberger et Mabry, 2020 ; Rossi, Freeman et Lipsey, 2019 ; Smith et Brandon, 2008). Si les évaluateurs juniors qui ont de l'expérience en tant que juniors collaborateurs reconnaissent le caractère rigide du contexte évaluatif fédéral, ils ne semblent pas nécessairement percevoir l'évaluation externe comme inscrite dans un climat d'obligation (d'exécution). Ils comprennent que beaucoup de paramètres méthodologiques sont initialement prescrits dans la demande de propositions (questions d'évaluation, méthodes, sources d'information, critères d'échantillonnage, etc.), mais à la différence des évaluateurs subalternes, ils n'ont pas une perception limitée de l'ampleur du rôle de

leur équipe dans la planification de l'évaluation. Ils comprennent l'aspect crucial de la collaboration étroite avec l'unité d'évaluation interne, qu'ils perçoivent, à l'instar des évaluateurs experts, comme un allié, un agent facilitateur ou encore un aiguilleur méthodologique plutôt qu'une autorité prescriptive. Si en toile de fond de la description de la démarche de planification de l'évaluation on observe comme point de référence les directives évaluatives fédérales (ou pratiques évaluatives standards du gouvernement fédéral), on retrouve au cœur de leurs explications des préoccupations liées aux caractéristiques de la situation évaluative (état de l'*evaluand*, besoins des utilisateurs ciblés, ressources disponibles, etc.), et à la faisabilité, la fiabilité/représentativité et l'utilité de l'évaluation. Par ailleurs, la participation à la conception de l'évaluation semble également renforcer les capacités techniques de l'évaluateur novice ainsi que la qualité de l'évaluation. En effet, armé d'une connaissance et d'une compréhension profonde du devis d'évaluation et du contexte évaluatif, l'évaluateur est mieux équipé, il est plus en contrôle de l'instrumentation, des activités de collecte et d'analyse de données, et donc plus à même de recueillir et de produire des informations utiles et justes.

La transition d'une pratique d'exécutant qui travaille en silo à une pratique contextualisée où l'évaluateur novice se perçoit comme « responsabilisé » (comme un réel contributeur) représente un moment charnière dans son développement professionnel tant sur le plan de ses capacités que de son identité professionnelle d'évaluateur de programme. À titre de collaborateur, il contribue aux discussions d'ordre méthodologique, à la révision du contenu de l'instrumentation ainsi qu'à la coordination des procédures de collecte/gestion/analyse des informations. Nous avons vu qu'à la différence des exécutants, il a la possibilité d'interagir avec les clients, plus précisément l'équipe d'évaluation interne, afin notamment d'obtenir de plus amples informations, d'accéder à certaines sources d'information, ou encore de leur demander des conseils/rétroactions pour veiller à l'utilité des informations ou des outils produits de l'évaluation. Il peut également réaliser la synthèse, l'interprétation et la communication des résultats (production de la matrice de résultats générale, rédaction de la première ébauche du rapport final). Et, si l'évaluateur sénior gestionnaire du projet demeure responsable de l'ensemble des livrables, et qu'il révise le contenu de tous les produits de l'évaluation, ce dernier semble prendre le temps de faire part de sa rétroaction à l'évaluateur junior collaborateur, d'expliquer les différentes modifications ou suggestions apportées (partage de ses savoirs expérientiels). L'encadrement de la pratique du jeune évaluateur semble donc s'inscrire davantage dans une perspective d'accompagnement, de mentorat, plutôt

que de microgestion. On remarque que l'expérience de collaborateur semble amorcer une prise de conscience de soi en tant qu'évaluateur de programme chez l'évaluateur novice, avec une augmentation à la fois de la confiance en ses compétences, mais aussi de la vocalisation de ses besoins en matière d'apprentissages expérientiels afin d'être prêt à mener à bien un projet d'évaluation lorsque le moment se présentera (volet de compétences *pratique réflexive*). On observe ici de l'importance d'offrir le plus tôt possible aux évaluateurs novices l'occasion de participer à des projets d'évaluation du début à la fin, d'accroitre les occasions d'observation et de contribution sur le plan de la conception et de la gestion de l'évaluation dans le cadre de petits projets où le devis est plutôt simple. Des tâches cruciales qui, dans le présent contexte, sont réalisées par les évaluateurs experts ; et que les jeunes évaluateurs ne réaliseront pas avant leur 5e année d'expérience professionnelle. Nous étayons la discussion sur ce point dans la section qui suit.

Le développement de la capacité à préparer, expliciter, négocier et justifier les choix méthodologiques opérés dans le cadre de la conception du devis d'évaluation

La qualité de l'évaluation est intrinsèquement reliée à la capacité de l'évaluateur à mobiliser des savoirs pratiques en situation réelle, son expertise, ses valeurs et son jugement afin d'opérer les choix méthodologiques les plus appropriés au regard des caractéristiques de l'environnement évaluatif, et des contraintes inhérentes au terrain (Bamberger et Mabry, 2020 ; McDavid, Huse et Hawthorn, 2019 ; McPherson, 2015 ; Greene, Lipsey, Schwandt, Smith et Tharp, 2007 ; Greene, 2006 ; King, 2003). Pour les évaluateurs experts participant à cette recherche doctorale, la conception de l'évaluation est tributaire de l'expérience de l'évaluateur, particulièrement face à la nature complexe d'une telle entreprise dans un contexte évaluatif hautement contraignant. C'est pourquoi ce sont eux qui s'en chargent. Les décisions méthodologiques engagées durant la conception de l'évaluation s'inscrivent dans la réalisation d'un diagnostic approfondie de la situation évaluative (et de ses caractéristiques), à savoir : l'état de l'*evaluand*, les contraintes politico-organisationnelles, les ressources nécessaires et disponibles pour mener à bien l'évaluation tant sur le plan des intrants que de l'accès aux données, etc. On fait référence ici au cycle itératif d'analyse du programme et de l'évaluabilité, et de la conception de la démarche d'évaluation (phase de *scoping-design).* Une démarche animée de préoccupations pratiques et méthodologiques qui portent sur la faisabilité et la légitimité (fiabilité, représentativité et utilité) de l'évaluation. Pour ce faire, les évaluateurs s'appuient sur des théories spontanées ou

implicites développées dans et par l'expérience (McPherson, 2015 ; Tourmen et Droyer, 2013 ; Brown Urban et Trochim, 2009). Les travaux de Tourmen (2009, 2008, 2007) sur les pratiques des évaluateurs de politiques publiques montrent que les évaluateurs expérimentés mobilisent des ressources cognitives pour :

1) analyser, interpréter une situation évaluative donnée (*stratégies de diagnostic* – mobilisation de réseau d'indices à examiner, de schèmes d'interprétation des indices/théories implicites ou savoirs tenus pour vrai sur le réel) ;

2) agir en conséquence (*stratégies d'action* – mobilisation de règles d'action pour adapter les choix des finalités, modalités et outils d'évaluation à la situation diagnostiquée ; et *stratégies d'arbitrage/de compromis* – mobilisation de champs de concepts-en-acte ayant traits à la faisabilité et la légitimité, qui permettent de trancher entre différentes options et faire preuve de souplesse dans les choix méthodologiques.).

L'explicitation des pratiques réelles des évaluateurs experts montre que la conception de l'évaluation nécessite la mobilisation de ressources personnelles composées de savoirs méthodologiques procéduraux et expérientiels, de connaissances spécifiques à l'environnement professionnel (programmes publics, système politico-organisationnel du gouvernement fédéral, politique évaluative, logique des acteurs publics, gouvernance de l'évaluation externe de programmes publics, etc.), et de savoir-faire cognitifs acquis dans et par l'expérience (indispensables à l'analyse de la situation et à la résolution de problèmes, à la conception et à la réalisation de l'évaluation). Aussi, le développement professionnel des évaluateurs novices doit être axé sur l'acquisition et le renforcement à la fois de compétences méthodologiques, mais aussi de leur capacité à diagnostiquer une situation évaluative, effectuer les choix méthodologiques les plus appropriés en situation, en tenant compte de la faisabilité et de la légitimité de la démarche évaluative et de ses conclusions.

D'après nos observations, l'accent mis sur le développement des capacités analytiques et de formulation du jugement évaluatif semble constituer une première étape dans le développement professionnel des évaluateurs novices. Les évaluateurs experts rapportent aussi faire de leur mieux pour expliciter la démarche évaluative, les caractéristiques du projet et du contexte évaluatif, etc., de manière à situer le plus possible la pratique et enrichir les connaissances des évaluateurs subalternes qui exercent en silo (discussions informelles, réunion de préparation de la collecte de l'information). Au fil de l'enrichissement de l'expérience des évaluateurs novices exerçant dans

une capacité d'exécutant on remarque une amélioration des connaissances générales en évaluation de programme, des savoirs spécifiques à l'environnement évaluatif dans le contexte de l'évaluation externe de programmes publics fédéraux. Au demeurant, les expériences les plus propices à l'apprentissage des évaluateurs novices sont celles où ils participent à l'ensemble des étapes d'un projet d'évaluation. En effet, nous avons vu que les évaluateurs novices avec une expérience de collaborateur étaient en mesure de décrire avec précision la démarche et le processus décisionnel engagés dans la conception de l'évaluation. De même, ils démontraient une bonne connaissance des choix méthodologiques opérés par les évaluateurs experts en réponse aux caractéristiques de la situation évaluative, favorisant ainsi le développement de ressources cognitives mobilisables en situation.

Pour développer la capacité de diagnostic/pronostic et l'acquisition de règles d'action qui permettent l'adaptation en situation, il est important, (1) d'offrir aux évaluateurs novices de nombreuses opportunités de suivre le déroulement entier d'une évaluation afin qu'ils puissent observer les résultats, en tirer des leçons et comprendre l'enchainement global des étapes ; (2) de leur permettre d'assister à la réalisation d'évaluation dans des situations variées ; et (3) de leur offrir l'aide de professionnels expérimentés pour identifier les indices caractéristiques d'une situation évaluative, savoir s'y adapter et anticiper certaines difficultés (Tourmen, 2009, 2007). McPherson (2015) parle du rôle essentiel du mentorat, de la transmission des théories implicites et des schèmes opératoires des évaluateurs experts aux évaluateurs novices ou moins expérimentés. Ce sont d'éléments dont les évaluateurs séniors gestionnaires participants à cette recherche doctorale sont conscients. En outre, s'ils essaient autant que possible d'inclure les évaluateurs juniors « avancés » dans l'ensemble des étapes de projets d'évaluation, la réalité de terrain, avec la multitude de projets d'évaluation réalisés de manière simultanée au sein du cabinet-conseil, et le besoin de mains-d'œuvre pour la collecte des données, semble causer des difficultés sur le plan logistique. De même, les échéanciers très serrés de la plupart des évaluations de programmes gouvernementaux dont le mandat s'inscrit dans les exigences de la politique évaluative fédérale signifient que dès l'obtention du contrat l'équipe évaluative externe manque déjà de temps, et la conception de l'évaluation doit se faire rapidement afin d'accorder le plus de temps possible à la collecte et l'analyse des informations. La démarche de cadrage de l'évaluation s'inscrit dans une collaboration étroite et une communication quasi constante entre les gestionnaires de l'évaluation interne et externe, ce qui semble laisser peu de place à l'accompagnement des évaluateurs novices

à travers ce processus. Les évaluateurs experts aimeraient avoir davantage de temps pour travailler en tandem avec les évaluateurs novices, mais les circonstances ne sont malheureusement pas toujours propices. En revanche, à travers de nombreux échanges informels et moments de discussion d'ordre méthodologique dans le cadre de réunions de l'équipe évaluative externe, les évaluateurs experts veillent à expliquer les raisons quant à leurs choix méthodologiques, et prennent le temps de répondre aux questions et de recueillir les suggestions des évaluateurs novices collaborateurs. Si ultimement, dans le présent contexte, les tâches de construction de la matrice évaluative, de l'instrumentation, etc., sont du ressort de l'évaluateur sénior gestionnaire, les évaluateurs juniors collaborateurs impliqués sur le projet depuis le début ont l'occasion de contribuer aux discussions quant au contenu des éléments du devis d'évaluation. Et ils participent également à la révision de l'instrumentation.

Selon nous, assister à la conception de l'évaluation (incluant la rencontre de travail initiale avec les clients), donner sa rétroaction, faire part de ses suggestions méthodologiques, et bénéficier de l'aide et du partage des savoirs expérientiels des évaluateurs experts permet aux jeunes évaluateurs de progresser du stade de novice au stade de débutant avancé sur le plan de leur développement professionnel. Cependant, afin qu'ils puissent progresser au stade de professionnel qualifié, et ainsi être en mesure d'entreprendre seuls les étapes de conception, de gérer et de s'adapter efficacement à des situations contraignantes et potentiellement imprévisibles (Benner et Tanner, 1987 ; Dreyfus et Dreyfus ; 1980), les évaluateurs novices doivent entreprendre des actions de diagnostic et de régulation (McPherson, 2015 ; Tourmen, 2009). Qu'ils s'approprient la démarche de conception, et les différents outils, et qu'ils acquièrent des savoirs opérationnels et tacites. Qu'ils poursuivent l'acquisition de savoir-faire cognitifs en mettant à l'essai les théories spontanées ou implicites qu'ils ont acquises grâce à leurs propres observations, et celles qu'on leur a transmises (explicitations des évaluateurs experts, formation initiale, guides méthodologiques, etc.). Au regard des propos des évaluateurs experts, les tâches de conception initiale des éléments clés du devis d'évaluation semblent être davantage confiées aux évaluateurs en transition vers le statut de *sénior collaborateur* (à partir de 5 ans d'expérience), où ces derniers entrent dans la phase de développement des compétences du volet *gestion de projet*. Ces derniers, dont nous n'avons pas été en mesure d'explorer la pratique, seraient selon la nature et l'ampleur des projets d'évaluation, des cogestionnaires de projets, ce qui inclurait également la communication régulière avec les clients tout au long de la démarche évaluative, le tout sous la supervision générale de

l'évaluateur sénior gestionnaire qui demeure le principal responsable de l'évaluation et de la relation avec les clients. La dimension « relation client » soulève ici un autre point essentiel sur le plan du développement de la capacité des évaluateurs novices à concevoir une démarche d'évaluation appropriée et utile. En effet, la pratique des évaluateurs séniors gestionnaires de l'évaluation externe implique de nombreux temps d'interaction plus ou moins formels avec le chef de l'évaluation interne, voire avec les utilisateurs ciblés, auxquels l'évaluateur junior collaborateur (ou encore d'autres membres de l'équipe évaluative) ne semble pas assister. Notamment : les temps de consultation avec le chef de l'évaluation interne afin de discuter de la faisabilité des choix méthodologiques, les temps de travail avec les gestionnaires de programme pour la définition/validation des indicateurs de mesure de performance et l'identification des sources de données, les temps de rétroaction des parties prenantes clés (membres du comité d'évaluation interne), ou encore les temps de négociation.

La capacité de l'évaluateur à concevoir un devis d'évaluation faisable, fiable et utile, repose sur la collaboration efficace avec les parties prenantes clés (ici, le chef de l'évaluation interne et les utilisateurs ciblés), ces derniers constituant une source d'information et de rétroaction indispensable dans le cadre de l'évaluation de la faisabilité et du cadrage de l'évaluation. En outre, les contraintes situationnelles associées à l'évaluation externe de programmes publics représentent des défis pratiques et méthodologiques importants qui nécessitent que l'évaluateur externe entreprenne un processus de négociations des décisions méthodologiques entre les exigences politico-organisationnelles, les contraintes de temps et d'argent, les problèmes d'accessibilité des données ainsi que la protection de la rigueur scientifique de l'évaluation (Bamberger et Mabry, 2020 ; Rossi, Freeman et Lipsey, 2019). Dans le présent contexte, la ressource interne privilégiée est le chef de l'évaluation à l'interne, et la construction du devis d'évaluation nécessite que l'évaluateur gestionnaire de l'évaluation externe communique régulièrement avec ce dernier. Pour les évaluateurs experts et les évaluateurs novices collaborateurs, il est un allié essentiel, un aiguilleur quant à la faisabilité et la pertinence/justesse des propositions méthodologiques, un agent de liaison/consultation auprès des utilisateurs ciblés, un facilitateur d'accès aux sources d'information, etc. À titre d'autorité décisionnelle et de superviseur général du projet d'évaluation, le gestionnaire de l'évaluation à l'interne est là pour s'assurer que le devis d'évaluation permet de répondre de manière appropriée aux questions évaluatives fondamentales de la politique évaluative fédérale, et prévenir les risques de détournement ou de manipulation de l'évaluation par les hauts

responsables et les gestionnaires de programme. Cependant, ce dernier peut faire preuve de résistance à l'égard de l'équipe évaluative externe, sa présence devenant alors contreproductive, voire une source de mésévaluation (p. ex. imposer le traitement d'une question d'évaluation alors que l'évaluateur expert externe indique que d'après son expérience et les caractéristiques de la situation, les informations nécessaires pour y répondre de manière appropriée ne seront pas accessibles). Parmi les potentielles sources de résistance, on trouve le manque d'expérience de l'évaluateur interne qui peut susciter chez ce dernier un manque de confiance et (ou) de compréhension des propositions méthodologiques de l'évaluateur gestionnaire de l'évaluation externe. Un évaluateur externe qui est d'ailleurs souvent plus expérimenté que lui. Entre alors en jeu la capacité de l'évaluateur externe à faire preuve de crédibilité, à communiquer clairement et efficacement les informations, à créer un climat relationnel de confiance, à reconnaitre et gérer de manière appropriée la manifestation de signes de résistance afin d'assurer la rigueur et l'utilité de l'évaluation (Marchand, 2020 ; Dubeau, Bérubé, Lafantaisie, Coutu et Turcotte, 2018 ; Patton, 2018 ; Stevahn et King, 2016 ; Hawkins, 2012 ; Alkin, 2005 ; Taut et Alkin, 2003 ; Donaldson, Gooler et Scriven, 2002). L'importance pour l'évaluateur d'établir et maintenir une dynamique d'implication constructive avec les parties prenantes clés, un dialogue de qualité, de faire preuve de flexibilité et d'adaptabilité, d'adopter une posture de conseiller, etc., montre le caractère essentiel du volet de compétences *relations interpersonnell*es – la collaboration, la communication, la négociation, la gestion et la résolution de conflit (Marchand, 2020 ; Dumaine, 2018 ; Patton, 2018 ; King et Alkin, 2016). Or, dans un contexte de pratique où la dimension relative à la gestion de la relation avec les clients demeure inconnue ou méconnue des évaluateurs novices, ce dernier n'assistant pas aux nombreux échanges entre les gestionnaires de l'évaluation interne et externe, nous nous interrogeons quant au développement de telles capacités chez les jeunes évaluateurs. On fait face ici à une préoccupation qui rejoint celle de Chouinard et *coll.* (2017) qui dans le cadre de leur étude ont observé le manque de capacité des étudiants-évaluateurs à faire suffisamment preuve d'autorité et de crédibilité auprès des parties prenantes. De telles observations montrent la nécessité de former les jeunes évaluateurs à la dimension sociopolitique de l'évaluation, de les préparer au mieux au rôle de gestionnaire de la relation avec les clients, un rôle essentiel de l'évaluateur de programme.

5.2 Discussion sur les limites et pistes de recherche futures

Bien que l'étude de cas multiples offre des avantages indéniables quant à la richesse des données amassées et qu'elle ouvre la voie à l'étude de phénomènes complexes peu explorés, le potentiel de généralisation des résultats et observations s'avère très limité (Gagnon, 2012).

En effet, compte tenu du fait que cette recherche doctorale fut menée auprès d'un échantillon restreint d'évaluateurs de programme qui exercent dans un contexte très spécifique, il n'y a pas de possibilité de généralisation des résultats obtenus. En outre, la nature descriptive et exploratoire de cette étude nous a permis de dresser des portraits de pratiques des évaluateurs de programme selon leur niveau d'expertise, dans un contexte complexe empreint de contraintes importantes qui sont de potentielles sources d'écueils méthodologiques qui entravent la faisabilité et l'utilité de l'évaluation. Notre analyse des pratiques des évaluateurs de programme soulève des points importants en ce qui a trait au développement professionnel des jeunes évaluateurs face à la sociopolitisation de l'évaluation, dans un environnement où (les premières années) ils travaillent dans l'exécution de tâches en silo. Cette étude ouvre ainsi des pistes de recherche et des pistes pédagogiques intéressantes (que nous explorons à différents moments dans cette section ainsi que la suivante qui porte sur les recommandations).

Les critères d'échantillonnage et les exigences du devis méthodologique ont fortement compliqué le recrutement des participants, ce qui a ultimement conduit cette recherche doctorale à s'inscrire dans un contexte très spécifique, celui des cabinets de conseil dont l'une des principales activités est la conduite d'évaluations de programmes externes, commanditées par des départements ou agences du secteur public canadien (programmes fédéraux à l'échelle du pays, d'une province ou d'un territoire). Un contexte qui ne fut pas sans son lot de contraintes, voire de limites. <u>En effet</u> :

- **Les évaluations qui ont fait l'objet des activités de collecte de données de cette recherche ont été réalisées sous la supervision d'un chef de l'évaluation interne qui semble responsable de la planification initiale de l'évaluation dans la demande de propositions.** Outre son statut d'autorité décisionnelle, il est également le collaborateur privilégié de l'équipe évaluative externe dans la conception de l'évaluation, à titre d'aiguilleur méthodologique, d'agent de liaison entre l'évaluateur sénior gestionnaire de l'évaluation externe et les utilisateurs ciblés. Il aurait donc été pertinent de consulter les évaluateurs internes qui ont travaillé sur les projets d'évaluation qui ont l'objet de l'activité d'explicitation afin d'avoir une appréhension et une compréhension plus complète des caractéristiques des situations évaluatives décrites par les évaluateurs participants.

De plus, la non-consultation des évaluateurs internes limite considérablement la discussion sur l'examen de l'incidence des données de mesure de rendement et des données de gestion financière sur la fiabilité et l'utilité de l'évaluation du rendement (efficacité, efficience et économie), et les écueils que nous avons perçus sur le plan du calibrage de l'évaluation comme prescrit dans la demande de propositions. En effet, rappelons que malgré l'existence d'une mesure de rendement de programme, voire d'un système d'information de gestion, nous avons observé de nombreuses lacunes sur le plan des données de mesure de rendement disponibles pour l'évaluation (évaluations de programme discutées réalisées entre 2013 et 2016). Des difficultés qui sont liées à la nature incomplète, insuffisante ou non fiable des données, à la faiblesse de certains indicateurs de mesure. Une observation qui rejoint celles émises à la suite de l'évaluation de la Politique d'évaluation fédérale de 2009 selon lesquelles :

i. parmi les organisations interrogées sur la période de 2012-2013, la moitié rapportait des données de mesure de rendement insuffisantes. Et si la plus grande majorité des programmes avaient une stratégie de mesure de rendement, l'utilité d'environ la moitié des indicateurs de mesure a été remise en question ;
ii. les résultats de l'enquête évaluative des capacités de 2013-2014 des fonctions de résultats du Secrétariat, ont montré que sur les 33 évaluations approuvées entre avril et septembre 2013, 30 ont connu de gros écueils méthodologiques à cause de l'insuffisance des données de mesure de rendement, même si plus de la moitié des programmes avait implanté une stratégie de mesure du rendement ;
iii. les écueils sur le plan des données ont été signalés par un certain nombre d'agences gouvernementales lors de la soumission annuelle obligatoire par les services d'évaluation d'un rapport sur l'état de la mesure du rendement (Secrétariat du Conseil du Trésor du Canada, 2015).

Tout ceci nous interpelle, car malgré les problèmes reconnus sur le plan de la collecte, la gestion et la structuration des données de mesure de rendement et des données de gestion financière, et de l'incidence que cela peut avoir sur la faisabilité et la fiabilité de l'évaluation, la planification interne de la démarche d'évaluation et des intrants ne semble pas nécessairement contingente de l'état des données de programme. Or, à des fins d'optimisation des ressources et de l'efficacité des évaluations mandatées, les directives de la politique évaluative recommandent notamment qu'une analyse de risque soit réalisée en amont de la planification initiale de

l'évaluation (prescrite dans la demande de propositions). Soit, à l'image de la phase de *scoping-design* entreprise par les évaluateurs experts gestionnaires de l'évaluation à l'externe, on recommande que l'unité d'évaluation interne examine de manière approfondie la qualité et l'étendue des données sur le rendement du programme (et autres données de programme existantes), afin de planifier de manière appropriée (pertinente et faisable) le recours à la collecte de nouvelles données. Aussi, à la lumière des constats émis dans le rapport de mise au point sur l'évaluation de l'efficacité des programmes fédéraux (Bureau du vérificateur général du Canada, 2013) ainsi que le rapport d'évaluation de la Politique d'évaluation de 2009 (Secrétariat du Conseil du Trésor du Canada, 2015), nous pouvons émettre l'hypothèse que les écueils perçus pourraient être dus à des limites sur le plan de la capacité évaluative des unités de l'évaluation des ministères concernés dans le cas des évaluations abordées dans le cadre de cette étude doctorale. Parmi les potentielles limites citées dans le rapport d'évaluation de la Politique évaluative de 2009, on trouve notamment le fait que :

i. les exigences de la politique concernant le calendrier et l'orientation ne donnent pas aux unités d'évaluation ministérielles la marge de manœuvre voulue et nécessaire pour faire pleinement état des besoins en matière de planification des évaluations ;
ii. la rigidité sur le plan de la portée et de la fréquence des évaluations ainsi que le manque de ressources (budgets insuffisants, besoins de main-d'œuvre, etc.) feraient en sorte qu'il est difficile pour les ministères de coordonner la planification des évaluations avec d'autres fonctions de surveillance. Ou encore de les planifier selon la nature du programme et des besoins des utilisateurs ciblés afin de maximiser l'utilité de l'évaluation et alléger le fardeau ;
iii. la manifestation d'un « comportement d'aversion au risque » peut limiter l'utilisation par les ministères des assouplissements prévus par la Politique – soit des stratégies qui facilitent le cadrage de l'évaluation, le calibrage de l'effort évaluatif selon les questions évaluatives fondamentales traitées. Ces assouplissements semblent perçues comme des menaces aux évaluations du Cadre de responsabilisation de gestion (*reddition de compte > faisabilité, fiabilité et utilité de l'évaluation*).

Par ailleurs, Bourgeois (2020) souligne un autre facteur lié à la capacité évaluative des ministères qui pourrait selon nous entrer en jeu ici, à savoir le niveau de compétence et d'expérience des évaluateurs internes. Dans des perspectives de recherche future, il serait donc

intéressant d'examiner les pratiques des évaluateurs novices et experts qui exercent dans les unités d'évaluation ministérielles. Et de se concentrer plus particulièrement sur les tâches relatives à la préparation et au suivi de projets d'évaluation externe afin d'apprécier pleinement leur perception des facteurs contextuels qui exercent une influence sur l'évaluation.

- **Afin d'assurer la comparaison et la fiabilité des résultats de cette étude doctorale, les rapports d'évaluation sélectionnés pour les différentes activités de collecte de données portent sur des projets qui se sont déroulés lorsque la Politique d'évaluation de 2009 du gouvernement fédéral était toujours en vigueur, et non la Politique sur les résultats de 2016.** Il importe de préciser que selon notre compréhension, la nouvelle Politique semble faire preuve d'un allègement sur le plan des exigences et du caractère prescriptif de certains éléments selon la nature des programmes. Ainsi, une agence gouvernementale donnée n'est plus dans l'obligation d'évaluer tous ses programmes, tous les cinq ans, la Politique ayant fixé un seuil moyen quinquennal des dépenses réelles des programmes pour déterminer les attentes en matière d'évaluation. Notons que les exigences restent globalement les mêmes pour les programmes de subventions et de contributions qui se situent au-dessus dudit seuil restent les mêmes que dans le cadre de la Politique de 2009 (obligation d'évaluation quinquennale de la pertinence, de l'efficacité et de l'efficience du programme). En revanche, pour les programmes en dessous du seuil des dépenses, la Politique recommande de réaliser une évaluation périodique afin de rendre compte de l'utilisation, de la pertinence, de l'efficacité, etc. Mais que le choix des objets d'évaluation ou axes de questionnement devait se faire à la lumière de la pertinence pour les buts de l'évaluation et les besoins d'information des utilisateurs ciblés. Par ailleurs, l'examen informel du contenu des cadres méthodologiques (et des limites) communiqué dans une vingtaine de rapports d'évaluations dont le mandat répondait à l'obligation d'évaluation quinquennale de la Politique de 2016 nous a permis d'observer une amélioration sur le plan de la disponibilité des données de mesure de rendement et des données de gestion financière (évaluations réalisées entre 2017 et 2020). Cependant, des problèmes sur le plan de la gestion, de la pertinence et de l'exhaustivité des données de programme semblent perdurer. En effet, plus de la moitié des rapports font mention d'écueils, tels que : le manque d'uniformité des rapports des bénéficiaires de financement, le manque d'exhaustivité des données pour un cycle de financement, la remise en question de l'intégrité de certaines données retranscrites, l'absence de données pour certains indicateurs de mesure, le manque de cohérence dans la saisie des données

d'un programme de subvention à un autre, la nécessité de créer une base de données de mesure de rendement global (à l'échelle du programme), etc.

- **À la lumière du fonctionnement des équipes évaluatives externes, où l'évaluateur sénior associé du cabinet de conseil (expert) est le gestionnaire du projet, et l'évaluateur novice (qu'il soit junior exécutant ou collaborateurs) réalise l'évaluation conçue par ce dernier, il aurait été pertinent de sélectionner les rapports d'évaluation qu'ils avaient en commun pour l'activité d'explicitation. Cela aurait permis d'appréhender les perceptions de chacun sur une même situation évaluative rencontrée.** Cependant, cela compliquait les choses sur le plan du recrutement et (ou) de la passation, puisque :
    - d'une part cela nécessitait que les évaluateurs juniors et séniors qui ont accepté de participer à cette étude partagent un rapport d'évaluation commun et communicable ;
    - et d'autre part il y avait la possibilité qu'un évaluateur sénior soit responsable de deux projets d'évaluation, chacun impliquant un évaluateur junior participant différent, ce qui aurait nécessité qu'il participe à deux entrevues d'explicitation (le devis initial nécessitait déjà que les participants libèrent cinq heures sur leur temps de travail, à minima, ce qui était exigeant compte tenu de leur disponibilité).

  Par ailleurs, la pratique professionnelle des évaluateurs séniors associés du cabinet de conseil semble inclure le soutien du développement professionnel des évaluateurs juniors. Aussi, nous aurions aimé examiner davantage la dimension relative à la formation continue des jeunes évaluateurs sur le terrain, aux occasions d'apprentissages expérientiels, au raisonnement derrière les différents statuts octroyés aux évaluateurs juniors au fur et à mesure de leur développement professionnel, de leur progression. Si nous avons tenté d'aborder certains de ces éléments, la nature de l'activité d'explicitation et son caractère chronophage n'était pas propice à une telle discussion. En outre, afin d'enrichir (et de compléter) les observations émises dans le cadre de cette étude doctorale, et de renforcer la proposition des recommandations sur le plan de la formation continue des jeunes évaluateurs sur le terrain, il serait intéressant, par exemple, d'organiser des groupes d'analyse de pratique « novices-novices », mais aussi « novices-experts ».

Enfin, afin de saisir pleinement les multiples facettes des situations évaluatives rencontrées par les évaluateurs, les démarches diagnostiques et le raisonnement méthodologique, il aurait été

idéal d'observer la pratique des participants directement sur le terrain. D'assister notamment aux rencontres de travail avec les commanditaires, aux échanges avec le chef de l'évaluation interne, aux réunions de l'équipe évaluative externe (discussion méthodologique, préparation de la collecte/analyse des informations, etc.), etc. Et de procéder après la réalisation du projet, à l'activité d'explicitation. De plus, nous l'avons mentionné plus tôt, il aurait été pertinent de choisir des projets d'évaluation communs aux évaluateurs novices et experts participants, et d'inclure des entrevues auprès des évaluateurs internes gestionnaires des projets concernés. En outre, le choix d'analyser la pratique des évaluateurs de programme, novices et experts, en croisant leurs pratiques recommandées (ou bonnes pratiques) et leurs pratiques réelles, constituait pour nous un bon moyen pour renforcer la représentativité des portraits de pratiques dressés dans le cadre de cette étude exploratoire. Et, le recours à deux activités d'analyse réflexive reconnues pour leur contribution à l'analyse de pratiques professionnelles (l'exercice de métaévaluation d'un projet d'évaluation tierce, et l'entrevue d'explicitation à l'appui d'un référentiel d'analyse de l'activité qui a fait l'objet d'une validation de contenu auprès d'experts) représente pour nous l'une des forces de notre devis de recherche qualitative. *Malgré tout, il importe de souligner les potentiels écueils sur le plan des données collectées dans le cadre de l'entrevue d'explicitation.*

- **La nature et l'enchainement connus des activités de recherche (passation 1/métaévaluation d'un devis tierce ; passation 2/entrevue d'explicitation) a potentiellement biaisé le choix du rapport d'évaluation que les évaluateurs participants ont soumis pour analyse dans le cadre de l'activité d'explicitation.** En effet, on ne peut pas exclure ici la dimension relative à la désirabilité sociale, plus particulièrement dans le cas des évaluateurs experts qui avaient un nombre important d'options quant au rapport à nous partager ; et qui à titre de gestionnaires des projets d'évaluation ont probablement été enclins à en choisir un qu'ils jugeaient exemplaire. Cependant, à la lumière de nos observations, ces derniers semblent avoir davantage choisi une évaluation avec des éléments spécifiques qu'ils désiraient mettre en lumière, tant en termes de bonnes pratiques que pour illustrer la complexité, et les enjeux méthodologiques face à la situation évaluative rencontrée et les contraintes qui l'habitaient. Notons que nous avions bien pris le soin d'expliquer les objectifs de l'activité, notamment la volonté de rendre compte de la réalité de terrain, de mettre en exergue le poids des contraintes et les stratégies entreprises pour les appréhender et gérer au mieux, les problèmes méthodologiques rencontrés, etc. Dans un cas, par exemple, cela a signifié le choix d'une évaluation où les séries de données de mesure de

rendement du programme étaient complètes et d'une grande qualité, portaient sur d'excellents indicateurs quantitatifs, et comprenaient des informations administratives et des données financières, ce qui a permis de réaliser une démonstration rigoureuse et empirique du rendement du programme (efficacité, efficience). L'évaluateur souhaitait partager avec nous d'une rare occasion, qui a également permis l'évocation des écueils habituellement rencontrés. Par ailleurs, les vas et viens observés dans le discours des évaluateurs entre ce qu'ils aimeraient faire, ce qu'ils devraient faire, ce qu'ils ont pu faire, ce qu'ils ne peuvent pas faire, ce qui devrait être fait, etc., et pourquoi, qu'on se trouve au niveau de l'analyse critique d'une évaluation tierce, du partage de leurs expériences vécues, ou encore du projet d'évaluation qui fait l'objet de l'entrevue d'explicitation, font toute la richesse des résultats.

- **Les données collectées au cours de l'entretien d'explicitation auprès des évaluateurs juniors exécutants se restreignent à l'exécution des choix méthodologiques opérés par l'évaluateur sénior gestionnaire du projet d'évaluation externe.** Par conséquent, les résultats portaient dans la mesure du possible sur leurs explications des activités méthodologiques auxquelles ils ont pris part et des décisions prises par leurs supérieurs. Dans certains cas, plus précisément chez les évaluateurs exécutants avec le plus d'expérience (voir une expérience antérieure de junior collaborateur), il n'était pas toujours facile d'identifier si les informations transmises étaient fondées sur une connaissance concrète de ce qui s'est réellement passé, ou si elles étaient basées sur des connaissances antérieures, sur ce qu'ils perçoivent comme des procédés systématiques ou encore des pratiques prescrites par le contexte évaluatif. En ce qui concerne le cas de l'évaluateur junior collaborateur, même si la conception de l'évaluation repose sur les épaules de l'évaluateur sénior, sa participation au projet d'évaluation dans son entièreté (incluant les différentes rencontres de travail durant la conception de l'évaluation, la révision de l'instrumentation, la préparation et la coordination des activités de collecte/analyse de données, etc.), lui a permis de nous expliciter avec précision et justesse les choix méthodologiques opérés dans le cadre de la démarche d'évaluation. Ainsi, si ultimement nous n'étions pas en mesure de comparer les pratiques recommandées et les pratiques réelles des évaluateurs novices, les données obtenues nous ont permis de mettre en lumière des éléments très intéressants sur le plan du développement professionnel des jeunes évaluateurs sur le terrain, et leurs besoins en termes d'occasion d'apprentissage expérientiel, et autres sources d'apprentissage.

En guise de « mots de la fin », nous allons aborder quelques éléments de recommandations à visées « pédagogiques » (soutien au développement professionnel sur le terrain).

5.3 Quelques pistes de recommandations sur le plan du soutien au développement professionnel des évaluateurs de programme novices sur le terrain

L'évaluation est une entreprise à la fois technique et sociopolitique et les choix méthodologiques de conception de l'évaluation reposent sur le diagnostic approfondi d'une situation évaluative donnée afin de faire les choix les plus appropriés en contexte. En effet, les résultats de cette étude doctorale montrent que la pertinence du devis d'évaluation repose sur la capacité des évaluateurs à réaliser un diagnostic efficace de la situation évaluative rencontrée afin de revoir et d'opérationnaliser les paramètres de l'évaluation initialement prescrits par les clients, et d'assurer la faisabilité et la qualité de l'évaluation. Pour mener à bien ce diagnostic et concevoir une évaluation tangible et légitime (dont les résultats sont fiables, représentatifs de l'objet d'évaluation, et utiles), les évaluateurs s'appuient sur des savoirs pratiques ancrés en situation réelle, sur l'articulation de théories spontanées ou implicites développées dans et par l'expérience.

Dans le contexte de pratique des évaluateurs participants à cette étude doctorale, l'activité de conception de l'évaluation est du ressort de l'évaluateur expert (évaluateur sénior gestionnaire du projet d'évaluation externe). Dès lors, il importe de : 1) mobiliser diverses stratégies afin de permettre aux évaluateurs novices d'acquérir des compétences méthodologiques, mais aussi de développer leurs capacités à diagnostiquer une situation évaluative pour effectuer les choix méthodologiques les plus appropriés en contexte ; 2) préparer les jeunes évaluateurs à exercer en situation, à mobiliser des stratégies adaptatives d'ordre méthodologique appropriées face aux contraintes situationnelles rencontrées, tout en veillant à la qualité et à l'utilité des évaluations dont ils auront la responsabilité, ou partageront la responsabilité (dans le futur) ; 3) les préparer, sur le lieu de travail, à l'entreprise du cadrage de l'évaluation.

Au demeurant, les résultats montrent que les situations les plus propices à l'apprentissage des évaluateurs novices sur le terrain sont celles où ils participent à l'ensemble des étapes d'un projet d'évaluation et bénéficient du mentorat de l'évaluateur expert gestionnaire du projet. On rejoint ici les observations de Tourmen (2009, 2007) selon lesquelles pour développer chez les évaluateurs débutants la capacité de diagnostic/pronostic de situations évaluatives données et l'acquisition de règles d'action qui permettent l'adaptation, il importe de : 1) leur offrir de

nombreuses occasions de suivre le déroulement entier d'une évaluation, afin qu'ils puissent observer les résultats, en tirer des leçons et comprendre l'enchainement global des étapes ; 2) leur permettre d'assister à la réalisation d'évaluation dans des situations variées ; 3) leur offrir l'aide de professionnels expérimentés pour identifier les indices caractéristiques d'une situation évaluative, pouvoir s'y adapter et anticiper de potentiels problèmes.

Le troisième point souligne le rôle essentiel de la transmission directe, orale, des ressources cognitives sur le lieu de travail avec l'explicitation des théories implicites, des raisonnements, des schèmes opératoires mobilisés par les évaluateurs experts pour prendre des décisions méthodologiques appropriées en contexte (McPherson, 2015 ; Tourmen et Droyer, 2013). Cependant, la réalité de terrain ne semble pas toujours des plus propices pour multiplier et varier les occasions pour les évaluateurs novices d'assister à un projet d'évaluation du début à la fin. De même, elle peut limiter de manière plus ou moins importante les efforts de mentorat entrepris par les évaluateurs experts durant les projets qu'ils dirigent (p. ex. les tentatives de travail en tandem « junior-sénior », l'explicitation et les discussions en temps opportun du raisonnement en amont des diverses décisions méthodologiques opérées, etc.).

Nous proposons deux types d'activités de développement professionnel qui peuvent être mises en place au sein des organismes de travail, mais aussi dans le cadre d'activités proposées par les associations professionnelles (conférence, ateliers de formation continue, etc.) : les groupes d'analyse de pratiques et les analyses de cas. Soit des activités qui mobilisent deux éléments essentiels au développement de savoir-faire cognitifs - la réflexivité et la transmission des savoirs tacites des évaluateurs expérimentés.

L'organisation de groupes d'analyse de pratiques ou de groupes de discussion réflexifs

La conception des savoir-faire cognitifs et des théories implicites repose, entre autres, sur un retour réflexif vis-à-vis des situations rencontrées, d'une sorte de mise en confrontation avec les connaissances théoriques et procédurales transmises. Il y a un va-et-vient entre la théorie et le terrain qui invite à examiner, à remettre en question, à revisiter les concepts et à reformuler des théories (Rog, 2015 ; Christie et Rose, 2003). Il y a également le besoin de rendre explicite et accessible aux praticiens, les savoirs pratiques et les actions implicites mobilisés en situation (Finlay, 2008). Les activités d'analyse de pratiques collaboratives ou de groupes de discussion réflexifs semblent de très bonne occasion pour entreprendre un tel processus. Que ce soit de manière plus ou moins formelle dans le cadre d'échanges avec un superviseur ou un mentor,

comme on l'a vu dans notre étude ou que ce soit dans le cadre de l'animation régulière de groupes d'analyse de pratiques.

L'organisation d'activités d'analyse de pratiques collaboratives, de groupes de discussion réflexifs peut se faire dans une perspective de partage des savoirs expérientiels des évaluateurs expérimentés avec les jeunes évaluateurs, et de retour sur l'activité, de l'évaluation de la pratique après les faits, et des leçons retenues. De telles activités peuvent venir renforcer l'effort d'explicitation et de partage des savoirs ou des théories implicites, des schèmes opératoires et des règles d'action mobilisées en situation par les évaluateurs expérimentés. Elles peuvent faciliter l'analyse, la transmission et la construction/co-construction/reconstruction de stratégies de résolution de problème qui sont situées et efficaces. Il peut s'agir, par exemple, de l'organisation au sein de l'organisme de travail : 1) de séances de réunion-bilan formelles à la fin des projets, où les membres de l'équipe reviennent ensemble sur l'activité ; 2) de faire un retour sur le déroulement du projet, les freins et les leviers rencontrés, les stratégies mises en place pour gérer les contraintes situationnelles, les écueils inévitables, et les leçons apprises ; 3) de permettre aux jeunes évaluateurs de poser des questions, de faire part de leurs observations, etc. L'équipe pourrait réaliser un exercice commun de métaévaluation centrée sur le devis d'évaluation à partir des rapports techniques — rapport méthodologique, matrice évaluative, matrice(s) de résultat(s), limites, etc.

- de groupes de discussion réflexifs réguliers de type « *que faire si ?* » entre les évaluateurs novices et les évaluateurs experts.

Notons que l'organisation de groupes de discussion de type « *que faire si ?* » dans le cadre de conférences, d'ateliers organisés par les associations professionnelles ou encore dans le cadre de baladodiffusions constitue pour nous une ressource pédagogique et professionnelle importante.

L'organisation d'activité de formation continue de type étude de cas

L'enseignement centré sur les cas constitue une stratégie prometteuse de l'enseignement en situation réelle, qui est applicable dans le contexte d'activités de formation continue organisées sur le lieu de travail ou dans le cadre d'ateliers offerts par les associations professionnelles. La modélisation proposée par Bourgeois, Wildschut et Steinberg (2021) (Figure 34, p. 364), offre un aperçu de ses contributions clés à la formation en évaluation, qui renforcent par ailleurs d'autres stratégies recommandées telles que le travail de terrain et le mentorat. Parmi les contributions mises de l'avant, on trouve le fait que :

i. l'enseignement centré sur les cas permet une exposition répétée à une gamme variée de scénarios évaluatifs du monde réel ;
ii. l'enseignement centré sur les cas demande aux apprenants de penser comme des praticiens et examiner les caractéristiques des situations évaluatives rencontrées dans les cas. Dans le cadre de projets d'évaluation, il s'agit de la gestion des contraintes situationnelles que nous avons observées (exigences politico-organisationnelles, besoins des utilisateurs ciblés, problèmes d'accès aux données, intrants limités, etc.) et du caractère faisable, fiable et utile de la méthodologie évaluative mobilisée. Ces mises en situation permettraient aux apprenants de développer de nouvelles stratégies méthodologiques en réponse aux caractéristiques situationnelles, en s'appuyant sur diverses approches évaluatives et méthodes, mais aussi sur les expériences de leur formateur et les propositions de leurs pairs ;
iii. la multiplicité et la diversité des cas associées à la discussion de groupe sur des questions complexes et les réponses possibles optimiseraient selon nous le développement des ressources cognitives et des capacités des jeunes évaluateurs pour exercer en situation réelle.

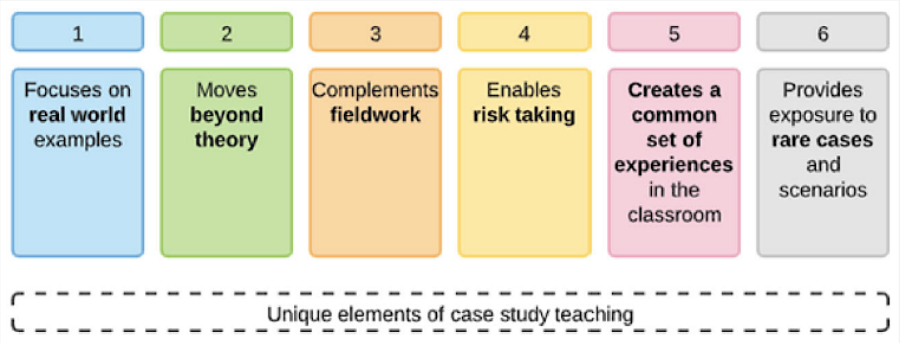


Figure 34. *Les contributions de l'enseignement centré sur les cas (Bourgeois, Wildschut et Steinberg, 2021, p.41)*

Enfin, l'étude exploratoire menée par Bourgeois, Wildschut et Steinberg (2021) auprès d'enseignants et de formateurs en évaluation de programme, montre que ces derniers reconnaissent la contribution de cette approche ainsi que son fort potentiel sur le plan du développement des compétences liées aux volets — pratique professionnelle, pratique contextuelle (ou *analyse situationnelle*), habiletés techniques/méthodologiques, et pratique de la gestion.

Références

Abma, T. A. (2000). Stakeholder Conflict: A Case Study. *Evaluation and Program Planning, 23*, 199-210.

Alexandre, M. (2013). La rigueur scientifique du dispositif méthodologique d'une étude de cas multiple. *Recherche qualitative, 32*(1), 26-56.

Alkin, M. C. (2011). Evaluation Essentials: From A to Z. *Overview Chart*. New York, NY : The Guilford Press.

Alkin, M. C. (2005). Utilization of evaluation. Dans S. Mathison (dir.), *Encyclopedia of evaluation* (p.435-439). Thousand Oaks, CA : Sage.

Alkin, M. C. (1990). *Debates on evaluation*. Newbury Park, CA: Sage.

Alkin, M. et Coyle, K. (1988). Thoughts on evaluation misutilization. *Studies in Educational Evaluation, 14*, 331–340.

Alkin, M. C., Hofstetter, C. H. et Ai, X. (1998). Stakeholder concepts in program evaluation. Dans A. Reynolds et H. Walberg (dir.), *Advances in educational productivity* (p. 87-113). Greenwich, CT: JAI Press.

Alkin, M. C. et King, J. A. (2017). Definition of Evaluation Use and Misuse, Evaluation Influence, and Factors Affecting Use. *American Journal of Evaluation, 38*(3), 434-450.

Alkin, M. C. et King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation, 37*, 568–579.

American Evaluation Association (2014). What is evaluation? Récupéré sur site de l'AEA : https://www.eval.org/Portals/0/What%20is%20evaluation%20Document.pdf

Anadòn, M. (2006). La recherche dite « qualitative » : de la dynamique de son évolution aux acquis indéniables et aux questionnements présents. *Recherche qualitative, 26*(1), 5-31.

Anadón, M. et Guillemette, F. (2007). La recherche qualitative est-elle nécessairement inductive? *Recherches Qualitatives,* Hors-série, 5, 26-37.

Anadón, M. et Savoie-Zajc, L. (2009). L'analyse qualitative des données. *Recherches qualitatives, 28*(2), 1-7.

Bamberger, M. et Mabry, L. (2020). *RealWorld Evaluation. Working under Budget, Time, Data, and Political Constraints* (3e éd.). Thousand Oaks, CA: Sage.

Bamberger, M., Rugh, J. et Mabry, L. (2012). *RealWorld Evaluation. Working under Budget, Time, Data, and Political Constraints* (2e éd.). Thousand Oaks, CA: Sage.

Baribeau, C. et Royer, C. (2012). L'entretien individuel en recherche qualitative : usages et modes de présentation dans la Revue des sciences de l'éducation. *Revue des sciences de l'éducation, 38*(1), 23-45.

Baxter, P. et Jack, S. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report, 13*(4), 544-559.

Beaudoin, A., Lefrançois, R. et Ouellet, F. (1986). Les pratiques évaluatives : enjeux, stratégies et principes. *Service social, 35*(1-2), 188-213.

BetterEvaluation (2014). *BetterEvaluation Rainbow Framework and Planning Tool*. Récupéré de : https://www.betterevaluation.org/tools-resources/rainbow-framework-brief-overview

Binet, A. (1916). New Methods for the Diagnosis of the Intellectual Level of Subnormals. Dans E. S. Kite (trad.), *The development of intelligence in children*. Vineland, NJ : Publications of the Training School at Vineland. (Original publié en 1905 dans *L'Année Psychologique, 12*, 191-244.).

Bourgeois, I., Wildschut, L. et Steinberg, M. (2021). How does teaching with cases support the development of evaluation competencies? *New Directions for Evaluation*, *2021*(172), 37-51.

Brandon, P. R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation, 19*, 325–337.

Bryson, J. M., Patton, M. Q. et Bowman, R.A (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning, 34*, 1-12.

Bureau du vérificateur général du Canada (2013). *Rapport du vérificateur général du Canada. Chapitre 1 : rapport Le Point sur l'évaluation de l'efficacité des programmes*. Récupéré sur le site du Gouvernement du Canada : https://publications.gc.ca/site/fra/439144/publication.html

Chen, H. T. (2015). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2e éd.) (p. 1-94). Thousand Oaks, CA : Sage.

Chen, H.-T. (1994). Theory-driven evaluations: Need, difficulties, and options. *Evaluation Practice, 15*(1), 79-82.

Chouinard, J. A., Boyce, A. S., Hicks, J., Jones, J., Long, J., Pitts, R. et Stockdale, M. (2017). Navigating Theory and Practice Through Evaluation Fieldwork: Experiences of Novice Evaluation Practitioners. *American Journal of Evaluation*, 38(4), 493-506.

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluations, 97*, 7-35.

Christie, C. A. et Fleischer, D. N. (2010). Insight into evaluation practice: A content analysis of methods used in evaluation studies published in North American evaluation focused journals. *American Journal of Evaluation, 31,* 326–346.

Christie, C. A. et Rose, M. (2003). Learning about evaluation through dialogue: Lessons from an informal discussion group. *American Journal of Evaluation, 24*, 235–243.

Christie, C. A. et Alkin, M. C. (1999). Further reflections on evaluation misutilization. *Studies in Educational Evaluation, 25*, 1-10.

Cook, T. D., Scriven, M., Coryn, C. L. S. et Evergreen, S. D. H. (2010). Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven. *American Journal of Evaluation, 31*, 105–117.

Cook, T. D. (1997). Lessons Learned in Evaluation over the Past 25 Years. Dans E. Chelimski et W. R. Shadish (dir.), *Evaluation for the 21st Century* (p. 30-52). Thousand Oaks, CA : Sage.

Cousins, J. B. (2004). Commentary: Minimizing Evaluation Misuse as Principled Practice. American Journal of Evaluation, 25(3), 391-397

Cousins, J. B. (2003). Utilization effects of participatory evaluation. Dans T. Kelligan et D. L. Stufflebeam (dir.), *International Handbook of Educational Evaluation* (p. 245-266). Dordrecht, the Netherlands: Springer.

Cousins, J. B. et Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*, 397-418.

Cousins, J.B. et Leithwood, K.A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research, 56*, 331-365.

Cousins, J. B., Whitmore, E. et Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation, 34*, 7-22.

Cousins, J. B. et Whitmore, E. (1998). Framing participatory evaluation. Dans E. Whitmore (dir.), *Understanding and practicing participatory evaluation. New Directions in Evaluation, 80* (p. 3–23). San Francisco, CA : Jossey Bass.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, *64*, 672-683.

Darabi, A. (2002). Teaching Program Evaluation: Using a System Approach. *American Journal of Evaluation*, 23(2), 219-228.

Datta, L.-E. (2001). Coming Attractions. *American Journal of Evaluation, 22*(3), 403-8.

Denzin, N. K. et Lincoln, Y. S. (2011). Introduction: The Discipline and Practice of Qualitative Research. Dans N. K. Denzin et Y.S. Lincoln (dir.), *The SAGE Handbook of Qualitative Research* (5e éd.) (p. 1-20). Thousand Oaks, CA: Sage.

Dillman, L. M. (2012). Evaluator Skill Acquisition: Linking Educational Experiences to Competencies. *American Journal of Evaluation, 34*(2), 270-285.

Donaldson, S. I. et Christie, C. A. (2005). The 2005 Claremont debate: Lipsey versus Scriven. Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation, 2*(3), 60–77.

Donaldson, S. I., Gooler, L. E. et M. Scriven (2002). Strategies for Managing Evaluation Anxiety: Toward a Psychology of Program Evaluation. *American Journal of Evaluation, 23*(3), 261-74.

Donaldson, S. I. et Scriven, M. (2002). Diverse visions for program evaluation: Should we integrate or embrace diversity? Dans S. I. Donaldson et M. Scriven (dir.), *Evaluating social programs and problems: Visions for the new millennium* (p. 3-16). Mahwah, NJ : Erlbaum.

Dubeau, D., Bérubé, A., Lafantaisie, V., Coutu, S. et Turcotte, G. (2018). Les conditions gagnantes et les défis pour une participation active des parties prenantes dans un contexte d'évaluation de programmes d'intervention. *Revue canadienne en évaluation de programme, 33*(2), 202-222.

Dubois, C. A, Champagne, F. et Bilodeau H. (2011). Historique de l'évaluation. Dans A. Brousselle, F. Champagne, A-P. Contandriopoulos et Z. Hartz (dir.), *L'évaluation. Concepts et méthodes* (2e éd.) (p. 28-48). Montréal, Québec : Les Presses de l'Université de Montréal.

Dufays, J.-L. (2001). Quelle(s) méthodologie(s) pour les recherches en didactique de la littérature ? Esquisse de typologie et réflexions exploratoires. *Enjeux,* (51-52), 7-29.

Dumaine, F. (2018). La traversée du funambule : l'engagement dans la neutralité. Dans M. Hurteau, I. Bourgeois et D. Houle (dir.), *L'évaluation de programme axée sur la rencontre des acteurs* (p. 203-222). Québec, Québec : Les Presses de l'Université du Québec.

Fetterman, D. M. (2005). Empowerment evaluation. Dans S. Mathison (dir.), Encyclopedia of evaluation (p. 125–129). Thousand Oaks, CA: Sage.

Fetterman, D. M. (1994). Empowerment evaluation. *Evaluation Practice, 15*, 1–15.

Finlay, L. (2008) Reflecting on "Reflective Practice. *Practiced-based Professional Learning Paper 52,* 1–27. Récupéré de The Open University : https://oro.open.ac.uk/

Fishman, D. B. (1992). Postmodernism Comes to Program Evaluation: A Critical Review of Guba and Lincoln's Fourth Generation Evaluation. *Evaluation and Program Planning, 15,* 263-70.

Fitzpatrick, J. L. (2012). Commentary: Collaborative evaluation within the larger evaluation context. *Evaluation and Program Planning, 35*, 558-563.

Fitzpatrick, J. L., Christie, T. A. et Mark, M. M. (dir.) (2008). *Evaluation in action: Interviews with Expert Evaluators* (p. 1-17). Thousand Oaks, CA: Sage.

Fournier, D. M. (2005). Evaluation. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 140). Thousand Oaks : Sage.

Fleischer, D. N. et Christie, C. A (2009). Evaluation use: Results from a survey of U.S American Evaluation Association members. *American Journal of Evaluation, 30*, 158–175.

Freeman, R. E. (1984). *Strategic management: a stakeholder approach*. Boston, MA : Pitman.

Gagnon, Y.-C. (2011). *L'étude de cas comme méthode de recherche* (2e éd.) (p. 1-54). Québec, Québec : Les presses de l'Université du Québec.

Geva-May, I. et Thorngate, W. (2003). Reducing Anxiety and Resistance in Policy and Programme Evaluations: A Socio-Psychological Analysis. *Evaluation, 9*(2), 205-227.

Ghere, G., King, J. A., Stevahn, L. et Minnema, J. (2006). A Professional Development Unit for Reflecting on Program Evaluator Competencies. *American Journal of Evaluation, 27*(1), 108-123.

Greene J. C. (2015). The Emergence of Mixing Methods in the Field of Evaluation. *Qualitative Health Research, 25*(6), 746-750.

Greene, J. C. (2006). Evaluation, democracy, and social change. Dans I. F. Shaw, J. C. Greene et M. M. Mark (dir.), *The sage handbook of evaluation* (p. 118-160). Thousand Oaks, CA : Sage.

Greene, J. C. (2005a). Context. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 83-85). Thousand Oaks, CA : Sage.

Greene, J. C. (2005b). Stakeholder Involvement. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 398). Thousand Oaks, CA: Sage.

Greene, J. C. (2005c). Mixed Method. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 255-256). Thousand Oaks, CA: Sage.

Greene, J. C. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review, 12*, 91-116.

Greene, J. C., Lipsey, M. W., Schwandt, T. A., Smith, N. L. et Tharp, R. G. (2007). Method choice: Five discussant commentaries. *New Directions for Evaluation, 113*, 111-127.

Guba, E. G. et Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA : Sage.

Holbrook, A. M. et Chen, W.-Y. (2017). Learning by doing: an experiential approach to program evaluation. *Social Work Education, 36*(1), 62–74.

Hatry, H.-P, Newcomer, K. E. et Wholey, J. S. (2015). Evaluation Challenges, Issues, and Trends. Dans K. E. Newcomer, H.-P. Hatry et J. S. Wholey (dir.), *Handbook of Practical Program Evaluation* (4e éd.) (p. 816-833). San Francisco, CA : Jossey-Bass.

Hawkins, P. (2012). Successful evaluation management: Engaging mind and spirit. *Canadian Journal of Program Evaluation, 25*(3), 27-36.

Huberman, A. M. et Miles, M. B. (1994). Data management and analysis methods. Dans N. K. Denzin et Y. S. Lincoln (dir.), *The SAGE Handbook of Qualitative Research* (1re éd.) (p. 428-444). Thousand Oaks, CA : Sage.

Hurteau, M. et Houle, S. (2012). La production du jugement crédible : une synthèse des contribution. Dans M. Hurteau, S. Houle et F. Guillemette (dir.), *L'évaluation de programme axée sur le jugement crédible* (p. 175-192). Québec, Québec : Presses de l'Université du Québec.

Hurteau, M. et Houle, S. (2005). Évaluation de programme et recherche évaluative : des activités distinctes. *Mesure et évaluation en éducation, 28*(3), 83-95.

Hurteau, M., Houle, S., Marchand, M.-P., Ndinga, P., Guillemette, F. et Schleifer, M. (2012). Les processus de production et de crédibilisation du jugement en évaluation. Dans M. Hurteau, S. Houle et F. Guillemette (dir.), *L'évaluation de programme axée sur le jugement crédible* (p. 77-100). Québec, Québec : Presses de l'Université du Québec.

Karsenti, T. et Demers, S. (2018). L'étude de cas. Dans T. Karsenti et L. Savoie-Zajc (dir.), *La recherche en éducation. Étapes et approches* (4e éd.) (p. 291-316). Montréal, Québec : Presses de l'Université de Montréal.

King, J. A. (2007). Making sense of participatory evaluation. Dans S. Mathison (dir.), *Enduring issues in evaluation: The 20th anniversary of the collaboration between NDE and AEA: Vol. 114*. *New Directions in Evaluation* (p. 83-86). San Francisco, CA: Jossey Bass.

King, J. A. (2003). The challenge of studying evaluation theory. *New Directions for Evaluation, 97*, 57-68.

King, J. A. et Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. Thousand Oaks, CA: Sage.

King, J.A., Stevahn, L., Ghere, G. et Minnema, J. (2001). Toward a Taxonomy of Essential Evaluator Competencies. *American Journal of Evaluation, 22*, 229-247.

Kinney-Walker, A. (2015). *The Role of Student Identity in Upward Bound Graduates Persistence in College* (thèse de doctorat, University of Nebraska-Lincoln, États-Unis). Récupéré de Educational Administration: Theses, Dissertations, and Student Research. 225. http://digitalcommons.unl.edu/cehsedaddiss/225

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. Dans V. Caracelli et H. Preskill (dir.), *The expanding scope of evaluation use*. *New Directions for Evaluation, 88*, 5-23.

Kundin, D. M. (2010). A Conceptual Framework of How Evaluators Make Everyday Practice Decisions. *American Journal of Evaluation, 31*(3), 347-362.

Lamari, M., Poulin-Larivière, L. et Lemieux K. (2016). Praxis de l'évaluation de l'action gouvernementale. Dans M. Lamari et J. L. Jacob (dir.), *Praxis de l'évaluation et de la révision des programmes publics : approches, compétences et défis* (p. 1-32). Québec, Québec : Presse de l'Université du Québec.

LaVelle, J. M. (2019). Educating evaluators 1976–2017: An expanded analysis of university-based evaluation education programs. *American Journal of Evaluation, 41*(4), 494–509.

Le Boterf, G. (2011). Apprendre à agir et interagir en professionnel compétent et responsable. *Éducation permanente, 188*, 97-112.

Le Boterf, G. (2002). *Ingénierie et évaluation des compétences* (4e éd.) (p. 44-63). Paris, France : Éditions d'Organisation.

Lemaire, D. (2016). Trente ans d'institutionnalisation de l'évaluation au gouvernement fédéral : quel bilan? Dans M. Lamari et J. L. Jacob (dir.), *Praxis de l'évaluation et de la révision des programmes publics : approches, compétences et défis* (p. 331-346). Québec, Québec : Presse de l'Université du Québec.

Levin-Rozalis, M. et Rosenstein, B. (2003). A mentoring approach to the One-year evaluation course. *American Journal of Evaluation, 24(2)*, 245-259.

Marchand, M.-P. (2020). *Vers une meilleure contextualisation de l'évaluation crédible : identification des facteurs d'influence de la crédibilité à partir d'expériences évaluatives de parties prenantes* (Thèse de doctorat, Université du Québec à Montréal, Canada). Récupéré de l'archive de publications électroniques de l'UQAM : https://archipel.uqam.ca/14084/

Mark, M. M. (2001). Evaluation's Future: Furor, Futile, or Fertile? *American Journal of Evaluation, 22*(3), 457-79.

Martinez, C. (1997). L'entretien d'explicitation comme instrument de recueil de données. *Les journaux du GREX,* (21), 2-7.

McDavid, J. C., Huse, I. et Hawthorn, L. R. L. (2019). *Program evaluation and performance measurement: an introduction to practice* (3e éd.). Thousand Oaks, CA : Sage.

McPherson, C. L. (2015). *Implicit and Explicit Evaluation Theories of Expert Evaluation Practitioners* (thèse de doctorat, University of South Alabama, États-Unis). Disponible sur ProQuest Dissertations and Thesis (Publication No 3737837).

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education. Revised and expanded from "Case Study Research in Education"* (p. 151-219). San Francisco, CA : Jossey-Bass.

Mertens, D. M. (2015). Philosophical assumptions and program evaluation. *SpazioFilosofico, 15*, 75-85.

Miles, M. B. et Huberman, A. M. (2003). *Analyse des données qualitatives: recueil de nouvelles méthodes* (M. H. Rispan, trad. et J.-J. Bonniel, rév.) (p. 307-436). Bruxelles, Belgique : De Boeck. (Original publié en 1994 dans *Qualitative Data Analysis: An Expended Sourcebook* (2e éd.). Thousand Oaks, CA : Sage.

Monsen, L. (2002). School-based Evaluation in Norway: Why Is It So Difficult to Convince Teachers of Its Usefulness? Dans D. Nevo (dir.), *School-based Evaluation: An International Perspective* (p. 73-88). Oxford, UK : JAI Elsevier Science.

Morris, D. B. (2002). The inclusion of stakeholders in evaluation: benefits and drawbacks. *Revue Canadienne d'évaluation de programme, 17*(2), 49-58.

Mucchielli, A. (2007). Les processus intellectuels fondamentaux sous-jacents aux techniques et méthodes qualitatives. *Recherches qualitatives,* Hors-série, 3, 1-27.

Patton, M. Q. (2018). L'intervention auprès des parties prenantes : le rôle et les responsabilités de l'évaluateur à titre d'animateur. Dans M. Hurteau, I. Bourgeois et S. Houle (dir.), *L'évaluation de programme axée sur la rencontre des acteurs* (p. 95-113). Québec, Québec : Presses de l'Université du Québec.

Patton, M. Q. (2015a). Misuse: The shadow side of use. Dans C. A. Christie et A. T. Vo (dir.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (p. 131-147). Charlotte, NC: Information Age.

Patton, M. Q. (2015b). *Qualitative research and evaluation methods* (3e éd). Thousand Oaks, CA : Sage.

Patton, M. Q. (2012). A utilization-focused approach to contribution analysis. *Evaluation, 18*(3), 364-377.

Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4e éd.) (p. 3-58). Thousand Oaks, CA : Sage.

Patton, M. Q. (2005). Misuse of Evaluations. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 255-256). Thousand Oaks, CA : Sage.

Patton, M. Q. (1997). *Utilization-Focused Evaluation* (3e éd.). Thousand Oaks, CA : Sage.

Patton, M. Q. (1982). *Practical Evaluation* (p. 11-30). Thousand Oaks, CA : Sage.

Perret, V. et Séville, M. (2003). Fondements épistémologiques de la recherche. Dans R.-A. Thiétart (dir.), *Méthodes de recherche en management* (2e éd.) (p. 13-33). Paris, France : Dunod.

Poth, C., Lamarche, M. K., Yapp, Sulla, E. et Chisamore, C. (2014). Toward a Definition of Evaluation Within the Canadian Context: Who Knew This Would Be So Difficult? *Revue canadienne d'évaluation de programme, 29*(3), 87-103.

Poupart, J. (2012). L'entretien de type qualitatif : réflexion de Jean Poupart sur cette méthode. À partir des propos recueillis et rassemblés par Nadège Broustau et Florence Le Cam. *Sur le journalisme, About journalism, Sobre jornalismo, 1*(1), 60-71.

Poupart, J. (1997). L'entretien de type qualitatif : considérations épistémologiques, théoriques et méthodologiques. Dans J. Poupart, J.-P. Deslauriers, L.-H. Groulx, A. Laperrière, R. Mayer et A. P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques* (p. 173-209). Bourcheville, Québec : Gaëtan Morin.

Pourtois, J.-P. et Desmet, H. (1988). *Épistémologie et instrumentation en sciences humaines*. Bruxelles, Belgique : Mardaga.

Preskill, H. et Torres, R. T. (1999). Building Capacity for Organizational Learning through Evaluative Inquiry, *Evaluation 5*(1), 42-60.

Ridde, V. et Dagenais, C. (2012). Introduction générale à l'évaluation de programme. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme* (2e éd.) (p. 12-32). Montréal, Québec : Presses de l'Université de Montréal.

Rog, D. J. (2005). Design, Evaluation. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 115-117). Thousand Oaks, CA : Sage.

Rossi, P. H., Freeman, H. E. et Lipsey, M. (dir.) (2019). *Evaluation: A systematic approach* (8e éd.) (p. 32-82). Thousand Oaks, CA: Sage.

Rossi, P. H., Freeman, H. E. et Lipsey, M. (dir.) (2004). *Evaluation: A systematic approach* (7e éd.). Thousand Oaks, CA: Sage.

Roy, S. N. (2009). L'étude de cas. Dans B. Gauthier (dir.), *Recherche sociale de la problématique à la collecte de données* (5e éd.) (p. 199-225). Québec, Québec : Presses de l'Université du Québec.

Sanders, J. R. et Nafziger, D. N. (2011). A Basis for Determining the Adequacy of Evaluation Designs. *Journal of Multidisciplinary Evaluation, 7*(15), 44-78.

Savoie-Zajc, L. (2018). La recherche qualitative/interprétative en éducation. Dans T. Karsenti et L. Savoie-Zajc (dir.), *La recherche en éducation. Étapes et approches* (4e éd.), (p. 191-218). Montréal, Québec : Presses de l'Université de Montréal.

Savoie-Zajc, L. (2009). L'entrevue semi-dirigée. Dans B. Gauthier (dir.), *Recherche sociale de la problématique à la collecte de données* (5e éd.) (p. 337-360). Québec, Québec : Presses de l'Université du Québec.

Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, Hermeneutics and Social Constructivism. Dans N. K. Denzin et Y. S. Lincoln (dir.), *The SAGE Handbook of Qualitative Research* (2e éd.) (p. 189-214). Thousand Oaks, CA: Sage.

Scriven, M. (2016). Roadblocks to recognition and revolution. *American Journal of Evaluation, 7*(1), 27-44.

Scriven, M. (2005). Metaevaluation. Dans S. Mathison (dir.), *Encyclopedia of Evaluation* (p. 250-252). Thousand Oaks, CA: Sage.

Scriven, M. (1997). Truth and objectivity in evaluation. Dans E. Chelimsky et W.R. Shadish (dir.), *Evaluation for the 21st century: A handbook* (p. 477–500). Thousand Oaks, CA: Sage.

Scriven, M. (1991). *Evaluation Thesaurus*. Thousand Oaks, CA : Sage.

Secrétariat du Conseil du Trésor du Québec (2018). *Cadre d'évaluation : instruction destinée aux ministères et aux organismes*. Récupéré sur le site du Gouvernement du Québec : https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/cadre_evaluation.pdf

Secrétariat du Trésor du Canada (2016). *Politique sur les résultats*. Récupéré sur le site du Gouvernement du Canada : https://www.tbs-sct.canada.ca/pol/doc-fra.aspx?id=31300

Secrétariat du Conseil du Trésor du Canada (2015). *Calibrage des évaluations – Facteurs à prendre en considération*. Récupéré sur le site du Gouvernement du Canada : https://www.canada.ca/fr/secretariat-conseil-tresor/services/verifications-evaluations/calibrage-evaluations.html

Secrétariat du Conseil du Trésor du Canada (2015). *Évaluation de la Politique sur l'évaluation de 2009*. Récupéré sur le site du Gouvernement du Canada : https://www.canada.ca/fr/secretariat-conseil-tresor/services/verifications-evaluations/centre-excellence-en-evaluation/evaluation-politique-evaluation-2009.html#toc-find-1

Secrétariat du Conseil du Trésor du Canada (2009). Directive sur la fonction d'évaluation. Récupéré sur le site du Gouvernement du Canada : https://www.tbs-sct.canada.ca/pol/doc-fra.aspx?id=15681

Secrétariat du Conseil du Trésor du Canada (2009). *Politique sur l'évaluation*. Récupéré sur le site du Gouvernement du Canada : https://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=15024&section=html

Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, *19*(1), 1-19.

Shadish, W. R. (1995). Philosophy of science and the quantitative-qualitative debates: thirteen common errors. *Evaluation and Program Planning, 18*(1), 63-75.

Shadish, W. R, Cook, T. D. et Leviton, L. C. (dir.) (1991). *Foundations of Program Evaluation: Theories of Practice* (p. 17-36). Thousand Oaks, CA: Sage.

Shaw, I. G. R., Greene, J. C. et Mark, M. M. (dir.). (2006). *The SAGE Handbook of Evaluation* (p. 1-30). Thousand Oaks, CA: Sage.

Smith, N. L. (2008). Fundamental Issues in Evaluation. Dans N. L. Smith et P. R. Brandon (dir.), *Fundamental Issues in Evaluation* (p. 1-25). New York, NY : The Guilford Press.

Smith, N. L. et Brandon, P. R. (dir.) (2008). *Fundamental Issues In Evaluation* (p. i-xvi). New York, NY : The Guilford Press.

Société canadienne d'évaluation (2018). *Référentiel des compétences professionnelles requises à l'exercice d'évaluation de programmes au Canada*. Récupéré sur le site de la SCÉ : https://evaluationcanada.ca/txt/2_competences_cdn_pratique_evaluation.pdf

Société canadienne d'évaluation (2005). *Qu'est-ce que l'évaluation ?* Récupéré sur le site de la SCÉ : https://evaluationcanada.ca/sites/default/files/def_evaluation_sce_201510.pdf

Stake, R. (1998). Case Studies. Dans N. K. Denzin et Y. S. Lincoln (dir.), *Strategies of Qualitative Inquiry*. Thousand Oaks, CA: Sage.

Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage.

Stevahn, L. et King, J. A. (2016). Evaluation and facilitation. Facilitating interactive evaluation practice: Engaging stakeholders constructively. *New Directions for Evaluation, 149*, 67-80.

Stevahn, L., King, J. A., Ghere, G. et Minnema, J. (2005). Establishing Essentials Competencies for Program Evaluators. *American Journal of Evaluation, 26*(1), 43-59.

Stufflebeam, D. L. (2013). The CIPP Evaluation Model: Status, Origin, Development, Use, Theory. Dans M. Alkin (dir.), *Evaluation Roots. A Wider Perspective of Theorists' Views and Influences* (2e éd.). Thousand Oaks, CA: Sage.

Stufflebeam, D. L. (2001). The Metaevaluation Imperative. *American Journal of Evaluation, 22*, 183–209.

Stufflebeam, D. L. (1978). Metaevaluation: An overview. *Evaluation & the Health Professions, 1*(2), 146–163.

Stufflebeam, D. L. (1971). The Relevance of the CIPP Evaluation Model for Educational Accountability. Paper presented at the Annual meeting of the American Association of School Administrators (Atlantic City, N.J., February 24, 1971).

Stufflebeam, D. L. et Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2e éd.) (p. 1-105). San Francisco, CA: Jossey-Bass.

Stufflebeam, D. L. et Madaus, G. F. (2000). Program Evaluation: An historical Overview. Dans D. L. Stufflebeam, G. F. Madaus et T. Kellaghan (dir.), *Evaluation models: Viewpoints on educational and human services evaluation* (2e éd.) (p. 3-10). Norwell, MA: Kluwer.

Stufflebeam, D. L. et Shinkfield, A. J. (2007). *Evaluation theory, models, and applications* (p. 1-61). San Francisco, CA: Jossey-Bass.

Taut, S. (2008). What have we learned about stakeholder involvement in program evaluation? *Studies in Educational Evaluation, 34,* 224-230.

Taut, S. et Alkin, M.C. (2003). Program Staff Perceptions of Barriers to Evaluation Implementation. *American Journal of Evaluation, 24*(2), 213-226.

Taut, S. et Brauns, D. (2003). Resistance to Evaluation: A Psychological Perspective. *Evaluation, 9*(3), 247-264.

Thiebaut, G.-C., Brousselle, A., Contandriopoulos, A.-P., Champagne, F. et Hartz, Z. (2011). Introduction. Positionner l'évaluateur face aux enjeux de l'évaluation. Dans A. Brousselle, F. Champagne, A-P. Contandriopoulos et Z. Hartz (dir.), *L'évaluation. Concepts et méthodes* (2e éd.) (p. 11-23). Montréal, Québec : Les Presses de l'Université de Montréal.

Tourmen, C. (2009). Evaluator's decision making: the relationship between theory, practice, and experience. *American Journal of Evaluation, 22*(1), 71-79.

Tourmen, C. (2008). Les compétences des évaluateurs de politiques publiques. *Formation emploi, 104*, 53-65.

Tourmen, C. (2007). *Les compétences des évaluateurs : le cas des évaluateurs de politiques publiques* (thèse de doctorat non publiée). Université Pierre Mendès-France, Grenoble, France.

Tourmen, C. et Droyer, N. (2013). Dialogues entre théories spontanées et académiques de l'évaluation. *Mesure et Evaluation en Education, 36*(3), 25-49.

Brown Urban, J. et Trochim, W. (2009). The Role of Evaluation in Research—Practice Integration Working Toward the ''Golden Spike''. *American Journal of Evaluation*, *30*(4), 538-553.

Van der Maren, J.-M. (2010). La maquette d'un entretien. Son importance dans le bon déroulement de l'entretien et dans la collecte de données de qualité [Notes de chercheurs en méthodologies qualitatives]. *Recherches qualitatives, 29*(1), 129-139.

Vermersch, P. (2004). Aide à l'explicitation et retour réflexif. *Éducation Permanente,* (160), 71-80.

Vermersch, P. (2003). *L'entretien d'explicitation* (2e éd.). Paris, France : ESF.

Vermersch, P. (1994). *L'entretien d'explicitation*. Paris, France : ESF.

Vermersch, P. (1991). L'entretien d'explicitation. *Les cahiers de beaumont,* (52bis), 63-70.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation, 19*(1), 21-33.

Widmer, T., Lamber, C. et Bachman, N. (2000). *Standards d'évaluation de la Société suisse d'évaluation (Standards SEVAL)*. Récupéré du site de la Société suisse d'évaluation, section Standards SEVAL : http://www.seval.ch/fr/Standards/index.cfm

Yarbrough, D. B., Shulha, L. M., Hopson, R. K. et Caruthers, F. A. (2011). *The Program Evaluation Standards. A Guide for Evaluators and Evaluation Users* (3e éd.). Approuvé par le Joint Committee on Standards for Educational Evaluation (JCSEE). Thousand Oaks, CA : Sage.

Yin, R. K. (1994). *Case Study Research: Design and Methods* (2e éd.). Thousand Oaks, CA : Sage.

Yin, R. K. (2003). *Application of Case Study Research* (2e éd.). Thousand Oaks, CA : Sage.

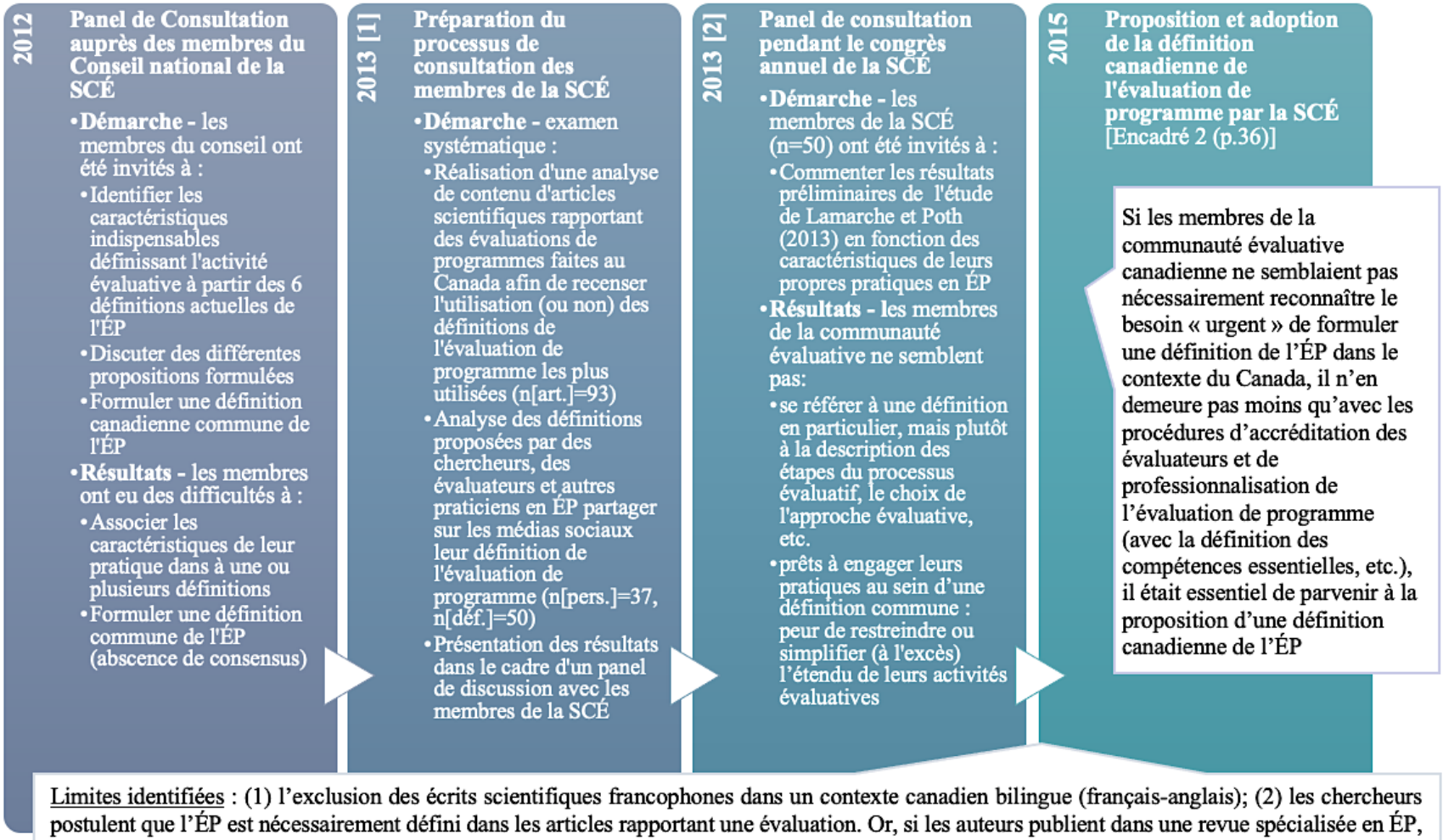
ANNEXES

Annexe 1. *Le pragmatisme et les approches évaluatives « participatives » contemporaines*

Quelques théoriciens pragmatistes (Newcomer, 2015 ; Christie et Alkin, 2013, 2014 ; Martens et Wilson, 2011)	
Branche de la méthode : C. Weiss (1970)	
La politisation de l'évaluation	Pour Weiss, que l'évaluateur le veuille ou non, l'évaluation est à la fois une activité de « recherche » et une activité politique : il importe de considérer l'influence des considérations politiques sur la conception et la mise en œuvre des programmes et de leur évaluation, les théories (à la fois importantes et éphémères) qui sous-tendent et soutiennent les programmes publics, l'utilisation différenciée des résultats évaluatif (améliorer le programme, produire des résultats compétitifs face aux intérêts politiques, etc.) et sur la manière dont l'évaluation pourrait contribuer à l'enrichissement des connaissances sur les interventions sociales.
Les méthodes pour évaluer l'efficacité des programmes [Amélioration de l'utilisation des méthodes de recherche propres aux sciences sociales, en évaluation]	Si Weiss reconnait le statut de *Gold Standard* du devis expérimental, elle précise que son utilisation n'est pas nécessairement appropriée en évaluation. De plus les mesures d'impact ne sont pas les plus pertinentes pour vérifier et renforcer l'efficacité des services et programmes sociaux. L'auteure suggère que : «*More resources should be allocated to evaluations that compare the effectiveness of variant conditions within programs (...) and begin to explain which elements and sub-elements are associated with more or less success. Such an approach produces data of interest across a wide range of programs and has high utility in pointing direction for further program development.* » (p.62-63). Elle recommande le recours à la fois aux méthodes qualitatives et aux méthodes quantitatives, et ce, en s'assurant de la qualité des procédures méthodologiques et la divulgation intégrale de leurs limites. La complexité du contexte évaluatif freine la possibilité pour l'évaluateur de garantir une parfaite objectivité, aussi le recours à de multiples méthodes de collectes de données est le compromis idéal pour générer des résultats à la fois rigoureux et utiles pour le programme.
Branche de l'utilisation	
D. Stufflebeam	**Le modèle CIPP (Context-Input-Process-Product) en tant qu'approche évaluative centrée sur le processus de prise de décision (1971) :** engage les parties prenantes en s'assurant que les questions d'évaluation répondent aux besoins de ces derniers, et que les informations évaluatives les assistent efficacement dans leurs prises de décision (Stufflebeam, 2013).
H. S. Preskill [Influencée par les travaux de Patton (1982) et Cousins (1998)]	**Apprentissage transformatif (1999) et *Capacity-Building Evaluation (2005)* :** processus où les individus, les équipes et les organisations identifient les informations dont ils ont besoin pour atteindre leurs objectifs. Afin de faciliter cet apprentissage, les évaluateurs devraient élargir les frontières entre eux et les parties prenantes, et leur permettre de saisir le processus d'obtention des résultats d'évaluation, et la manière dont ces derniers peuvent être impliqués (Christie et Alkin, 2013 ; Cousins, 2003, 1998 ; Preskill, 1994 ; Patton, 1982).
Quelques approches évaluatives contemporaines pragmatistes (Branche de l'utilisation)	

Évaluation axée sur l'utilisation (Patton, 2008, 2005, 1982)	- **Prémices :** l'évaluation devrait être jugée par son utilité et son utilisation réelle, et par conséquent, l'évaluateur doit promouvoir l'utilisation de l'évaluation en adaptant sa démarche aux facteurs situationnels et organisationnels spécifiques au contexte évaluatif. - **Objectif :** encourager l'utilisation de l'évaluation par les principaux utilisateurs attendus ou *parties prenantes clés* (mandataires, gestionnaires, etc.). - **Processus :** les parties prenantes clés sont impliquées, via un processus de consultation, dans la conception et le déroulement de l'évaluation afin de leur permettre de développer un sentiment d'adhésion, de reconnaitre l'utilité de l'évaluation et d'utiliser les résultats produits.
Évaluation de type *Empowerment* (Fetterman, 2005, 1994)	- **Prémices :** l'utilisation des concepts, des techniques et des résultats de l'évaluation peut favoriser l'amélioration et l'autodétermination. - **Objectif** : aider les individus à s'aider eux-mêmes et à améliorer leurs programmes, en recourant à une forme d'autoévaluation (et de réflexion). - **Processus :** l'évaluation est conduite par les participants du programme eux-mêmes (clients, consommateurs, membres du personnel, etc.) et un évaluateur externe joue le rôle de facilitateur.
Évaluation participative (Cousins et Whitmore, 1998)	- **Prémices :** les parties prenantes ont exprimé le désir de participer à l'amélioration de leur programme à travers son évaluation. - **Objectif :** inviter l'équipe de professionnels et les utilisateurs du programme à participer, avec l'équipe d'évaluation, à l'amélioration du programme afin de favoriser l'utilisation des résultats évaluatifs. - **Processus :** L'ensemble de la démarche évaluative (déterminer les questions et le type d'évaluation, le cadre méthodologique, la collecte et l'analyse des données, la communication des résultats) peut être partagé entre les évaluateurs et les professionnels (voir les clients) du programme.

Annexe 2. *Le processus de formulation de la définition canadienne de l'évaluation de programme*



Limites identifiées : (1) l'exclusion des écrits scientifiques francophones dans un contexte canadien bilingue (français-anglais); (2) les chercheurs postulent que l'ÉP est nécessairement défini dans les articles rapportant une évaluation. Or, si les auteurs publient dans une revue spécialisée en ÉP, ces derniers ne vont pas la définir, mais ils vont davantage décrire l'approche évaluative utilisée; et (3) le devis de recherche prévoyait une enquête auprès des utilisateurs ciblés de de l'évaluation, mais pour des raisons budgétaires ils ont dû annuler. Cependant, les parties prenantes font partie intégrante de la communauté évaluative et, selon nous, il est essentiel de s'assurer que les praticiens et les utilisateurs de l'évaluation partagent une définition, une vision commune, si on veut promouvoir l'utilité, l'utilisation et la crédibilité de l'évaluation de programme.

Annexe 3. *Les définitions actuelles de l'évaluation de programme*[27]

Auteur(s)	Définition
Patton (1997)	The systematic collection and analysis of information about program activities, characteristics, and outcomes to make judgements about the program, improve program effectiveness, and/or inform decisions about future programming. Program Evaluation: (1) involves the systematic collection and analysis of information; (2) focuses on a broad range of topics (accessibility, comprehensiveness, Integration, cost, efficiency, effectiveness); (3) is designed for a variety of uses (management, accountability, planning).
Fournier (2005)	Evaluation is an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs value, merit, worth, significance or quality of a program, product, person, or plan. Conclusions made in evaluations encompass both an empirical aspect and a normative aspect. It is the value feature that distinguishes evaluation from other types of inquiry such as basic science research, clinical epidemiology, investigative journalism, or public polling.
Yarbrough, Shulha, Hopson et Caruthers (2011)	The systematic investigation of the quality of programs, projects, subprograms, subprojects, and/or of their components or elements, together or singly for the purposes of decision making, judgements, conclusion, findings, new knowledge, organizational development, and capacity building in response to the needs of identified stakeholders leading to improvement and/or accountability in the users' program and systems ultimately contributing to organizational or social value.
Scriven (1991)	Evaluation refers to the process of determining merit, worth, or value of something, or the product of that process (…). The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of evaluands on these standards; some on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations.
Preskill et Torres (1999)	Evaluative inquiry is an ongoing process for investigating and understanding critical organizational issues. It is an approach to learning that is fully integrated with an organization's work practices, and as such, it engenders (a) organization members' interest and ability in exploring critical issues using evaluation logic, (b) organization members' improvement in evaluative processes, and (c) the personal and professional growth of individuals within the organization.
Rossi, Lipsey et Freeman (2019, 2004)	Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs. It draws on the techniques and concepts of social science disciplines and is intended to be useful for improving programs and informing social action ameliorating social problems.

[27] Les portions de phrase rédigées et soulignées en bleu correspondent aux éléments clés qu'on retrouve dans la définition « proxy » de l'évaluation de programme en contexte canadien de la SCÉ (2015).

Annexe 4. *La Taxonomie des compétences essentielles de l'évaluateur de programme (Stevahn, King, Ghere et Minnema, 2005) (Traduction libre)*

1.	**Pratique professionnelle**
1.1	Applique les normes professionnelles de l'évaluation
1.2	Agit de manière éthique et s'efforce de faire preuve d'intégrité et d'honnêteté dans la conduite des évaluations
1.3	Transmet des approches et des compétences évaluatives personnelles aux potentiels clients
1.4	Respecte les clients, les répondants, les participants du programme et les autres parties prenantes
1.5	Considère le bien-être général et public dans la pratique de l'évaluation
1.6	Contribue aux connaissances de base de l'évaluation
2.	**Enquête systématique (ou volet technique/opérationnel)**
2.1	Comprend/Maitrise les connaissances de base de l'évaluation (termes, concepts, méthodes, théories, hypothèses)
2.2	Connait les méthodes quantitatives
2.3	Connait les méthodes qualitatives
2.4	Connait les méthodes mixtes
2.5	Effectue des recensions des écrits
2.6	Spécifie la théorie du programme
2.7	Encadre les questions d'évaluation
2.8	Développe des conceptions d'évaluation
2.9	Identifie les sources de données
2.10	Collecte des données
2.11	Évalue la validité des données
2.12	Évalue la fiabilité des données
2.13	Analyse les données
2.14	Interprète les données
2.15	Émet des jugements
2.16	Élabore des recommandations
2.17	Justifie les décisions prises tout au long de l'évaluation
2.18	Rapporte les procédures d'évaluation et les résultats
2.19	Note les points forts et les limites de l'évaluation
2.20	Réalise des métaévaluations
3.	**Analyse situationnelle**
3.1	Décrit le programme
3.2	Détermine l'évaluabilité du programme
3.3	Identifie les intérêts des parties prenantes concernées
3.4	Répond aux besoins d'information des utilisateurs ciblés
3.5	Résout les conflits
3.6	Examine le contexte organisationnel de l'évaluation
3.7	Analyse les considérations politiques relatives à l'évaluation
3.8	S'occupe des questions liées à l'utilisation de l'évaluation
3.9	S'occupe des questions liées au changement organisationnel
3.10	Respecte le caractère unique du site d'évaluation et du client
3.11	Reste ouvert aux commentaires des autres
3.12	Modifie l'étude au besoin
4.	**Gestion de projet**
4.1	Répond aux appels de propositions
4.2	Négocie avec les clients avant le début de l'évaluation
4.3	Rédige des accords formels
4.4	Communique avec les clients tout au long de l'évaluation
4.5	Budgète l'évaluation
4.6	Justifie le coût compte tenu des besoins d'information
4.7	Identifie les ressources nécessaires pour l'évaluation (informations, expertise, main d'œuvre, instruments, etc.)
4.8	Utilise la technologie appropriée
4.9	Supervise les autres personnes impliquées dans la réalisation de l'évaluation
4.10	Forme les autres personnes impliquées dans la conduite de l'évaluation
4.11	Effectue l'évaluation de manière non perturbatrice
4.12	Présente le travail en temps opportun
5.	**Pratique réflexive**
5.1	Conscient de soi en tant qu'évaluateur (connaissances, aptitudes, attitudes)
5.3	Réfléchit à sa pratique personnelle d'évaluation (compétences et domaines de croissance)
5.4	Poursuit son perfectionnement professionnel en évaluation
5.5	Poursuit son développement professionnel dans des domaines de contenus pertinents
5.6	Établit des relations professionnelles pour améliorer l'évaluation pratique
6.	**Compétences interpersonnelles**
6.1	Mobilise des compétences de communication écrite
6.2	Mobilise des compétences de communication verbale/d'écoute
6.3	Mobilise des compétences de négociation
6.4	Mobilise des compétences en résolution de conflits
6.5	Facilite des interactions interpersonnelles constructives (travail d'équipe, animation de groupe, traitement)
6.6	Démontre une compétence interculturelle

Annexe 5. *Certificats d'approbation éthique*

Numéro de dossier: 04-15-12 **Date (mm/jj/aaaa):** 06/05/2015

Université d'Ottawa University of Ottawa
Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

Certificat d'approbation éthique

CÉR Sciences sociales et humanités

Chercheur principal / Superviseur / Co-chercheur(s) / Étudiant(s)

Prénom	Nom de famille	Affiliation	Rôle
Eric	Dionne	Éducation / Education	Superviseur
Maud	Mediell	Éducation / Education	Étudiante-chercheuse

Numéro du dossier: 04-15-12

Type du projet: Thèse de doctorat

Titre: Comparaison du processus décisionnel des évaluateurs novices et experts

Date d'approbation (mm/jj/aaaa)	Date d'expiration (mm/jj/aaaa)	Approbation
06/05/2015	06/04/2016	Ia

(Ia: Approbation complète, Ib: Autorisation préliminaire de libération de fonds de recherche)

Conditions Spéciales / Commentaires:
Le présent certificat est valide pour le recrutement des participants et la collecte des données à l'Université d'Ottawa. Il sera mis à jour lorsque vous soumettrez les certificats d'approbation éthique de l'Université du Québec à Montréal et de l'École Nationale d'Administration publique.

550, rue Cumberland 550 Cumberland Street
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada
(613) 562-5387 • Téléc./Fax (613) 562-5338
http://www.recherche.uottawa.ca/deontologie/ http://www.research.uottawa.ca/ethics/index.html

Numéro de dossier: 04-15-12 **Date (mm/jj/aaaa):** 03/17/2017

Université d'Ottawa University of Ottawa
Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

Certificat d'approbation déontologique

CÉR Sciences sociales et humanités

Chercheur principal / Superviseur / Co-chercheur(s) / Étudiant(s)

Prénom	Nom de famille	Affiliation	Rôle
Eric	Dionne	Éducation / Education	Superviseur
Maud	Mediell	Éducation / Education	Étudiant-chercheur

Numéro du dossier: 04-15-12

Type du projet: Thèse de doctorat

Titre: Comparison du processus décisionnel des évaluateurs novices et experts

Date de renouvellement (mm/jj/aaaa)	Date d'expiration (mm/jj/aaaa)	Approbation
06/05/2016	06/04/2017	Renouvellement

Conditions Spéciales / Commentaires:
N/A

550, rue Cumberland, pièce 154 550 Cumberland Street, room 154
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada
(613) 562-5387 • Téléc./Fax (613) 562-5338
www.recherche.uottawa.ca/deontologie/ www.research.uottawa.ca/ethics/

Numéro de dossier: 04-15-12 **Date (mm/jj/aaaa):** 06/14/2017

Université d'Ottawa University of Ottawa
Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

Certificat d'approbation déontologique

CÉR Sciences sociales et humanités

Chercheur principal / Superviseur / Co-chercheur(s) / Étudiant(s)

Prénom	Nom de famille	Affiliation	Rôle
Eric	Dionne	Éducation / Éducation	Superviseur
Maud	Mediell	Éducation / Éducation	Étudiant-chercheur

Numéro du dossier: 04-15-12

Type du projet: Thèse de doctorat

Titre: Comparison du processus décisionnel des évaluateurs novices et experts

Date de renouvellement (mm/jj/aaaa)	Date d'expiration (mm/jj/aaaa)	Approbation
06/05/2017	06/04/2018	Renouvellement

Conditions Spéciales / Commentaires:
N/A

550, rue Cumberland, pièce 154 550 Cumberland Street, room 154
Ottawa (Ontario) K1N 6N5 Canada Ottawa, Ontario K1N 6N5 Canada
(613) 562-5387 • Téléc./Fax (613) 562-5338
www.recherche.uottawa.ca/deontologie/ www.research.uottawa.ca/ethics/

UQÀM
Université du Québec à Montréal

CIEREH
Comité institutionnel d'éthique de la recherche avec des êtres humains

No du certificat:: A-UofO-04-15-12

CERTIFICAT D'ÉTHIQUE

Le *Comité institutionnel d'éthique de la recherche avec des êtres humains* de l'UQAM, a examiné le protocole de recherche suivant et jugé conforme aux pratiques habituelles et répond aux normes établies par le Cadre normatif pour l'éthique de la recherche avec des êtres humains de l'UQAM (juin 2012).

Protocole de recherche

Chercheur(e) principal(e) : Maud Medeill
Programme d'études : Doctorat en éducation
Unité de rattachement : Faculté d'éducation, Université d'Ottawa
Directeur de recherche : Éric Dionne, Faculté d'éducation, Université d'Ottawa
Titre du protocole de recherche : *Comparaison du processus décisionnel des évaluateurs novices et experts*
Organisme de financement (le cas échéant): s/o

Modalités d'application

Le présent certificat est valide pour le projet tel que soumis au CIEREH. Les modifications importantes pouvant être apportées au protocole de recherche en cours de réalisation doivent être communiquées au comité[1].

Tout évènement ou renseignement pouvant affecter l'intégrité ou l'éthicité de la recherche doit être communiqué au comité.

Toute suspension ou cessation du protocole (temporaire ou définitive) doit être communiquée au comité dans les meilleurs délais.

Le présent certificat d'éthique est valide jusqu'au **6 juillet 2016**. Selon les normes de l'Université en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique. Le rapport d'avancement de projet (renouvellement annuel ou fin de projet) est requis pour le **6 juin 2016**. Vous recevrez automatiquement un premier courriel de rappel trois mois avant la date d'échéance du certificat.

[redacted]

Maria Nengeh Mensah, Ph.D.
Professeure
Présidente

6 juillet 2015
Date d'émission initiale du certificat

[1] Modifications apportées aux objectifs du projet et à ses étapes de réalisation, au choix des groupes de participants et à la façon de les recruter et aux formulaires de consentement. Les modifications incluent les risques de préjudices non-prévus pour les participants, les précautions mises en place pour les minimiser, les changements au niveau de la protection accordée aux participants en termes d'anonymat et de confidentialité ainsi que les changements au niveau de l'équipe (ajout ou retrait de membres).

Annexe 6. *Matériel de recrutement des participants*

6.1 Lettre de présentation

Présentation de la chercheuse principale

Maud Mediell, doctorante, Faculté d'éducation
Courriel :
Sous la direction du professeur Eric Dionne, Faculté d'éducation
Courriel :

Titre du projet

Comparaison du processus décisionnel des évaluateurs novices et experts

But de l'étude

Le but de l'étude est de mieux documenter comment les évaluateurs de programme, novices et experts, appréhendent le processus d'évaluation. Cette recherche doctorale vise à étudier les pratiques des évaluateurs et, plus particulièrement, les choix méthodologiques qu'ils mettent en œuvre. Elle vise également à rendre compte de l'influence des contraintes inhérentes au terrain sur lesdits choix. Enfin, la finalité de cette étude est de pouvoir offrir aux praticiens de l'évaluation, des propositions méthodologiques afin de répondre de façon efficace à la réalité de terrain, aux contraintes et de favoriser ainsi la réalisation d'évaluations de programme efficientes et de qualité.

Pour le recrutement des novices – débutants avancés *(« juniors »)* :

- Détenir un diplôme en évaluation de programme ou avoir suivis une formation interne en évaluation de programme.
- Avoir entre une et trois années d'expérience (ou moins) en évaluation de programme.
- Parler français et/ou anglais

Pour le recrutement des experts *(« seniors associés »)* :

- Détenir un ou plusieurs diplômes en évaluation de programme, en recherche en sciences sociales, etc.
- Avoir au moins dix ans d'expérience en évaluation de programme.
- Démontrer un certain degré de contribution au domaine ***(optionnel)***
- Parler français et/ou anglais

Bienfaits de la participation à étude :

Participer à cette recherche présente plusieurs bénéfices :

- échanger avec une chercheuse-évaluatrice de programme, qui s'intéresse aux problèmes que nous, praticiens de l'évaluation, rencontrons sur le terrain.
- réaliser une métaévaluation d'une évaluation de programme que les participants ont réalisé, soit une pratique réflexive essentielle lorsque l'on est évaluateur de programme
- participer à l'enrichissement des connaissances et contribuer au développement des solutions afin de réduire l'écart entre la théorie et la pratique de terrain. Ce qui, par ailleurs, entrainera de meilleures évaluations, de meilleurs praticiens ainsi qu'une meilleure image de notre discipline.

La participation à cette recherche implique de :

1. Fournir un ou plusieurs rapports d'évaluation récents ainsi que les offres de services correspondant aux demandes faites par les clients, afin que la chercheuse réalise une métaévaluation de ce dernier.
2. Réaliser, au cours d'un premier entretien semi-dirigé (enregistré), la lecture d'un article rapportant une évaluation de programme, et en discuter avec la chercheuse. Le participant bénéficiera d'un temps de préparation nécessaire (la veille de l'entrevue) afin d'être en mesure de réaliser l'exercice à l'orale dans le cadre d'une entrevue semi-dirigée. Cette activité devrait durée 90 minutes.
3. Participer à un entretien semi-dirigé (enregistré), de type descriptif, dans le cadre duquel le participant reviendra avec la chercheuse sur les éléments clés du rapport d'évaluation qui a été fourni. L'objectif de cet entretien est de revenir sur le déroulement de l'évaluation de programme. L'entretien durera entre 45 et 60 minutes.

6.2 Formulaire de consentement

Faculté d'éducation
Faculty of Education
613-562-5804
1-800-860-8577
613-562-5963
education.uOttawa.ca
145 Jean-Jacques Lussier
Ottawa ON K1N 6N5
Canada

FORMULAIRE DE CONSENTEMENT

Titre du projet :

Comparaison du processus décisionnel des évaluateurs novices et experts

Chercheuse principale :
Maud Mediell, doctorante, Faculté d'éducation
Courriel : [redacted]
Sous la direction du professeur Eric Dionne, Faculté d'éducation
Pavillon Lamoureux, local 408
145 rue Jean-Jacques Lussier
K1N6N5, ON, Ottawa, Canada
[redacted]
Courriel : [redacted]

Invitation à participer:

Je suis invité(e) à participer à la recherche doctorale nommée ci-haut et qui est menée par Maud Mediell.

But de l'étude:

Le but de l'étude est de mieux documenter comment les évaluateurs de programme, novices et experts, appréhendent le processus d'évaluation. Cette recherche doctorale vise à étudier les pratiques des évaluateurs et, plus particulièrement, les choix méthodologiques qu'ils mettent en œuvre. Elle vise également à rendre compte de l'influence des parties prenantes sur lesdits choix. Enfin, la finalité de cette étude est de pouvoir offrir aux praticiens de l'évaluation, des propositions méthodologiques afin de répondre de façon efficace à la réalité de terrain, aux contraintes et de favoriser ainsi la réalisation d'évaluations de programme efficientes et de qualité.

Participation:

Ma participation à cette recherche se fera en trois temps :

1. Fournir un ou plusieurs rapports d'évaluation récents ainsi que les offres de services correspondant aux demandes faites par les clients, afin que la chercheuse réalise une méta-évaluation de ce dernier.
2. Réaliser, au cours d'un premier entretien semi-dirigé (enregistré), je vais lire un article rapportant une évaluation de programme, et en discuter avec la chercheuse. Je bénéficierais d'un temps de préparation nécessaire afin d'être en mesure de réaliser l'exercice à l'orale dans le cadre d'une entrevue semi-dirigée. Cette activité devrait durée entre 90 et 120 minutes.

3. Participer à un entretien semi-dirigé (enregistré), de type descriptif, dans le cadre duquel nous reviendrons avec la chercheuse sur les éléments clés du rapport d'évaluation que j'ai produit. L'objectif de cet entretien est de revenir sur le déroulement de l'évaluation de programme que j'ai réalisé. L'entretien durera entre 45 et 60 minutes.

Risque:

La chercheuse a précisé qu'il n'y a pas de risques prévisibles.

Bienfaits:

Ma participation à cette recherche me donne l'occasion de discuter, d'échanger avec une chercheuse, elle-même évaluatrice de programme, qui s'intéresse aux problèmes que nous, praticiens de l'évaluation, rencontrons sur le terrain. De plus, cela m'offre une chance de réaliser une méta-analyse de mon propre travail, de réfléchir sur la manière dont j'ai réalisé l'évaluation de programme faisant l'objet du rapport d'évaluation que j'ai fourni à la chercheuse. Enfin, je participe à l'enrichissement des connaissances et je contribue au développement des solutions que la chercheuse a pour objectifs de proposer afin de réduire l'écart entre la théorie et la pratique de terrain. Ainsi, cela entrainera de meilleures évaluations, de meilleurs praticiens ainsi qu'une meilleure image de notre discipline.

Confidentialité et anonymat:

J'ai l'assurance du chercheur que l'information que je partagerai avec elle/lui restera strictement confidentielle. Je m'attends à ce que le contenu ne soit utilisé que pour la rédaction de sa thèse et des publications futures.

L'anonymat et la confidentialité sont garantis de la façon suivante : une fois le recrutement réalisé, la chercheuse me demandera de m'auto-attribuer un code qui sera composé de trois lettres en s'assurant que cela ne donne pas d'indice sur leurs initiales, et un chiffre soit, par exemple, FAD7. Ainsi, elle sera la seule détentrice de cette information et personne ne sera en mesure de m'identifier. Elle anonymisera également le rapport d'évaluation. Aussi fera-t-elle référence au projet d'évaluation en faisant référence à son devis, par exemple : l'évaluation de l'implantation d'un programme communautaire de type *children after-school.* Enfin, elle ne fera jamais référence à mon identité, que soit ce soit dans sa thèse ou tout autre type de publications ou communications orales.

Conservation des données:

En ce qui concerne la protection physique des données, celles-ci seront conservées de manière sécuritaire, "sous clés", dans un casier dans le bureau du directeur de thèse de la chercheuse principale qui ferme également à clé,

et ce, pendant toute la durée de l'étude. Quant à la protection technique des données, celles-ci seront anonymisées/codées dans tous les documents relatifs à l'étude (enregistrements audio, fichiers de transcriptions, rapports d'évaluation, etc.), chaque participant devenant un "code" sur tous les documents qui sont relatifs à leur participation. Enfin, les données se trouveront sur le disque dur d'un seul ordinateur dans le bureau du directeur de thèse qui ferme à clé et elles seront conservées pendant 10 ans à partir de la fin du doctorat de la chercheuse principale à des fins de publications, de communications scientifiques, mais aussi pour une potentielle nouvelle recherche inscrite dans la lignée des résultats de la thèse de la doctorante (chercheuse principale). Elles seront détruites à la fin de cette période.

***Participation volontaire**:*

Ma participation à la recherche est volontaire et je suis libre de me retirer en tout temps, et/ou refuser de répondre à certaines questions, sans subir de conséquences négatives. Si je choisis de me retirer de l'étude, la chercheuse me demandera mon consentement afin d'utiliser mes données.

Acceptation:

Je, _____, accepte de participer à cette recherche menée par Maud Mediell de la Faculté d'éducation de l'Université d'Ottawa, laquelle est supervisée par le directeur de la chercheuse, le professeur Eric Dionne.

Pour tout renseignement additionnel concernant cette étude, je peux communiquer avec la chercheuse ou son directeur.

Pour tout renseignement sur les aspects éthiques de cette recherche, je peux m'adresser au Responsable de l'éthique en recherche, Université d'Ottawa, Pavillon Tabaret, 550, rue Cumberland, pièce 154, (613) 562-5387 ou ethics@uottawa.ca.

Il y a deux copies du formulaire de consentement, dont une copie que je peux garder.

Signature du participant :
Date :

Signature de la chercheuse :

Date:

Annexe 7. *Instrumentation de la collecte de données numéro 1 – Référentiel d'analyse de l'activité et Guide d'entretien métaévaluatif*

Référentiel d'analyse de l'activité

Afin de s'assurer que nous aborderons tous les éléments nécessaires afin de répondre aux objectifs de la première collecte de données, nous avons préparé un référentiel d'analyse de l'activité (tableau ci-dessous), à partir duquel nous avons conçu notre guide d'entretien de retour sur l'activité. Ce référentiel nous servira également pour la prise de notre durant l'entrevue.

L'entretien de retour sur l'activité a 3 objectifs sur lesquelles seront basés les axes de relances dans le cadre de la méta-analyse réalisée oralement par les participants :

- Comprendre ce qui constitue, pour les participants, une bonne pratique évaluative.
- Explorer avec eux le processus en amont d'une prise de décision en ce qui a trait à la méthodologie élaborée (idéale) et la méthodologie qui a été mise en place.
- Comprendre le rôle que peuvent jouer les contraintes inhérentes au terrain (telles que l'implication des parties prenantes ainsi que la manière dont ils appréhendent leur relation avec ces dernières) et l'influence que cela peut avoir sur le processus évaluatif.

Section du rapport d'EP discutée	Analyse critique (points forts, points faibles, remarques, etc.)	Suggestions pour amélioration (ce que l'interviewé aurait fait)	Processus décisionnel (suggestions quant aux raisons justifiant les choix effectués durant l'EP)	Autre(s)
Etc.				

Guide d'entrevue de retour sur l'activité ou *d'entretien métaévaluatif*

Ce guide d'entretien de retour sur l'activité est basé sur le référentiel d'analyse d'activité. L'utilisation du guide, mais aussi du référentiel, demeurera souple puisque nous privilégierons la parole du participant et nous relancerons uniquement lorsque c'est nécessaire (afin d'être en mesure de répondre à nos objectifs).

Je vous ai sollicité aujourd'hui pour un premier entretien dans le cadre de votre participation à la recherche doctorale que je mène sur le processus décisionnel des évaluateurs de programme.

Rappel : le but de l'étude est de mieux comprendre comment les évaluateurs de programme appréhendent le processus d'évaluation. Cette recherche doctorale vise à étudier les pratiques des évaluateurs et, plus particulièrement, les choix méthodologiques qu'ils mettent en œuvre. La finalité de cette étude est de pouvoir offrir aux praticiens de l'évaluation, des propositions méthodologiques afin de répondre de façon efficace à la réalité de terrain, aux contraintes et de favoriser ainsi la réalisation d'évaluations de programme de qualité.

Déroulement de l'entretien : aujourd'hui vous allez d'abord me faire part de l'analyse critique que vous avez préparée concernant le contenu du devis de l'évaluation qui fait l'objet du rapport que je vous aie envoyé hier ; puis nous reviendrons ensemble sur les différents éléments clés de votre analyse, vos commentaires, vos suggestions pour améliorer la méthodologie évaluative, etc. Cet entretien est enregistré, et il devrait durer entre 60 et 80 minutes. Avez-vous des questions ou des préoccupations dont vous aimeriez me faire part avant de commencer ?

1ère partie de l'entretien : le participant est invité à réaliser oralement son analyse critique du plan d'évaluation (incluant la méthodologie) présenté dans le rapport d'évaluation.

I. Pouvez-vous me faire part de votre analyse critique du plan d'évaluation (incluant la méthodologie) [ou devis d'évaluation] présenté dans le rapport d'évaluation ?

À ce niveau-ci de l'entrevue, les relances ne seront que des demandes de clarification. Nous demeurerons concentrés sur la demande de détails.

Cette section concerne particulièrement l'objectif numéro 1 de l'entrevue de retour sur activité : comprendre ce qui constitue, pour les participants, une bonne pratique évaluative. Si l'on se réfère au référentiel d'analyse de l'activité, nous nous situons au niveau de l'axe 1 portant sur l'analyse critique (points forts, points faibles, suggestions pour amélioration, etc.).

2e partie de l'entretien : le participant est invité à répondre à différentes questions se rapportant aux axes du référentiel d'analyse d'activité qui n'ont pas encore été abordés.

II. Merci pour cette analyse. J'aimerais revenir à présent sur certains éléments avec vous.

À ce niveau-ci de l'entrevue, nous reviendrons sur chacune des sections du rapport mises en exergue par l'interviewé et nous poserons les questions suivantes (rappel – cette entrevue est semi-dirigée, aussi en fonction de la discussion, les questions ne seront pas nécessairement posées dans cet ordre ou formulées exactement de cette manière) :

i. Au regard de vos différents commentaires, qu'auriez-vous fait à la place des évaluateurs du programme dont vous venez de critiquer le rapport ?
ii. Selon vous, quelles sont les raisons qui ont pu pousser les évaluateurs du programme à prendre ces décisions ?
iii. Au regard des différents éléments que nous avons abordés aujourd'hui, avez-vous d'autres commentaires dont vous aimeriez me faire part ?

Cette section concerne particulièrement les objectifs numéro 2 et 3 de l'entrevue de retour sur l'activité : (2) explorer avec eux le processus en amont d'une prise de décision en ce qui a trait à la méthodologie élaborée (idéale) et la méthodologie qui a été mise en place ; et (3) comprendre le rôle que peuvent jouer les contraintes inhérentes au terrain (telles que l'implication des parties prenantes ainsi que la manière dont ils appréhendent leur relation avec ces dernières) et l'influence que cela peut avoir sur le processus évaluatif. Si l'on se réfère au référentiel d'analyse de l'activité, nous nous situons au niveau de l'axe 2 portant sur les suggestions pour amélioration (ce que l'interviewé aurait fait), et de l'axe 3 s'attardant sur le processus décisionnel (suggestions quant aux raisons justifiant les choix effectués durant l'EP faisant l'objet de l'analyse critique).

Annexe 8. *Procédure de validation de contenu de la grille d'analyse de rapports d'évaluation centrée sur le devis d'évaluation*

Sélection et profils des experts

Afin de s'assurer de la validité du contenu de la grille d'analyse de rapports d'évaluation, nous avons recruté des experts en mesure et évaluation, en évaluation de programme au fédéral et en évaluation de programme en milieu éducationnel et communautaire ; soit trois domaines d'expertise en évaluation de programme. Nous avons envoyé un courriel de recrutement (section ci-dessous) à six experts (au total) et trois ont accepté de réaliser la validation du contenu de la grille d'analyse de rapports d'évaluation. Le premier expert est professeur agrégé spécialisé en mesure et évaluation, et le second expert travaille dans un département d'évaluation de programme au gouvernement fédéral et participe à la formation des nouveaux employés en partenariat avec la Société canadienne d'évaluation. Enfin, le troisième expert est professeur émérite spécialisé en recherche et en évaluation de programmes éducationnels et communautaires. Au regard de l'analyse de leurs commentaires concernant les items de la grille et de la complémentarité de ces derniers, nous n'avons pas invité davantage d'experts à participer à la validation du contenu de la grille.

Courriel de recrutement envoyé aux experts

Par le présent courriel, je sollicite votre expertise afin de réaliser la validation de contenu d'une grille d'évaluation de rapports d'évaluation de programme élaborée dans le cadre de mon projet de thèse doctorale, sous la direction du professeur Eric Dionne.

S'intitulant *Comparaison du processus décisionnel des évaluateurs novices et experts,* cette recherche doctorale vise à étudier les pratiques des évaluateurs et, plus particulièrement, les choix méthodologiques qu'ils mettent en œuvre. Elle vise également à rendre compte de l'influence des parties prenantes sur lesdits choix. Enfin, la finalité de cette étude est d'offrir aux praticiens de l'évaluation des propositions méthodologiques afin de répondre de façon efficace à la réalité de terrain, aux contraintes et de favoriser ainsi la réalisation d'évaluations de programme efficientes et de qualité.

Si vous acceptez d'accorder du temps, de 45 minutes à 1h, pour valider le contenu des items de cet instrument de collecte de données, voici quelques consignes :

1. Je vous invite à consulter le dossier électronique joint à ce courriel – vous trouverez un document décrivant la méthodologie d'élaboration de la grille ainsi que la grille elle-même et deux rapports d'évaluation que vous pouvez utiliser, si vous le désirez, dans le cadre du processus de validation du contenu de la grille (optionnel).
2. Je vous invite, dans le cadre de la validation du contenu, à remplir le formulaire électronique inclus dans le dossier. Je vous demande d'utiliser la case commentaire à la suite de chacun des items, pour me faire part de vos critiques et suggestions pour amélioration des items que vous jugerez nécessaire de modifier.

Merci pour toute l'aide que vous pourrez m'apporter.

Maud Mediell

Validation du contenu de la grille

Comme nous l'avons mentionné dans le courriel de recrutement (ci-dessus), nous avons converti la grille d'analyse de rapports d'évaluation en formulaire électronique afin de faciliter le

processus de validation de contenu ainsi que l'analyse des évaluations des experts. En guise d'introduction du formulaire nous avons décrit l'instrument et les consignes relatives au processus de validation de la grille. Nous avons demandé aux experts d'utiliser la case commentaire à la suite de chacun des items, pour nous faire part de leurs critiques et suggestions pour l'amélioration des items qu'ils jugeaient nécessaire de modifier.

Une fois le processus de validation du contenu de la grille terminé, nous avons analysé les évaluations (les critiques et les suggestions) du contenu des items de la grille faites par les trois experts. Aussi, vous trouverez à la page suivante un tableau résumant le processus décisionnel concernant chaque item de la grille au regard des commentaires des experts (Tableau du processus de modification à la page suivante), suivi de la nouvelle version de l'instrument (Annexe 9).

Tableau de modification des items de la grille de métaévaluation du devis d'évaluation « post rétroaction » des experts

Thème 1 : précision concernant la description de la démarche d'évaluation		
1.1 Description des objectifs de l'évaluation	**Commentaires concernant l'item 1.1.1 Le mandat d'évaluation est clairement décrit**	**Décision concernant l'item**
	É.1 : je ne suis pas certaine que le terme mandat soit utilisé, en général on parle de contexte ou de problématique. Il peut y avoir confusion entre mandat et objectifs.	Le contexte fait partie du mandat d'évaluation. Aussi, une évaluation est mandatée par des parties prenantes et en fonction du contexte, des objectifs de l'évaluation (des raisons d'être de l'évaluation) et du cahier des charges, l'évaluateur définit le plan d'évaluation. Enfin, nous sommes consciente que selon les circonstances, les objectifs de l'évaluation sont difficiles à définir, cependant cela n'empêche pas de formuler des objectifs plus larges, mais clairement décrits. **Nous conservons l'item tel quel** : **1.1.1 Le mandat d'évaluation est clairement décrit**
	É.2 : Critère sensé et forcément inclus dans nos rapports d'évaluation; il sert à légitimer la démarche et le pourquoi de l'évaluation produite; au gouvernement fédéral, le mandat tire son origine du Conseil du Trésor par le biais de la politique d'évaluation qui sera bientôt revue et remplacée par la politique sur les résultats.	
	É.3 : Étant donné que le mandat d'une évaluation formative pourrait être assez vague au début du processus d'évaluation et même au moment de rédiger le rapport, peut-être que vous pourriez formuler cet item (et peut-être d'autres) de la façon suivante: "Le mandat d'évaluation a été décrit dans le rapport aussi clairement que possible dans les circonstances, étant donné les objectifs quelque peu vagues de l'évaluation".	
	Commentaires concernant l'item 1.1.2 Les objectifs d'évaluation répondent adéquatement au mandat d'évaluation	**Décision concernant l'item**
	É.1 : voir mon commentaire précédent. Il faudrait plutôt regarder la clarté et la précision des objectifs, afin de pouvoir vérifier ensuite l'adéquation entre le devis et les objectifs.	Nous sommes d'accord avec les évaluateurs É.1 et É.2, aussi au regard de leurs commentaires, voici les **nouveaux items que nous proposons**: **1.1.2 Les objectifs de l'évaluation sont clairs** **1.1.3 Les objectifs d'évaluation sont précis**
	É.2 : Les objectifs, focus et scope d'une évaluation donnée sont forcément inclus et inter-reliés avec les programmes d'architecture des activités; au fédéral, ce critère sera automatiquement présent ou l'évaluation sera rejetée aussitôt qu'à la phase du devis d'évaluation.	

	É.3 : La précision et clarté des objectifs risquent d'influencer la précision du mandat. Donc, le sens exact du mot '"adéquatement" pourrait varier avec les circonstances de l'évaluation.	
1.2 Justification du choix du devis d'évaluation	**Commentaires concernant l'item 1.2.1 Le type d'évaluation choisi est adéquat pour répondre aux objectifs d'évaluation**	**Décision concernant l'item**
	É.1 : N/A	L'évaluateur É.3 soulève un bon point, le terme adéquat peut porter à confusion (et ce commentaire a été mentionné pour les deux items précédents), aussi voici **notre nouvelle proposition concernant cet item**: **1.2.1 le type d'évaluation choisi est pertinent pour répondre aux objectifs d'évaluation**
	É.2 : Au fédéral, le type d'évaluation choisi est un peu forcé; l'évaluateur doit s'assurer qu'une l'évaluation de la pertinence et de la performance des programmes est faite selon des critères bien spécifiques. La réponse à ce critère serait probablement toujours positive en ce qui concerne les évaluations fédérales.	
	É.3 : C'est bien -- mais encore le sens exact du mot '"adéquat" pourrait varier avec les circonstances de l'évaluation (e.g., avec la précision/clarté des objectifs et du mandat).	
	Commentaires concernant l'item 1.2.2 Le choix du type d'évaluation est justifié par l'auteur	**Décision concernant l'item**
	É.1 : N/A	Nous voulons que les praticiens soient en mesure d'utiliser cette grille de méta-évaluation quel que soit leur domaine de pratique, aussi nous sommes consciente que certains items ne seront pas nécessairement pertinents pour tous les types d'évaluation (ou n'auront pas nécessairement le même poids). De surcroit, si le Secrétariat du Trésor prodigue des directives en matière du choix du type d'évaluation, ce dernier souligne l'importance de clairement décrire le choix du type d'évaluation (dans son guide pour rédiger un rapport d'évaluation de qualité). Aussi, **nous conservons l'item tel quel**: **1.2.2 Le choix du type d'évaluation est justifié par l'auteur**
	É.2 : Comme indiqué précédemment, le choix du type d'évaluation est un peu forcé et demande très peu de justifications de la part de l'évaluateur. Nos évaluations sont encadrées et bien règlementées. Ce critère risqué d'être non applicable pour l'ensemble des évaluations fédérales.	
	É.3 : Bien.	
	Commentaires concernant l'item 1.2.3 Le devis d'évaluation est fidèle au type d'évaluation choisi	**Décision concernant l'item**
	É.1 : le terme devis porte à confusion ici. Parle-t-on de la méthodologie?	Le terme devis d'évaluation fait référence au terme anglophone « *evaluation design* » qu'on définit comme l'empreinte

	É.2 : Cela dépend de ce que l'on entend par le terme "devis d'évaluation". Chez nous, le devis et un document appelé "workplan" élaboré spécifiquement pour l'évaluation d'un programme; chaque évaluation possède un workplan qui est le premier document à produire dans la conduite d'un processus d'évaluation. Comme nous avons très peu de flexibilité pour choisir le type d'évaluation, le devis devra nécessairement refléter cette réalité. Il serait dangereux de considérer la portion du rapport portant sur la méthodologie comme étant le devis d'évaluation, si cette hypothèse représente une option.	méthodologique de l'évaluation. Aussi, au regard des commentaires il semble plus approprié de parler de plan d'évaluation celui-ci faisant explicitement référence à la méthodologie évaluative. Aussi, voici **notre nouvelle proposition concernant cet item**: **1.2.3 Le plan d'évaluation est fidèle au type d'évaluation choisi**
	É.3 : Bien.	
1.3 Présentation des questions d'évaluation	**Commentaires concernant l'item 1.3.1 Les questions d'évaluation correspondent adéquatement aux objectifs d'évaluation**	**Décision concernant l'item**
	É.1 : N/A	Dans la continuité des changements apportés à l'item 1.2.1, voici **notre nouvelle proposition concernant cet item**: **1.3.1 Les questions d'évaluation sont pertinentes au regard des objectifs d'évaluation**
	É.2 : Il s'agit d'un critère qui fait appel au jugement de ceux qui utiliseront la grille, donc qui risque de varier pour plusieurs individus évaluant un même rapport.	
	É.3 : La précision et clarté des objectifs et du mandat risquent d'influencer la précision des questions d'évaluation. Donc, le sens exact du mot '"adéquatement" pourrait varier avec les circonstances de l'évaluation.	
	Commentaires concernant l'item 1.3.2 Les questions d'évaluation sont claires	**Décision concernant l'item**
	É.1 : N/A	Afin d'assurer la rigueur scientifique d'une évaluation, de la définition de ses objectifs à la rédaction de ses résultats, les questions d'évaluation doivent être claires, et ce, quelles que soient les circonstances. Aussi, **nous conservons l'item tel quel**: **1.3.2 Les questions d'évaluation sont claires**
	É.2 : Encore une fois ici, ce qui peut sembler "claire" pour un évaluateur peut sembler très flou pour un autre évaluateur et ce pour l'évaluation d'un même rapport; c'est une question de jugement qui met en cause le facteur humain dont on ne contrôle pas tous les facteurs influençant la prise de décision. De plus, ce critère fait fi du facteur "contexte" qui peut souvent ne pas être toujours bien défini.	

	É.3 : La clarté des objectifs et du mandat risquent d'influencer la clarté des questions d'évaluation. Donc, le sens exact du mot '"clair" pourrait varier avec les circonstances de l'évaluation.	
	<u>Commentaires concernant l'item</u> 1.3.3 Les questions d'évaluation sont précises	**Décision concernant l'item**
	É.1 : N/A	Afin d'assurer la rigueur scientifique d'une évaluation, de la définition de ses objectifs à la rédaction de ses résultats, les questions d'évaluation doivent être précises, et ce, quelles que soient les circonstances. Aussi, **<u>nous conservons l'item tel quel</u>**: **1.3.3 Les questions d'évaluation sont précises**
	É.2 : Même commentaires que pour le point 1.3.2. Le terme "précis" peut évoquer des choses différentes pour différentes personnes.	
	É.3 : La précision des objectifs et du mandat risquent d'influencer la précision des questions d'évaluation. Donc, le sens exact du mot '"clairs" pourrait varier avec les circonstances de l'évaluation.	
1.4 Description de la méthodologie	**<u>Commentaires concernant l'item</u> 1.4.1 Les sources d'information et les procédures d'échantillonnage sont décrites**	**Décision concernant l'item**
	É.1 : s'il y a plusieurs questions, ce serait peut-être utile de connaitre cette réponse question par question plutôt que globalement (par exemple une réponse partiellement sans commentaire ne sera pas facile à analyser).	Au regard du commentaire de É.3 voici **<u>notre nouvelle proposition concernant cet item</u>**: **1.4.1 Les sources d'information et les procédures d'échantillonnage sont décrites et justifiée** Au regard du commentaire de É.1 **<u>nous avons inséré la remarque suivante dans la case commentaire</u> :** « Si le plan d'évaluation fait état de plusieurs sources d'information et procédures d'échantillonnage, veuillez utiliser cette case (commentaire) pour évaluer la description (et la justification) de chaque source d'information et la procédure d'échantillonnage de cette dernière ».
	É.2 : Être mis au courant des sources et des procédures d'échantillonnage des données peur s'avérer être un critère sensé qui peut améliorer la crédibilité à donner à un rapport. Toutefois, plusieurs s'abstiennent de trop développer ces éléments par souci de concision et parfois de cohérence; on peut ainsi retrouver sommairement ces informations dans la section méthodologie d'un rapport donné.	
	É.3 : Bien	
	<u>Commentaires concernant l'item</u> 1.4.2 Les procédures de sélection et de développement de l'instrumentation sont décrites	**Décision concernant l'item**

	É.1 : pourquoi avoir regroupé le choix des instruments et leur développement dans la même question?	Au regard des commentaires de É.1 et É.3 **<u>nous proposons les nouveaux items suivants</u> :** **1.4.2 Les procédures de sélection de l'instrumentation sont décrites et justifiées** Au regard du commentaire de É.1 **<u>nous avons inséré la remarque suivante dans la case commentaire</u>** : « Si le plan d'évaluation fait état de la sélection de plusieurs instruments, veuillez utiliser cette case (commentaire) pour évaluer la description (et la justification) de la procédure de sélection de chaque instrument de collecte de données ». **1.4.3 Les procédures de développement de l'instrumentation sont décrites et justifiées** Au regard du commentaire de É.1 **<u>nous avons inséré la remarque suivante dans la case commentaire</u>** : « Si le plan d'évaluation fait état de procédures de développement de plusieurs instruments, veuillez utiliser cette case (commentaire) pour évaluer la description (et la justification) de la procédure de développement de chaque instrument de collecte de données ».
	É.2 : Elles peuvent être décrites, mais rencontrent-elles le but recherché? Sont-elles pertinentes et répondent-elles aux besoins d'expérimentation ou de types d'évidences à obtenir.	
	É.3 : Bien--mais pourrait-on ajouter "et justifiées" ?	
	<u>Commentaires concernant l'item</u> 1.4.4 Les procédures de collecte de données sont décrites	**Décision concernant l'item**
	É.1 : idem pour le fait qu'il peut y avoir plusieurs collectes.	Au regard du commentaire de É.3 **<u>voici notre nouvelle proposition concernant l'item</u>** : **1.4.4 Les procédures de collecte de données sont décrites et justifiées** Au regard du commentaire de É.1 **<u>nous avons inséré la remarque suivante dans la case commentaire</u>** : « Si le plan d'évaluation fait état de plusieurs procédures de collecte de données, veuillez utiliser cette case (commentaire) pour évaluer la description (et la justification) de chaque procédure de collecte de données ».
	É.2 : Il s'agit d'un critère légitime, mais que plusieurs réduisent à la plus simple expression par souci de concision des rapports.	
	É.3 : Bien--mais pourrait-on ajouter "et justifiées" ?	
	<u>Commentaires concernant l'item</u> 1.4.5 Les procédures d'analyse de données sont décrites	**Décision concernant l'item**

	É.1 : idem	Au regard du commentaire de É.3 **voici notre nouvelle proposition concernant l'item** : **1.4.5 Les procédures d'analyse de données sont décrites et justifiées** Au regard du commentaire de É.1 **nous avons inséré la remarque suivante dans la case commentaire** : « Si le plan d'évaluation fait état de plusieurs procédures d'analyse de données, veuillez utiliser cette case (commentaire) pour évaluer la description (et la justification) de chaque procédure d'analyse de données ».
	É.2 : Nous voyons plutôt dans un rapport le résultat de l'analyse des données bien plus qu'une description des procédures prises pour transformer les observations en conclusions et en recommandations. Ce critère peut sembler superflu.	
	É.3 : Bien--mais pourrait-on ajouter "et justifiées" ?	
Thème 2 : précision concernant les sources d'information et les procédures d'échantillonnage		
2.1 Description générale des sources d'information	**Commentaires conservant l'item 2.1.1 Les critères de sélection des différentes sources d'information sont clairement exposés**	**Décision concernant l'item**
	É.1 : N/A	Au regard du commentaire É.3, voici **notre nouvelle proposition pour cet item**: **2.1.1 Les critères de sélections des différentes sources d'information sont clairement justifiés**
	É.2 : confusion possible entre sources, informations et données de la part des répondants. This is a criterion very close from the # 1.4.1. It may sound like a duplication. We, in federal evaluations, do not usually develop those selection criteria of various sources of information that much.	
	É.3 : Bien--mais pourrait-on ajouter "et justifiés" ?	
	Commentaires concernant l'item 2.1.2 Les caractéristiques descriptives des sources d'information sont pertinentes au regard des objectifs et des questions d'évaluation	**Décision concernant l'item**
	É.1 : il faudrait formuler différemment. je ne comprends pas ce que signifie caractéristiques descriptives. là encore le caractère global de la question me semble poser problème si le répondant ne fournit pas de commentaire pour préciser un partiellement qui aurait été choisi.	**Nous conservons l'item tel quel**: **2.1.2 Les caractéristiques descriptives des sources d'information sont pertinentes au regard des objectifs et des questions d'évaluation**

	É.2: It might be interesting to link the descriptive characteristics versus the aims and the evaluation questions; nevertheless, is it really making an evaluation report more credible?	
	É.3 : Bien	
2.2 Description des procédures d'échantillonnage	**Commentaires concernant l'item 2.2.1 L'échantillonnage ou le choix des individus retenus est pertinent par rapport aux objectifs et aux questions d'évaluation**	**Décision concernant l'item**
	É.1 : les individus ne sont-ils pas des sources d'information?	Au regard du commentaire de É.1, **voici notre nouvelle proposition pour cet item**: **2.2.1 L'échantillonnage ou le choix des sources d'information retenus est pertinent par rapport aux objectifs et aux questions d'évaluation**
	É.2: This is a question of individual judgment requiring analysis from the person filling the PARE, this opening the door to all kinds of interpretations, levels of interpretation and different views of a same situation.	
	É.3 : Bien	
	Commentaires concernant l'item 2.2.2 Le type d'échantillonnage est précisé	**Décision concernant l'item**
	É.1 : N/A	Au regard du commentaire de É.2, voici **notre nouvelle proposition pour cet item**: **2.2.2 Le type de procédure d'échantillonnage est précisé**
	É.2 : que signifie type ici (aléatoire ou non, selon quelle méthode, ...)? ce serait à définir. This criterion becomes relevant if the evaluation process might have to be repeated with the same selection factors and the same scientific approach. Most of federal programs do not use sampling as the sampling reflect the program itself, the main stakeholders of the program and their users.	
	É.3 : Bien	
	Commentaires concernant l'item 2.2.3 L'auteur justifie et décrit le processus d'échantillonnage en ayant conscience de la représentativité de celui-ci	**Décision concernant l'item**
	É.1 : avoir conscience me semble peu précis. Donne des informations sur la représentativité de l'échantillon?	Au regard du commentaire de É.1, **voici notre nouvelle proposition pour cet item**:

	É.2: The assessment of the interpretation of this "justification and description" can be multiple and biased by the cultural, educational, experiential background of the meta-evaluator.	**2.2.3 L'auteur décrit le processus d'échantillonnage et justifie la représentativité de celui-ci**
	É.3 : Bien	
	<u>Commentaires concernant l'item</u> 2.2.4 L'auteur décrit les contraintes existantes relativement aux processus d'échantillonnage et mentionne les précautions qui ont été prises si l'échantillonnage n'est pas aléatoire	**Décision concernant l'item**
	É.1 : il y a deux questions dans une, cela pose problème. Il faudrait traiter le cas où l'échantillon est aléatoire dans une question, le cas où il ne l'est pas dans une autre. De plus, dans certains rapport il y a plusieurs échantillonnages, il serait alors impossible de répondre à la question.	Au regard des commentaires de É.1 et de É.2, **<u>voici les deux nouveaux items que nous proposons</u>**: **2.2.4 L'auteur décrit les contraintes existantes relativement aux processus d'échantillonnage** **2.2.5 L'auteur mentionne les précautions qui ont été prises si l'échantillonnage n'est pas aléatoire** En ce qui a trait aux situations dans le cadre desquels on retrouve plusieurs procédures d'échantillonnage, l'utilisateur de la grille peut utiliser la case commentaire pour les critiquer cas par cas.
	É.2: Valid criterion. It might also covers the limitations associated with the sampling process and the mitigation process used in case of non-randomized sampling.	
	É.3 : Bien	
Thème 3 : précision concernant la description du plan d'évaluation *(à la suite de la <u>modification de l'item 1.2.3</u> on ne parle plus de devis, mais de plan)*		
3.1 Description de la démarche méthodologique	**<u>Commentaires concernant l'item</u> 3.1.1 L'auteur présente les informations nécessaires pour bien rendre compte de la méthodologie**	**Décision concernant l'item**
	É.1 : la méthodologie a été abordée dans l'échantillonnage et dans les instruments. Cette question vise-t-elle autre chose? si oui il faut préciser quoi	Dans le cadre des thèmes précédents il était question de la description générale de la méthodologie. À présent nous rentrons dans les détails de cette dernière et nous nous

	É.2: The methodology is an important factor to better understand an evaluation and its results; it is important to have it clearly explained and understandable by a grade 10 college student (i.e. vocabulary used, expressions and process explanations, etc.). Answering no to this question might have the effect to lower the credibility of the whole report.	concentrons sur la qualité et la précision de sa description. Aussi, **<u>nous conservons l'item tel quel</u>**: **3.1.1 L'auteur présente les informations nécessaires pour bien rendre compte de la méthodologie**
	É.3 : Bien--mais pourrait-on ajouter "ainsi que de ses limites dans les circonstances de l'évaluation" ?	
	<u>Commentaires concernant l'item</u> 3.1.2 La démarche méthodologique est organisée	**Décision concernant l'item**
	É.1 : N/A	**<u>Nous conservons l'item tel quel</u>**: **3.1.2 La démarche méthodologique est organisée**
	É.2: Again, this is a question of judgment; in order to objectively assess this criterion, more information is necessary to define what is an "organized methodology".	
	É.3 : N/A	
3.2 Description de procédures d'obtention de données provenant d'angles différents (ex. recours à la triangulation)	**<u>Commentaires concernant l'item</u> 3.2.1 Le rapport fait état de procédures permettant de vérifier la stabilité des résultats dans le temps (triangulation temporelle)**	**Décision concernant l'item**
	É.1 : N/A	Nous voulons que les praticiens soient en mesure d'utiliser cette grille de méta-évaluation quel que soit le type d'évaluation et la méthodologie évaluative, aussi nous sommes consciente que certains items ne seront pas nécessairement pertinent dans tous les cas de figure. Aussi, **<u>nous conservons l'item tel quel et nous rajoutons un exemple entre parenthèse</u>**: **3.2.1 Le rapport fait état de procédures permettant de vérifier la stabilité des résultats dans le temps (triangulation temporelle [étude longitudinale])**
	É.2: I am not sure that every day's doers in evaluation will care about a temporal triangulation as a criterion for a good evaluation report.	
	É.3 : Un ou deux exemples pourraient être utiles pour la mise en œuvre de ce critère qui me semble pertinent mais pas tout à fait habituel (surtout quand il s'agit d'une évaluation formative).	
	<u>Commentaires concernant</u> l'item 3.2.2 Le rapport fait était de l'implication de plusieurs évaluateurs, intervenant de manière	**Décision concernant l'item**

	indépendante, au niveau de la collecte et/ou de l'analyse des données (accord inter-juges - triangulation des observateurs)	
	É.1 : N/A	**<u>Nous conservons l'item tel quel</u>:** **3.2.2 Le rapport fait était de l'implication de plusieurs évaluateurs, intervenant de manière indépendante, au niveau de la collecte et/ou de l'analyse des données (accord inter-juges - triangulation des observateurs)**
	É.2 : N/A	
	É.3 : Critère pertinent dans le cas d'une équipe d'au moins deux évaluateurs mais moins évident dans le cas d'une évaluation peu coûteux faite par un seul évaluateur.	
	<u>Commentaires concernant l'item</u> 3.2.3 Le rapport fait état de procédures d'obtention d'informations auprès de multiples informateurs et/ou diverses sources d'information objectives telles que des documents d'archives	**Décision concernant l'item**
	É.1 : N/A	**<u>Nous conservons l'item tel quel</u>:** **3.2.3 Le rapport fait état de procédures d'obtention d'informations auprès de multiples informateurs et/ou diverses sources d'information objectives telles que des documents d'archives (triangulation de sources ou d'informateurs)**
	É.2: This is the triangulation that is the most known, common and used. Nevertheless, other than in the methodology section of the report, triangulation of event and sources are more evaluable in the action (data collection and being concerns with having multiple sources) than in the definition (the details and explanations in the methodology).	
	É.3 : Bien.	
	<u>Commentaires concernant l'item</u> 3.2.4 Le rapport fait état du recours à de multiples méthodes de collectes de données, complémentaires ou contrastées	**Décision concernant l'item**
	É.1 : N/A	**<u>Nous conservons l'item tel quel</u>:** **3.2.4 Le rapport fait état du recours à de multiples méthodes de collectes de données, complémentaires ou contrastées (triangulation d'instruments ou de méthodes)**
	É.2: Comments made in 3.3.3 applies. The fact that the evaluation report stipulates that the concerns of multiple, complementary or contrasting methods have been taken into account may give credibility to the report. Nevertheless, it does not mean that the triangulation was relevant, well administered and conclusive for the end of the evaluation made.	

	É.3 : Bien.	
Thème 4 : <u>précision</u> concernant la collecte et le traitement des données		
4.1 Description de procédures garantissant la fiabilité des données recueillies	**4.1.1 Les procédures de collecte de données sont claires**	**Décision concernant l'item**
	É.1 : même commentaire que précédemment si plusieurs collectes sont utilisées	**<u>Nous conservons l'item tel quel</u>** : **4.1.1 Les procédures de collecte de données sont claires** Au regard du commentaire de É.1 **<u>nous avons inséré la remarque suivante dans la case commentaire</u>** : « Si le plan d'évaluation fait état de plusieurs procédures de collecte de données, veuillez utiliser cette case (commentaire) pour évaluer la clarté de chaque procédure de collecte de données ».
	É.2: If "devis d'évaluation" is the same thing as the "work plan" for the evaluation project, this is not necessarily transpiring in the final evaluation report. A "clear" data collection process may be interpreted differently by different people for a same statement and arriving to different assessment based on the human factor and on the context given by their background.	
	É.3 : Bien.	
	4.1.2 Le rapport fait état de procédures de vérification de la fiabilité des instruments	**Décision concernant l'item**
	É.1 : idem si plusieurs instruments sont utilisés	**<u>Nous conservons l'item tel quel</u>** : **4.1.2 Le rapport fait état de procédures de vérification de la fiabilité des instruments** Au regard du commentaire de É.1 **<u>nous avons inséré la remarque suivante dans la case commentaire</u>** : « Si le rapport d'évaluation fait état de plusieurs procédures de vérification de la fiabilité des instruments, veuillez utiliser cette case (commentaire) pour évaluer les procédures de vérification de la fiabilité de chaque instrument ».
	É.2: Reliability of instruments is rarely an element covered in the evaluation of the federal programs. This is a detailed procedure that is not required by the standards but that may be inserted in the methodology section. Most federal evaluator do not bother to include for the sake of saving space	
	É.3 : Il me semble que ce critère s'applique surtout aux instruments plus ou moins standardisés—qui sont mois pertinents à des évaluations formatives.	
	4.1.3 Les procédures d'analyse sont suffisamment décrites pour que l'on puisse les reproduire et obtenir les mêmes résultats	**Décision concernant l'item**
	É.1 : N/A	Au regard du commentaire de É.3, **<u>voici notre nouvelle proposition pour cet item</u>** : **4.1.3 Les procédures d'analyse sont suffisamment décrites et justifiées pour que l'on puisse les reproduire et obtenir les mêmes résultats**
	É.2: In an ideal world, the analysis process should be described enough in order to reproduce and obtain the same results for comparable evaluation processes. This criterion gives an insurance for reliability and validity.	

	É.3 : J'mettrais plutôt "suffisamment décrites et justifiées"	Au regard des commentaires précédents de É.1 **nous avons inséré la remarque suivante dans la case commentaire** : « Si le plan d'évaluation fait état de plusieurs procédures d'analyse de données, veuillez utiliser cette case (commentaire) pour évaluer la description (et la justification) de chaque procédure d'analyse de données afin de déterminer s'il est possible de les reproduire et d'obtenir les mêmes résultats ».
4.2 Description de procédures de contrôle des biais au niveau de la collecte et du traitement des données.	**4.2.1 Les sources d'information sont clairement spécifiées**	**Décision concernant l'item**
	É.1 : en quoi cette question est-elle différente de celle posée sur les sources dans une section précédente?	Au début de la grille nous discutons du niveau de précision de la description du processus de sélection et d'échantillonnage des sources d'information. Au niveau du sous-thème 4.2 il est question des diverses procédures de contrôle des biais. La spécification des sources d'information fait partie de ces procédures et c'est pour cette raison que nous mentionnons de nouveau la justification des sources d'information sélectionnées dans cette section-ci de la grille. Aussi, **nous conservons l'item tel quel** : **4.2.1 Les sources d'information sont clairement spécifiées**
	É.2 : To a certain extend, information sources might be identified; but is it really placing the evaluation report in a better situation in terms of avoiding experimental or evaluative bias?	
	É.3 : Bien.	
	4.2.2 Les inférences faites à partir des informations obtenues grâce aux instruments de collecte de données sont fiables	**Décision concernant l'item**
	É.1 : les inférences ne sont-elles pas plutôt issues des analyses des données?	**Nous conservons l'item tel quel** : **4.2.2 Les inférences faites à partir des informations obtenues grâce aux instruments de collecte de données sont fiables**
	É.2: We can be facing a question of bias in assessment caused by the human factor. How assessing that inferences from information obtained through data collection instruments are reliable? On which base? How to determine this specific reliability on an universal point of view?	
	É.3 : Bien.	
	4.2.3 Le rapport fait état de procédures de contrôle des biais	**Décision concernant l'item**
	É.1 : N/A	**Nous conservons l'item tel quel** :

	É.2: I usually see this mitigation aspect in the methodology section of the evaluation report. If this is what is targeted here, then the criterion is valid. Nevertheless, is stating this element in the evaluation report going to make the report more valid and reliable?	**4.3.3 Le rapport fait état de procédures de contrôle des biais**
	É.3 : J'ajouterais peut-être, "fait état de procédures et d'un esprit critique pertinents au contrôle des biais."	
4.3 Description de procédures de collecte et de traitement de données assurant la fidélité des informations obtenues vis-à-vis du programme (représentativité)	**4.3.1 Le rapport fait état de procédures de contrôle de la validité des inférences faites à partir de données obtenues à l'aide des instruments de collecte de données**	**Décision concernant l'item**
	É.1 : quelle différence avec 4.2.2?	Dans le cadre de l'item 4.2.2 il était question de la fiabilité des inférences faites à partir des données obtenues à l'aide des instruments de collecte de données. Ici, pour cet item, il s'agit de la validité des inférences faites à partir des données obtenues. Aussi, **nous conservons l'item tel quel** : **4.3.1 Le rapport fait état de procédures de contrôle de la validité des inférences faites à partir de données obtenues à l'aide des instruments de collecte de données**
	É.2: Point very specific. This is a very detailed criterion. Most of the reports I have seen in federal context do not specifically make such statements. It may sound that the evaluator producing this report insist to make sure that all elements have been put in place to prove the validity of the elements gathered through the evaluation process. The reader might become suspicious by this insistence on "insisting to proving validity" approach.	
	É.3 : Dans le cas de cet item, il s'agit peut-être davantage d'un esprit critique que de "procédures", ainsi que d'un souci de ne passer outre des données dans le cas d'inférences et de conclusions.	
	4.3.2 Les instruments de collecte de données sont appropriés pour les objectifs de l'évaluation	**Décision concernant l'item**
	É.1 : ici on retourne dans la méthodologie et l'instrumentation alors que les autres questions de cette section traitent de l'analyse.	Ici, il s'agit d'une question générale permettant de de déterminer si l'utilisation de ces instruments de collecte de données était, est ou sera bel et bien appropriés pour les objectifs d'évaluation. Aussi, **nous conservons l'item tel quel** : **4.3.2 Les instruments de collecte de données sont appropriés pour les objectifs de l'évaluation**
	É.2: Good criterion, but difficult to objectively assess. What is appropriate for you might not be appropriate for me so what would be the result obtained if it is due to the subjectivity of the meta-evaluators?	
	É.3 : Bien.	

	4.3.3 Le devis d'évaluation (incluant la méthodologie) permet de répondre adéquatement aux questions d'évaluation auxquelles il est censé répondre	**Décision concernant l'item**
	É.1 : cette question est d'un ordre très général, pourquoi est-elle ici?	Ici il s'agit d'une question générale portant sur l'ensemble du contenu du plan d'évaluation. Ainsi, au regard de l'analyse du contenu de chaque section du rapport d'évaluation rapportant le plan d'évaluation, il s'agit de déterminer si le plan d'évaluation a bel et bien permis, permet ou permettra de répondre adéquatement aux questions d'évaluation auxquels il est censé répondre. **Aussi, nous conservons l'item tel quel** : **4.3.3 Le devis d'évaluation (incluant la méthodologie) permet de répondre adéquatement aux questions d'évaluation auxquelles il est censé répondre**
	É.2: Relevant criterion. Nevertheless, how can we define the term "adequately"? How can it be made sure that there will be only one comprehension of the terms used in order to precisely assess the evaluation report?	
	É.3 : Bien.	
Thème 5 : limites méthodologiques mentionnées et expliquées		
Les limites méthodologiques sont prises en compte, le rapport fait état des obstacles techniques et logistiques.		**Décision concernant l'item**
	É.1 : N/A	**Nous conservons l'item tel quel**: **5.1 Les limites méthodologiques (menaces à la validité et fidélité) sont prises en compte (échantillonnage, limites des instruments, biais de l'évaluateur ou du contexte, données insuffisantes, etc.). Le rapport fait état des obstacles techniques et logistiques.**
	É.2: Relevant criterion. Almost essential to give some credibility to an evaluation report. Maybe, in addition to "limitations", the report should also make mention of the mitigation methods used to counter each limitation.	
	É.3 : Bien.	

Annexe 9. *Version finale de la grille d'analyse de rapports d'évaluation centrée sur le devis d'évaluation*

Thèmes	Critères principaux	Critères détaillés	Respect des critères				Commentaires
			N/A	Non	Part.	Oui	
1. Précision concernant la description générale de la démarche d'évaluation (cahier des charges) : ***[Les objectifs, les questions et la démarche de l'évaluation sont documentés avec suffisamment de précision pour qu'on puisse les identifier et les apprécier. Soit la description détaillée de :*** ***- l'organisation,*** ***- la collecte,*** ***- la mise au point, les réajustements,*** ***- la saisie et l'exploitation des données.]***	**1.1 Description des objectifs de l'évaluation**	1.1.1 Le mandat d'évaluation est clairement décrit	N/A	Non	Part.	Oui	
		1.1.2 Les objectifs d'évaluation sont clairs	N/A	Non	Part.	Oui	
		1.1.3 Les objectifs d'évaluation sont précis	N/A	Non	Part.	Oui	
	1.2. Justification du choix du devis d'évaluation	1.2.1 Le type d'évaluation choisi est pertinent pour répondre aux objectifs d'évaluation	N/A	Non	Part.	Oui	
		1.2.2 Le choix du type d'évaluation est justifié par l'auteur	N/A	Non	Part.	Oui	
		1.2.3 Le plan d'évaluation est fidèle au type d'évaluation choisi	N/A	Non	Part.	Oui	
	1.3. Présentation des questions d'évaluation	1.3.1 Les questions d'évaluation sont pertinentes au regard des objectifs d'évaluation	N/A	Non	Part.	Oui	
		1.3.2 Les questions d'évaluation sont claires	N/A	Non	Part.	Oui	
		1.3.3 Les questions d'évaluation sont précises	N/A	Non	Part.	Oui	
	1.4 Description de la méthodologie	1.4.1 Les sources d'information et les procédures d'échantillonnage sont décrites et justifiées	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état de plusieurs sources d'information et procédures d'échantillonnage, veuillez utiliser cette colonne (commentaire) pour évaluer la description (et la justification) de chaque source d'information et la procédure d'échantillonnage de cette dernière.

		1.4.2 Les procédures de sélection l'instrumentation sont décrites et justifiées	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état de la sélection de plusieurs instruments, veuillez utiliser cette colonne (commentaire) pour évaluer la description de chaque procédure de sélection des instruments.
		1.4.3 Les procédures de développement de l'instrumentation sont décrites et justifiées	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état du développement de plusieurs instruments, veuillez utiliser cette colonne (commentaire) pour évaluer la description des procédures de développement de chaque instrument.
		1.4.3 Les procédures de collecte de données sont décrites et justifiées	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état de multiples procédures de collecte de données, veuillez utiliser cette colonne (commentaire) pour évaluer la description de chaque procédure de collecte de données.
		1.4.4 Les procédures d'analyse de données sont décrites et justifiées	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état de multiples procédures d'analyse de données, veuillez utiliser cette colonne (commentaire) pour évaluer la description de chaque procédure d'analyse de données.
2. <u>Précision</u> concernant les sources d'information et les procédures d'échantillonnage ***[Les sources d'information utilisées dans l'évaluation sont décrites avec suffisamment de précision pour qu'on puisse apprécier la***	**2.1 Description générale des sources d'information**	2.1.1 Les critères de sélection des différentes sources d'information sont clairement justifiés	N/A	Non	Part.	Oui	
		2.1.2 Les caractéristiques descriptives des sources d'information sont pertinentes au regard des objectifs et des questions d'évaluation	N/A	Non	Part.	Oui	
	2.2 Description des procédures d'échantillonnage	2.2.1 L'échantillonnage ou le choix des sources d'information retenu est pertinent par rapport aux objectifs et aux questions d'évaluation	N/A	Non	Part.	Oui	

justesse des informations.]		2.2.2 Le type de procédure d'échantillonnage est précisé	N/A	Non	Part.	Oui	
		2.2.3 L'auteur décrit le processus d'échantillonnage et justifie la représentativité de celui-ci	N/A	Non	Part.	Oui	
		2.2.4 L'auteur décrit les contraintes existantes relativement aux processus d'échantillonnage	N/A	Non	Part.	Oui	
		2.2.5 L'auteur mentionne les précautions qui ont été prises si l'échantillonnage n'est pas aléatoire	N/A	Non	Part.	Oui	
3. <u>Précision</u> concernant la description du cadre méthodologique de l'évaluation	**3.1 Description de la démarche méthodologique**	3.1.1 L'auteur présente les informations nécessaires pour bien rendre compte de la méthodologie	N/A	Non	Part.	Oui	
		3.1.2 La démarche méthodologique est organisée	N/A	Non	Part.	Oui	
	3.3 Description de procédures d'obtention de données provenant d'angles différents (ex. recours à la triangulation)	3.3.1 Le rapport fait état de procédures permettant de vérifier la stabilité des résultats dans le temps (triangulation temporelle [ex. étude longitudinale])	N/A	Non	Part.	Oui	
		3.3.2 Le rapport fait était de l'implication de plusieurs évaluateurs, intervenant de manière indépendante, au niveau de la collecte et/ou de l'analyse des données (accord inter-juges - triangulation des observateurs)	N/A	Non	Part.	Oui	
		3.3.3 Le rapport fait état de procédures d'obtention d'informations auprès de multiples informateurs et/ou diverses sources d'information objectives telles que des documents d'archives	N/A	Non	Part.	Oui	

		(triangulation de sources ou d'informateurs)					
		3.3.4 Le rapport fait état du recours à de multiples méthodes de collectes de données, complémentaires ou contrastées (triangulation d'instruments ou de méthodes)	N/A	Non	Part.	Oui	
4. <u>Précision</u> concernant la collecte et le traitement des données ***[- La procédure de récolte d'informations a été développée, puis appliquée de manière à garantir la validité et la fiabilité des interprétations en vue de répondre aux questions d'évaluation.*** ***- Les informations qualitatives et quantitatives de l'évaluation sont analysées de manière appropriée et systématique, afin de répondre aux questions d'évaluation.]***	**4.1 Description de procédures garantissant la fiabilité des données recueillies** *[Le devis d'évaluation devrait inclure les procédures qui permettent de s'assurer que les informations collectées sont exactes. Ainsi, si l'évaluation est répliquée, les mêmes résultats seront obtenus.]*	4.1.1 Les procédures de collecte de données sont claires	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état de plusieurs procédures de collecte de données, veuillez utiliser cette colonne (commentaire) pour évaluer la clarté de chaque procédure de collecte de données.
		4.1.2 Le rapport fait état de procédures de vérification de la fiabilité des instruments	N/A	Non	Part.	Oui	Si le rapport fait état de plusieurs procédures de vérification de la fiabilité des instruments, veuillez utiliser cette colonne (commentaire) pour évaluer les procédures de vérification de la fiabilité de chaque instrument.
		4.1.3 Les procédures d'analyse sont suffisamment décrites et justifiées pour que l'on puisse les reproduire et obtenir les mêmes résultats	N/A	Non	Part.	Oui	Si le plan d'évaluation fait état de plusieurs procédures d'analyse de données, veuillez utiliser cette colonne (commentaire) pour évaluer la description (et la justification) de chaque procédure d'analyse de données afin de déterminer s'il est possible de les reproduire et d'obtenir les mêmes résultats.
	4.2 Description de procédures de contrôle des biais au niveau de la collecte et du traitement des données.	4.2.1 Les sources d'information sont clairement spécifiées	N/A	Non	Part.	Oui	
		4.2.2 Les inférences faites à partir des informations obtenues grâce aux instruments de collecte de données sont fiables	N/A	Non	Part.	Oui	

	[Le devis d'évaluation devrait contenir des procédures de contrôle des biais.]	4.2.3 Le rapport fait état de procédures de contrôle des biais	N/A	Non	Part.	Oui	
	4.3 Description de procédures de collecte et de traitement de données assurant la fidélité des informations obtenues vis-à-vis du programme (représentativité) *[Les informations collectées devraient représenter avec précision-justesse le programme évalué.]*	4.3.1 Le rapport fait état de procédures de contrôle de la validité des inférences faites à partir de données obtenues à l'aide des instruments de collecte de données	N/A	Non	Part.	Oui	
		4.3.2 Les instruments de collecte de données sont appropriés pour les objectifs de l'évaluation	N/A	Non	Part.	Oui	
		4.3.3 Le devis d'évaluation (incluant la méthodologie) permet de répondre adéquatement aux questions d'évaluation auxquelles il est censé répondre	N/A	Non	Part.	Oui	
5. <u>Limites méthodologiques</u> mentionnées et expliquées	**5.1 Les limites méthodologiques (menaces à la validité et fidélité) sont prises en compte (échantillonnage, limites des instruments, biais de l'évaluateur ou du contexte, données insuffisantes, etc.). Le rapport fait état des obstacles techniques et logistiques.**		N/A	Non	Part.	Oui	