

ÉCOLE NATIONALE D'ADMINISTRATION PUBLIQUE

STAGE AU SECRÉTARIAT DU CONSEIL DU TRÉSOR : RÉALISATION
D'UN BILAN DES PRATIQUES ÉVALUATIVES

RAPPORT DE STAGE PRÉSENTÉ
À MADAME LYNDA REY (CONSEILLÈRE ACADÉMIQUE) ET MONSIEUR
DANIEL OTIS (MANDANT)
COMME EXIGENCE FINALE DE L'ACTIVITÉ D'INTÉGRATION
ENP7969 S — STAGE

PAR
ALEXANDRE FORTIN

JUIN 2021

TABLE DES MATIÈRES

LISTE DES TABLEAUX	iv
LISTE DES SIGLES ET ABRÉVIATIONS.....	v
RÉSUMÉ.....	vi
INTRODUCTION.....	1
1. RECENSION DES ÉCRITS CONCERNANT LE CONTEXTE HISTORIQUE DE L'INSTITUTIONNALISATION DE L'ÉVALUATION AU QUÉBEC	3
1.1 Contexte historique et institutionnalisation de l'évaluation de programme au Québec ..	3
1.2 Les différents documents élaborés par le SCT depuis l'adoption de la Directive.....	8
2. DESCRIPTION DE L'ORGANISATION ET DU MANDAT.....	10
2.1 L'organisation : le SCT, le SSPALAP et la DGEAIGR	10
2.2 Le contexte du mandat.....	12
2.3 La nature du mandat	13
2.4 Les questions spécifiques du mandat	14
2.5 Les activités réalisées durant le stage	14
3. MÉTHODOLOGIE DE RECHERCHE ET D'INTERVENTION	16
3.1 Méthode de collecte, sources d'information et méthodes d'analyse des données	16
3.2 Limites méthodologiques et mesures d'atténuation adoptées	19
3.3 Matrice d'analyse	20
4. RÉSULTATS.....	22
4.1 Portrait général des évaluations.....	22
4.2 Traitement des enjeux de la pertinence, de l'efficacité et de l'efficience	25
4.3 Portrait des méthodes et sources utilisées dans les rapports.....	28
4.4 Méthodologie utilisée pour traiter les questions normalisées de l' <i>Outil</i>	30
4.5 Principaux constats du bilan.....	37
5. RECENSION DES ÉCRITS POST-STAGE	40
5.1 Le traitement des enjeux de pertinence, d'efficacité et d'efficience en évaluation.....	40
5.2 L'utilisation des groupes de discussion en évaluation.....	45
5.3 La triangulation des méthodes en évaluation	46

6. DISCUSSION ET RECOMMANDATIONS.....	46
CONCLUSION	50
BIBLIOGRAPHIE	51
ANNEXE 1 : ORGANIGRAMME DU SECRÉTARIAT DU CONSEIL DU TRÉSOR ET POSITIONNEMENT DE LA DGEAIGR AU 13 MAI 2021	55
ANNEXE 2 : LES QUESTIONS NORMALISÉES DE L’OUTIL D’ÉVALUATION DES PROGRAMMES ANALYSÉES LORS DE LA RÉALISATION DU BILAN	56
ANNEXE 3 : ORGANIGRAMME DES TÂCHES TEL QUE SPÉCIFIÉ DANS L’OFFRE DE SERVICE ET RÉALISATION EFFECTIVE RÉELLE	57
ANNEXE 4 : MÉTHODOLOGIE DÉTAILLÉE DU PROCESSUS DE CODAGE DES VARIABLES.....	58
ANNEXE 5 : LIMITES DU BILAN ET MESURES D’ATTÉNUATION ADOPTÉES ..	63

REMERCIEMENTS

Je tiens d'abord et avant tout à remercier mon mandant et responsable du stage Daniel Otis, qui m'a très bien accueilli et encadré tout au long du stage, et ce dans des circonstances parfois difficiles liées à la crise sanitaire. Les rencontres hebdomadaires et le suivi de près qu'il a effectué m'ont bien démontré qu'il avait réellement à cœur l'expérience de ses stagiaires.

Je remercie aussi ma conseillère académique Lynda Rey de m'avoir soutenu tout au long de la réalisation de mon stage et de m'avoir offert autant d'opportunités durant mon passage à l'ENAP. Lynda, tu auras été une professeure marquante lors de mon passage à la maîtrise. Je suis très reconnaissant pour tout le temps que tu as pris pour moi et je te souhaite de continuer d'être cette enseignante passionnée et généreuse que j'ai appris à connaître.

Je voudrais ensuite remercier ma collègue Ginette Côté, qui m'a beaucoup appuyé dans la réalisation de mon bilan. Son énergie positive et ses commentaires constructifs m'auront beaucoup aidé à relever des défis parfois complexes à résoudre.

Je remercie aussi mon conjoint Vincent qui m'a toujours fortement soutenu dans mon parcours et qui a eu à endurer quelques fois mon anxiété.

Finalement, je tiens à remercier mes parents de m'avoir toujours encouragé à trouver ma propre voie. Après plus de dix ans d'hésitations, je crois l'avoir trouvée.

LISTE DES TABLEAUX

Tableau 1 : Matrice d'analyse	21
Tableau 2 : Distribution des rapports d'évaluation déposés au SCT en 2019 et 2020	23
Tableau 3 : Statistiques descriptives — Rapports d'évaluation déposés au SCT en 2019 et 2020... 24	24
Tableau 4 : Traitement des enjeux et questions de pertinence, d'efficacité et d'efficience.....	27
Tableau 5 : Méthodes de collecte et sources utilisées dans les rapports d'évaluation.....	29
Tableau 6 : Méthodes de collecte et sources utilisées par question normalisée — Enjeu de la pertinence	33
Tableau 7 : Méthodes de collecte et sources utilisées par question normalisée — Enjeu de l'efficacité	34
Tableau 8 : Méthodes de collecte et sources utilisées par question normalisée — Enjeu de l'efficience	35
Tableau 9 : Taux de triangulation des méthodes selon la question normalisée.....	37
Tableau 10 : Questions spécifiques du bilan et principaux constats établis.....	37
Tableau A2.1 : Les questions normalisées de <i>l'Outil d'évaluation des programmes</i> analysées lors de la réalisation du bilan	56
Tableau A3.1 : Organigramme des tâches de la réalisation du bilan des pratiques évaluatives	57
Tableau A4.1 : Critères d'inclusion de la variable de la présence d'un comité d'évaluation et son type	58
Tableau A4.2 : Critères d'inclusion de la variable de l'origine du besoin d'évaluation	58
Tableau A4.3 : Combinaisons des sources d'information pour les sondages, les entrevues et les groupes de discussion	59
Tableau A4.4 : Critères d'inclusion des parties prenantes consultées.....	60
Tableau A4.5 : Critères d'inclusion et d'exclusion pour la question de l'efficacité opérationnelle (Q9)	61
Tableau A5.1 : Principales limites du bilan et mesures d'atténuation adoptées.....	63

LISTE DES SIGLES ET ABRÉVIATIONS

AS – Analyse stratégique

CE – Cadre d'évaluation

CSEP – Cadre de suivi et d'évaluation préliminaire

CT – Conseil du trésor

DGEAIGR – Direction de la gouvernance en évaluation, audit interne et gestion des risques

DGGGAR – Direction générale de la gouvernance en gestion axée sur les résultats

ENAP – École nationale d'administration publique

ETC. – Équivalent temps complet

GAR – Gestion axée sur les résultats

LAP – Loi sur l'administration publique

MO – ministère(s) et organisme(s)

PPBS – Planning, Programming and Budgeting System

PQI – Plan québécois des infrastructures

RE – Rapport d'évaluation

RPEP – Réseau des répondants en évaluation de programme

SCT – Secrétariat du Conseil du trésor

SSPALAP – Sous-secrétariat à la performance et à l'application de la Loi sur l'administration publique

VGQ – Vérificateur général du Québec

RÉSUMÉ

Ce rapport présente le bilan des pratiques évaluatives réalisé dans le cadre d'un stage effectué à la Direction de la gouvernance en évaluation, audit interne et gestion des risques (DGEAIGR) au sein du Secrétariat du Conseil du trésor (SCT). Le mandat avait pour objectif de dresser un portrait général des rapports d'évaluation déposés par les ministères et organismes québécois, de mesurer le taux de traitement des enjeux et des questions de pertinence, d'efficacité et d'efficience dans les évaluations ainsi que de répertorier les méthodes et les sources utilisées par les évaluateurs. À partir d'une analyse de contenu de 127 rapports déposés en 2019 et 2020, plusieurs constats ont pu être établis en lien avec ces objectifs. Notamment, le bilan a pu démontrer qu'une proportion relativement faible des évaluations s'appuyait sur un comité d'évaluation mixte, que la sous-traitance des évaluations par les ministères et organismes est une pratique plutôt rare et que l'enjeu de l'efficience est moins traité dans les rapports comparativement à ceux de pertinence et d'efficacité. Aussi, il a pu être constaté que les groupes de discussion sont beaucoup moins utilisés que d'autres méthodes de collecte de données et que les évaluateurs ne triangulaient pas toujours leurs méthodes de collecte de données pour répondre aux questions de *l'Outil d'évaluation des programmes*. Le rapport se termine sur une discussion sur les suites à donner, notamment en lien avec les tables rondes qui seront organisées dans les mois suivants la présentation au Réseau des répondants en évaluation de programme. Il ouvre finalement la porte à la réalisation future d'une méta-évaluation visant à mesurer différentes dimensions liées à la qualité des documents d'évaluation.

INTRODUCTION

Le présent document est le rapport final du stage que nous avons réalisé durant la session d'hiver 2021 à la Direction de la gouvernance en évaluation, audit interne et gestion des risques (DGEAIGR) au sein du Secrétariat du Conseil du trésor (SCT). Ce stage a été effectué dans le cadre de l'activité d'intégration obligatoire pour l'obtention de la maîtrise en administration publique à l'École nationale d'administration publique (ENAP) dont la concentration est l'évaluation de programmes.

Le stage s'est principalement axé sur la réalisation d'un bilan des pratiques évaluatives dans les rapports réalisés par les ministères et organismes (MO) au sein du gouvernement du Québec. Le rapport vise à présenter les fruits du travail réalisé et établir des liens avec les connaissances théoriques apprises lors du passage académique à l'ENAP. Il vise aussi, à partir des résultats issus du bilan, à élargir la réflexion en ancrant les constats établis dans la littérature grise et scientifique en évaluation de programmes. Le rapport de stage se divise donc en six principales sections.

La première section vise à présenter une recension des écrits concernant le parcours historique de l'institutionnalisation de l'évaluation de programme au Québec, allant du début des années 1970 à l'adoption de la Directive en 2014. Elle se termine par la présentation des principaux documents en matière d'évaluation de programmes qui ont été publiés dans les dernières années.

La seconde section présente plus en détail l'organisation d'accueil ainsi que le mandat qui nous a été confié. Pour la description de l'organisation, elle comprend à la fois le SCT, le Sous-secrétariat à la performance et à l'application de la Loi sur l'administration publique (SSPALAP) et la DGEAIGR. Pour la description du mandat, son contexte, sa nature et les

objectifs spécifiques sont d'abord présentés et ensuite, les activités réalisées dans le cadre du stage.

La troisième section vise à décrire précisément la méthodologie employée pour la réalisation du bilan. Elle comporte aussi les principales mesures d'atténuation des limites ainsi qu'une matrice d'analyse qui indique les différents indicateurs et sources de données utilisés pour répondre aux quatre questions spécifiques du bilan.

La quatrième section présente les principaux résultats du bilan. Les quatre volets du bilan sont abordés, soit le portrait général des rapports, le traitement des enjeux et des questions de pertinence, d'efficacité et d'efficience, la méthodologie utilisée dans les rapports ainsi que les méthodes et sources utilisées par question évaluative.

La cinquième section consiste en une recension post-stage, visant à ancrer les constats établis dans la littérature en évaluation de programme. Bien que cette étape ne fût pas exigée dans la réalisation du rapport de stage, nous avons cru pertinent de l'inclure puisqu'elle pourra contribuer à éclaircir certains constats et de renforcer les conclusions du bilan.

La sixième section vise à synthétiser les principaux constats et à discuter des principaux résultats ainsi qu'aux suites à donner après la présentation du bilan.

1. RECENSION DES ÉCRITS CONCERNANT LE CONTEXTE HISTORIQUE DE L'INSTITUTIONNALISATION DE L'ÉVALUATION AU QUÉBEC

Plutôt que de réaliser une seule recension des écrits, nous avons cru bon de la segmenter en deux parties, soit l'une présentée au début du rapport et l'autre à la suite de la section des résultats. La première visera le contexte historique de l'institutionnalisation de l'évaluation de programme au Québec ainsi que la documentation gouvernementale réalisée depuis la publication de la *Directive concernant l'évaluation des programmes dans les ministères et organismes* en 2014. La seconde concernera directement les constats qui découlent des résultats du bilan des pratiques évaluatives. Le but de cette segmentation est que ce rapport de stage soit plus utile auprès des décideurs. Cette section vise à présenter la première recension.

1.1 Contexte historique et institutionnalisation de l'évaluation de programme au Québec

Avant de présenter l'organisation dans laquelle nous avons réalisé notre stage et de préciser la nature du mandat, nous avons voulu aborder la question du parcours historique de l'institutionnalisation de l'évaluation au Québec. L'objectif ici est de mieux comprendre le contexte et l'environnement dans laquelle la DGEAIGR et le SCT ont émergé en tant qu'acteurs importants en évaluation.

L'évaluation de programme est une « démarche systématique de collecte et d'analyse de données et de renseignements sur les programmes afin d'améliorer ceux-ci ou de poser un jugement sur leur valeur et, ainsi, d'éclairer la prise de décision » (Secrétariat du Conseil du Trésor, 2014, p. 1011). Elle a débuté au Québec dans un contexte d'expansion considérable du rôle de l'État durant les décennies de 1960 et 1970, de la rationalisation des choix

budgétaires et de l'instauration du système de budgétisation de programme, que l'on appelle plus communément *Planning, Programming and Budgeting System* (PPBS). En effet, ces trois éléments « ont généré un besoin d'information pour l'allocation des ressources qui s'est traduit par le développement d'indicateurs pour recueillir des données sur les programmes » (Jacob et al., 2016). Dans ce contexte, les balbutiements de l'évaluation se sont déroulés dans un contexte d'essais-erreurs, où plusieurs tentatives d'institutionnalisation et de professionnalisation de l'évaluation ont eu lieu.

Marceau et al. (2000) considèrent que le début de l'histoire « officielle » de l'évaluation de programme au Québec a eu lieu en 1972, avec le début de l'application du PPBS (1972-1978) (p.486). Cette première période a consisté en l'implantation d'un système central d'allocation des ressources, où l'évaluation était intégrée au processus de budgétisation, et où le Conseil exécutif, sur recommandation du Conseil du trésor, exigeait des ministères la transmission de mémoires concernant des programmes particuliers (Ibid, p.487). Le contenu de ces mémoires intégrait principalement une description des besoins, une analyse des objectifs et une étude du choix des moyens utilisés et des solutions possibles (Ibid, p.487). Selon les auteurs, cette période fut un échec étant donné que les évaluations réalisées n'étaient pas en mesure de « déboucher sur une analyse critique de l'efficacité des programmes ». De plus, la pression était insuffisante sur les ministères pour que ces mémoires assument le rôle qu'ils devaient théoriquement avoir, soit celui de la contestation permanente (Ibid, p.488).

À la suite de cette période, les auteurs identifient une seconde tentative d'institutionnalisation de l'évaluation de programme, soit celle qui a eu lieu entre 1978 et 1983, et qui visait à centraliser les activités évaluatives au sein du Conseil du trésor afin de faciliter la réduction des dépenses (Ibid, p.488). Plus spécifiquement, quatre équipes d'évaluation correspondant aux quatre missions gouvernementales (sociale ; éducative et culturelle ; économique ; administrative et gouvernementale) avaient été créées. Aussi, un cadre pour le choix des programmes à évaluer avait été conçu et des questions d'évaluation précises et segmentées selon des catégories de programme avaient été élaborées (Ibid, p.489). Les auteurs

mentionnent que cet exercice n'a pas été considéré comme un grand succès malgré certaines réussites. Cela peut se refléter par le fait que cette période ait été « marquée par un arrêt des contrôles centraux sur l'évaluation » (Jacob et al., 2016).

En effet, la troisième période décrite par Marceau et al. (2000), entre 1984 et 1986, est celle de l'absence de mécanismes d'évaluation au sein du Conseil du trésor. En effet, elle fut marquée par une décentralisation des compressions budgétaires vers les ministères. Ces derniers devaient réaliser eux-mêmes des plans de rangement, visant à planifier leurs priorités et à ordonnancer lesdites coupures (Ibid, p.489). Au même moment, le Conseil du trésor a adopté le principe d'une politique-cadre en matière d'évaluation et a entrepris le lancement de consultations interministérielles sur le sujet (Ibid, p.491). Le projet en question reposait sur un partage des responsabilités entre les MO, responsables de l'exécution des évaluations, et le Conseil du trésor, responsable de la surveillance de l'application de la politique, de l'approbation des plans directeurs d'évaluation et des cadres analytiques d'évaluation (les « ancêtres » du PPE et du CSEP) ainsi que de la coordination des interventions interministérielles (Ibid, p.492). Malgré les bonnes intentions, le projet n'a finalement jamais été implanté, comme cela fut le cas aussi d'une autre tentative issue d'un rapport déposé en 1987 par le Centre de développement de la productivité dans le secteur public, qui proposait trois principes directeurs, dont l'adoption d'une politique d'évaluation de programme par le Conseil du trésor (Ibid, p.493).

Après ces multiples réformes, le Conseil du trésor a adopté en 1989 la *Directive concernant l'analyse de programme et la vérification interne*. Celle-ci, même si elle maintient la décentralisation de l'exécution des évaluations dans les MO, vient instaurer des mécanismes encadrant les fonctions d'évaluation et imposant certaines exigences telles que la transmission au Conseil du trésor d'un rapport sur l'état d'implantation de la fonction (Ibid, p.484). Cependant, le pouvoir du Conseil du trésor se limitait principalement à des considérations organisationnelles, les MO conservant une grande liberté dans le choix des

modalités d'évaluation (Ibid, p.485). Elle fut modifiée à deux reprises en 1994 (Arbour, 2016).

En 2000, la *Loi sur l'administration publique* (LAP) est adoptée à l'Assemblée nationale, instaurant un cadre de gestion axé sur les résultats visant notamment la reddition de comptes et à une utilisation optimale des ressources de l'État (L.R.Q., a-6.01, art. 2). On peut se rappeler que la gestion axée sur les résultats (GAR), selon le *Glossaire des termes usuels en mesure de performance et en évaluation*, est une « approche de gestion qui prend en considération les attentes exprimées par les citoyens en fonction des ressources disponibles et vise l'atteinte de résultats en fonction d'objectifs préalablement établis » (Leblanc-Constant et Bard, 2013, p. 11). La loi inclut deux principaux articles qui concernent l'évaluation de programme. D'abord, l'article 74 donne la possibilité au Conseil du trésor, avec approbation du Conseil des ministres, d'adopter une directive encadrant les MO sur la gestion des ressources humaines, budgétaires et matérielles « lorsqu'il estime qu'une question est d'intérêt gouvernemental ». Ensuite, l'article 75 de la loi donne de nouveaux pouvoirs au Conseil du trésor en matière d'évaluation de programme, ce dernier pouvant notamment « exiger la mise en place par un ministère ou un organisme de l'Administration gouvernementale d'un programme d'évaluation ». C'est ce dernier article qui faisait de la LAP, selon Jacob et al. (2016), l'instrument qui a permis au SCT de devenir « un acteur prédominant en évaluation », rôle renforcé par la suite avec l'adoption en 2014 de la *Directive concernant l'évaluation de programme dans les ministères et organismes*, qui relève de l'article 74. Enfin, l'article 78 de la LAP spécifie qu'un MO « doit fournir, sur demande du président du Conseil, tout renseignement utile à l'exercice des fonctions du président ou celles du Conseil du trésor » et qu'en vertu de cet élément, il « peut également exiger, aux mêmes fins, la préparation de documents », ce qui peut comprendre notamment des évaluations de programme.

En réponse à cette loi, le SCT a publié en 2002 des orientations concernant l'évaluation de programme dans une nouvelle directive et son document *L'évaluation de programme* :

document destiné aux dirigeants et dirigeantes de ministères et d'organismes, qui sont venus remplacer la Directive de 1994. Celui-ci visait plus spécifiquement à présenter les concepts fondamentaux de la discipline évaluative et les pratiques recommandées en la matière. Cependant, le document se limitait à assumer un rôle de guide des bonnes pratiques plutôt que constituer un document-cadre contraignant, les MO demeurant toujours responsables des évaluations (Secrétariat du Conseil du trésor, 2002, p. 4).

La décentralisation de la fonction évaluative demeure donc assez importante durant les années 2000 et au début de la décennie 2010. En effet, à l'exception des évaluations requises par le Conseil du trésor, le SCT « n'a pas exercé de façon systématique une surveillance à l'encontre des ministères eu égard à la fonction d'évaluation de programme ou des actions interministérielles » (Smits et Jacob, 2014, p. 85). Cependant, la flexibilité donnée en matière d'évaluation et les nouvelles responsabilités conférées par la LAP aux ministères afin d'instaurer un cadre de gestion axée sur les résultats ont « pu favoriser l'émergence d'initiatives de structuration infralégale se manifestant dans des règlements, des guides administratifs ou des politiques » (Arbour, 2016). En matière d'évaluation de programme, cela s'est traduit par la création de politiques ministérielles réalisées par les ministères eux-mêmes et encadrant leurs propres fonctions d'évaluation. Selon des estimations partielles, 16 ministères en possédaient une en 2011 selon un tableau du SCT repris par Smits et Jacob (2014, p. 82) et 13 ont mentionné en posséder une en 2012 selon Arbour (2016). Le contenu mentionné dans ces politiques peut varier selon les différents ministères, mais les principaux thèmes abordés concernent notamment le partage des responsabilités en évaluation, les questions traitées et la terminologie, la sélection des interventions, la planification annuelle, les étapes administratives et de production, la forme des rapports et les mécanismes de collaboration (Arbour, 2016). Aussi, en 2011, huit ministères s'étaient dotés d'un comité ministériel d'évaluation et seize d'un cadre de référence en évaluation de programme (Smits et Jacob, 2014, p. 82). En bref, cette période de décentralisation fut marquée par un soutien relativement peu important de la part des organismes centraux, cela s'incarnant par le fait

qu'il n'y avait qu'une seule personne responsable au SCT de l'encadrement des MO en évaluation de programme en 2011 (Ibid, p. 82).

Et puis finalement, en 2014, la *Directive concernant l'évaluation de programme dans les ministères et les organismes*, venant instaurer un encadrement accru des MO par rapport à leurs pratiques évaluatives ainsi qu'aux renseignements qu'ils doivent fournir au Conseil du trésor, a été adoptée par le gouvernement (Secrétariat du Conseil du Trésor, 2014). Plus particulièrement, la Directive vient préciser certaines obligations quant aux dépôts de certains documents et des informations disponibles à l'intérieur de ces derniers, tels que le plan pluriannuel d'évaluation de programme (PPE), le cadre d'évaluation (CE) et le cadre de suivi et d'évaluation préliminaire (CSEP). Elle confère aussi des responsabilités exclusives au SCT qui devra désormais notamment communiquer aux MO les besoins spécifiques d'information du Conseil du trésor, les consulter quant aux renseignements complémentaires à fournir et les modifications exigées, les assister dans l'élaboration de leurs CE et leurs CSEP ainsi que d'élaborer des outils pour les soutenir (Ibid, p.1013). En bref, entre l'approche centralisatrice adoptée initialement et celle de la décentralisation qui fut préconisée depuis les années 1980, la Directive se retrouve à mi-chemin entre les deux en créant un nouvel équilibre. Elle vise à la fois à maintenir une importante autonomie des MO concernant les modalités évaluatives et, en même temps, à accroître le rôle de surveillance du SCT en matière de la qualité des travaux évaluatifs au sein du gouvernement du Québec. Surtout, la Directive vient transformer le rôle du SCT, qui ne se limite désormais plus à la simple reddition de comptes, mais qui le transforme en tant qu'acteur important de soutien en matière de renforcement des capacités en évaluation de programme. Cette mutation s'illustre très bien par les différents documents et outils rédigés depuis son adoption.

1.2 Les différents documents élaborés par le SCT depuis l'adoption de la Directive

Depuis 2014, le SCT a élaboré différents documents afin de guider et d'encadrer les MO dans leurs travaux en évaluation de programme. D'un côté, cinq documents d'instructions obligatoires ont été réalisés, soit un pour chaque type des documents mentionnés dans la Directive (PPE, CSEP, CE et RE) (Secrétariat du Conseil du trésor, 2018c, 2017, 2018a, 2018d), en plus de l'*Outil d'évaluation des programmes* (Secrétariat du Conseil du trésor, 2018b). De l'autre côté, des guides sur le profil de compétences en évaluation de programme, l'élaboration des politiques publiques, l'évaluation des interventions interministérielles, la rédaction et le suivi des recommandations en évaluation ainsi qu'une trousse d'accompagnement à l'élaboration du PPE (Secrétariat du Conseil du trésor, 2019d, 2019b, 2019a, 2020b, s. d.) ont aussi été développés. Une multitude de séances d'information sur les instructions et les guides ont été organisées aussi par la DGEAIGR auprès des évaluateurs des MO.

Attardons-nous au document central au rôle de la DGEAIGR et en lien direct avec le mandat du stage, soit celui de l'*Outil d'évaluation des programmes*. Ce document précise les principales informations recherchées par le SCT dans les travaux d'évaluation en plus d'intégrer seize questions normalisées dont certaines sont obligatoires pour les évaluations requises par une décision du Conseil du trésor (voir annexe 2) (Secrétariat du Conseil du trésor, 2018b, p. 1). Il comporte des fiches qui viennent décrire les éléments de réponse recherchés et les sources de renseignements possibles pour chacun des enjeux couverts (pertinence de l'intervention, pertinence de la nature de l'intervention, cohérence interne, conformité opérationnelle, efficacité de la gouvernance, efficacité opérationnelle, efficacité de l'intervention et efficience relative de l'intervention). Aussi, des fiches complémentaires y sont incluses afin que les évaluateurs des MO soient en mesure de bien décrire les différentes composantes du programme ainsi que ses résultats prévus. En bref, il constitue, jusqu'à aujourd'hui, un document-cadre qui constitue le cœur de l'encadrement de l'évaluation de programme au sein du gouvernement du Québec.

2. DESCRIPTION DE L'ORGANISATION ET DU MANDAT

Après avoir décrit le parcours historique de l'institutionnalisation de l'évaluation au sein du gouvernement du Québec, la présente section vise à mieux décrire l'organisation dans laquelle le stage a été réalisé, notamment la mission et les principales responsabilités du SCT, du SSPALAP et de la DGEAIGR en lien avec l'évaluation de programme. Elle vise aussi à exposer le contexte du mandat en soulignant les enjeux qui ont mené à la nécessité de la réalisation d'un bilan des pratiques évaluatives. Enfin, elle a pour objectif de détailler le mandat du stage tel qu'il avait été prévu dans l'offre de service, y compris les questions spécifiques du bilan.

2.1 L'organisation : le SCT, le SSPALAP et la DGEAIGR

Le SCT a pour mission de soutenir les activités du Conseil du trésor et d'assister son président dans l'exercice de ses fonctions ainsi que le gouvernement dans ses responsabilités d'employeur (Secrétariat du Conseil du trésor, 2019c, p. 1). Par ses analyses et ses recommandations, il supporte le Conseil du trésor dans ses prises de décision afin d'assurer une saine gestion des fonds publics. Il soutient aussi les ministères et organismes (MO) quant à la mise en œuvre des orientations gouvernementales visant l'amélioration de la performance organisationnelle. Les responsabilités de soutien du SCT sont diverses et couvrent un large éventail d'actions et de domaines. Les principales fonctions touchent à la fois la préparation du budget annuel des dépenses et du Plan québécois des infrastructures (PQI) ; le soutien au gouvernement dans son rôle d'employeur, de régulateur et de coordonnateur des négociations de l'administration publique ; l'encadrement de la gestion des ressources humaines, financières et informationnelles et de la gestion contractuelle ; et la proposition de diverses recommandations sur les projets présentés par les MO (Ibid, p.1). Dans le cadre de ce rapport, la principale fonction de soutien du SCT qui retiendra notre

attention est celle « de faire évoluer le cadre de gestion axée sur les résultats et son application par les ministères et organismes » (Ibid, p.1).

En lien avec cette activité, un sous-secrétariat a été créé spécifiquement, soit le Sous-secrétariat à la performance et à l'application de la Loi sur l'administration publique (SSPALAP). Il a comme mandat de mettre en œuvre le cadre de gestion gouvernemental axé sur les résultats, visant la qualité des services aux citoyens, l'atteinte de résultats préalablement établis et favorisant l'imputabilité, la transparence et la reddition de compte des MO auprès de l'Assemblée nationale. Notamment, il a pour responsabilité de

(fournir) l'encadrement (politiques, directives et orientations) et le soutien (guides, meilleures pratiques et services-conseils) nécessaires (...) (et d'agir) à titre de répondant gouvernemental dans le cadre des fonctions qui visent à favoriser la performance en matière de planification stratégique, d'évaluation et de révision de programme, d'audit interne, de gestion des risques, de reddition de comptes publique et de déclaration de services aux citoyennes et citoyens (Secrétariat du Conseil du trésor, 2020a, p. 13).

À l'intérieur du SSPALAP, la Direction de la gouvernance en évaluation, audit interne et gestion des risques (DGEAIGR), de son côté, soutient la Direction générale de la gouvernance en gestion axée sur les résultats (DGGGAR) et le SSPALAP en matière d'évaluation de programme, d'audit interne et de gestion des risques. En lien avec le volet de l'évaluation, la Direction comporte d'une petite équipe d'une dizaine de personnes dédiées à l'exécution des responsabilités du SSPALAP en la matière. Sa principale mission est de soutenir, conseiller et encadrer les MO en matière d'évaluation de programme. Pour ce faire, la DGEAIGR agit sur différents plans. Sa principale responsabilité consiste à l'analyse et la rétroaction des documents réalisés par les MO et remis à la DGEAIGR à des fins d'amélioration continue. Les différents documents comprennent notamment les cadres de suivi et d'évaluation préliminaire (CSEP), les cadres d'évaluation (CE) et les rapports d'évaluation (RE). D'autres tâches font partie des responsabilités de la direction. Notamment, la DGEAIGR élabore différents documents, cadres, guides et d'autres documents visant à soutenir les MO dans leurs activités évaluatives. Aussi, la Direction réalise de temps en temps des mandats spécifiques visant les bonnes pratiques en évaluation de programme.

Finalement, une équipe volante de la DGEAIGR est responsable de soutenir les petits organismes (moins de 100 ETC. [équivalent temps complet]) ne disposant pas de fonction d'évaluation structurée. Cette équipe est notamment responsable de co-réaliser des évaluations en collaboration avec des responsables de ces MO.

Afin de mieux comprendre le positionnement de la DGEAIGR au sein du SCT, nous avons inséré, dans l'annexe 1, l'organigramme du SCT au 13 mai 2021, accompagné d'un encadré rouge autour de la DGEAIGR permettant de repérer la direction.

2.2 Le contexte du mandat

Depuis 2014, comme nous l'avons vu dans la section précédente, la Directive sert de cadre aux MO quant aux renseignements qu'ils doivent fournir au SCT concernant l'évaluation de leurs programmes (Secrétariat du Conseil du Trésor, 2014). Celle-ci a trois principaux objectifs, soit de contribuer à accroître l'utilisation des évaluations dans la prise de décision, d'améliorer la qualité des évaluations réalisées par les MO et de produire de la documentation pertinente, fiable et dans un format facilitant son utilisation au sujet des divers programmes (Ibid., p.1011). Cependant, des difficultés d'application de la Directive ont été rencontrées par les MO, le taux de conformité des MO à la Directive se situant à seulement 57 % lors de la mesure de départ du Plan stratégique de 2019-2023 (Secrétariat du Conseil du trésor, 2019c, p. 17). De plus, au-delà de la question de la conformité, des problématiques liées à la qualité des travaux évaluatifs et à la capacité organisationnelle des MO à bien les réaliser ont été identifiées par la DGEAIGR.

C'est dans ce contexte que plusieurs actions à réaliser par la DGEAIGR ont été prévues par le SCT dans la *Stratégie de gestion des dépenses de 2020-2021* dans le but notamment d'améliorer la qualité des évaluations réalisées par les MO. Il était notamment prévu qu'un bilan des PPE soit effectué, que les documents produits par les MO fassent l'objet d'une analyse et d'une rétroaction et qu'il y ait une bonification des services d'accompagnement,

de formations et d'outils en évaluation. Aussi, il était planifié qu'un bilan des pratiques d'évaluation des organisations soit réalisé et qu'une évaluation de la Directive soit entamée (Secrétariat du Conseil du trésor, 2020c, p. 29).

2.3 La nature du mandat

Le mandat concerne principalement l'avant-dernière action mentionnée au paragraphe précédent, c'est-à-dire que nous avons eu comme responsabilité principale la réalisation du bilan des pratiques d'évaluation. Plus précisément, le bilan avait pour principal objectif de dresser un portrait général des pratiques, méthodes et sources utilisées par les MO dans leurs évaluations remises au SCT. Il visait aussi à examiner le traitement des questions normalisées de *l'Outil d'évaluation des programmes* (Secrétariat du Conseil du trésor, 2018b) concernant les enjeux de la pertinence, de l'efficacité et de l'efficience et à répertorier les méthodes employées et les sources d'information utilisées par les MO pour répondre à chacune des questions. L'annexe 2 reprend le tableau décrivant les questions normalisées de *l'Outil* et leur caractère obligatoire en fonction du cycle de vie du programme et indique les questions qui devaient être analysées dans le bilan.

Il est à spécifier qu'au début du stage, nous avons eu à analyser la documentation existante afin de bien comprendre la nature du mandat qui nous a été confié. En effet, des travaux préliminaires avaient déjà été amorcés et portaient sur les enjeux de la pertinence et de l'efficacité. Une base de données regroupant les rapports d'évaluation (n=68) pour la période 2018-2019 a été constituée et des analyses descriptives ont été réalisées pour ces deux enjeux. Le mandat du stage était, à partir de la base de données précédente et des analyses rédigées, de réorienter l'analyse sur une présentation de données probantes à l'aide d'instruments de mesure de nature plus descriptive plutôt qu'appréciative.

2.4 Les questions spécifiques du mandat

Les questions spécifiques du mandat ont été divisées en quatre principaux volets et sont les suivantes :

1. Portrait général des évaluations :
 - 1.1. Quels sont les principaux éléments généraux et contextuels (distribution des rapports selon les MO, année du dépôt, dépôts préalables d'un CSEP ou d'un CE, comité d'évaluation, équipe d'exécution et origine du besoin d'évaluation) qui caractérisent les rapports d'évaluation déposés par les MO en 2019 et 2020 ?
2. Traitement des enjeux de la pertinence, de l'efficacité et de l'efficience :
 - 2.1. Combien de rapports ont abordé, d'une manière ou d'une autre, chacun des enjeux de la pertinence, de l'efficacité et de l'efficience ?
 - 2.2. Les questions normalisées liées aux enjeux de pertinence, d'efficacité et d'efficience de l'*Outil d'évaluation des programmes* ont-elles été traitées dans les rapports d'évaluation des MO ?
3. Portrait des méthodes et sources utilisées dans les rapports :
 - 3.1. Quelles méthodes de collecte de données et sources d'information les MO ont-ils utilisées dans leurs rapports d'évaluation ?
4. Méthodologie employée pour traiter les questions normalisées :
 - 4.1. Quelles ont été les méthodes et sources utilisées pour chacune des questions normalisées traitant la pertinence, l'efficacité et l'efficience ?
 - 4.2. Y a-t-il eu une triangulation des méthodes pour chacune des questions de l'Outil ?

2.5 Les activités réalisées durant le stage

L'annexe 3 présente l'organigramme des tâches qui avait été intégré dans l'offre de service au début du stage. Nous avons prévu les six principaux lots de travail. Premièrement, certaines tâches préliminaires avaient été planifiées dans le but de comprendre la nature du

stage, notamment la lecture des documents du SCT et ceux liés au bilan. Deuxièmement, il était question de la sélection des variables, de la détermination des ensembles catégoriels et de la planification des tableaux et graphiques. Troisièmement, l'élaboration de la grille d'analyse dans un fichier Excel était prévue, ce qui nécessitait notamment de fusionner les bases de données existantes. De plus, le codage de l'ensemble des rapports faisant partie de l'échantillon devait être réalisé. Quatrièmement, après avoir terminé l'analyse des rapports, nous devons concevoir les tableaux et les graphiques visant à présenter les données. La cinquième étape était de rédiger le document d'accompagnement Word du bilan. La sixième et dernière étape était de concevoir la présentation PowerPoint.

De manière générale, les tâches prévues reflètent les activités réelles qui ont été réalisées, à l'exception de quelques-unes. Aussi, certaines étapes n'ont pas pu être réalisées dans le cadre du stage étant donné que des difficultés ont été rencontrées par rapport au respect des délais, plus particulièrement celles de l'élaboration du fichier PowerPoint et de la présentation aux membres de l'équipe de la DGEAIGR. En effet, la taille de la base de données à analyser a été plus importante que prévu initialement et certaines vérifications imprévues (liées notamment la grille de détermination des questions traitées) ont nécessité des révisions itératives et ont contribué ainsi à ralentir le processus de codage. Cependant, l'ensemble des tâches prévues dans l'offre de service ont pu être finalement réalisées étant donné que nous avons obtenu un emploi contractuel à la suite du stage.

Il est à noter que durant le stage, nous avons pu contribuer à d'autres tâches en lien avec d'autres responsabilités de la DGEAIGR. Plus particulièrement, nous avons eu à participer aux travaux de l'équipe volante concernant le soutien à un MO spécifique pour l'élaboration d'un cadre d'évaluation (CE). Ces activités ne seront pas présentées dans le rapport de stage, qui se centre plus spécifiquement sur la réalisation du bilan des pratiques d'évaluation.

3. MÉTHODOLOGIE DE RECHERCHE ET D'INTERVENTION

Cette section présente la méthodologie utilisée dans le cadre de la réalisation du bilan. Elle reprend plusieurs parties de la section « méthodologie » rédigée dans le bilan, en ajoutant certaines précisions et en retirant certaines informations déjà mentionnées. La section se divise de la manière suivante. D'abord, la méthode de collecte de données, les sources d'information et les méthodes d'analyse sont abordées. Ensuite, les principales limites méthodologiques sont mentionnées brièvement et les mesures d'atténuation prises sont présentées. Finalement, une matrice d'analyse a été ajoutée afin de présenter de façon synthétique l'ensemble des volets, des questions spécifiques, les indicateurs et les sources d'information utilisés dans le cadre du bilan.

3.1 Méthode de collecte, sources d'information et méthodes d'analyse des données

3.1.1. Méthodes de collecte de données

Concernant les méthodes de collecte de donnée utilisées, le bilan s'est appuyé principalement sur une revue documentaire des rapports et des cadres d'évaluation réalisés par les MO. Aussi, des données du tableau de suivi ont été intégrées à la grille d'analyse, permettant de compléter la collecte de données. Aucune autre méthode de collecte primaire, telle que des entrevues ou des sondages, n'a été utilisée dans le cadre de ce bilan.

3.1.2. Constitution de la base de données et les sources d'information utilisées

Une recension de l'ensemble des rapports d'évaluation déposés entre 2019 et 2020 a été effectuée. Il est à noter que les rapports d'évaluation ayant été publiés en 2018 n'ont pas été pris en compte. Cette décision se justifie par le fait que l'*Outil d'évaluation des programmes*

a été publié par le SCT durant la même année et qu'un certain délai pouvait être nécessaire avant son utilisation et son application par les MO.

Afin de répertorier l'ensemble des rapports déposés en 2019 et 2020, la base de données interne à la DGEAIGR a été utilisée. Celle-ci consiste en un tableau de suivi qui indique des informations sur tous les différents types de documents reçus par la DGEAIGR. Les sources de données alimentant le tableau de suivi sont les plans pluriannuels d'évaluation (PPE) déposés annuellement au SCT. Trois filtres ont été utilisés :

- les rapports d'évaluation pour les types de documents,
- les années 2019 et 2020 pour la date de dépôt de la dernière version du rapport
- la version déposée, qui ne devait pas être une version de travail.

Au total, 129 rapports d'évaluation ont été répertoriés dans le tableau de suivi et ont été par la suite retrouvés dans le répertoire numérique commun où sont entreposés ceux-ci (N=129). 127 d'entre eux ont été transférés dans un dossier à des fins d'analyse (2 rapports ont été exclus pour des erreurs techniques tels qu'une mauvaise classification du type de document dans le tableau de suivi et un problème d'enregistrement de fichier) (n=127). Par la suite, pour chacun des rapports retenus, nous avons vérifié la présence d'un cadre d'évaluation dans le tableau de suivi. Au total, 109 rapports ont été précédés d'un cadre d'évaluation (CE). Tous les CE en question ont été aussi transférés dans un dossier à des fins d'analyse. Pour les 18 autres rapports, aucun cadre n'a été repéré dans la base de données interne de la DGEAIGR.

Comme il a été mentionné précédemment, les rapports d'évaluation sont la principale source d'information du bilan. Toutefois, les cadres d'évaluation ont aussi été utilisés pour certaines variables lorsque l'information n'était pas disponible dans les rapports. De plus, certaines variables spécifiques aux cadres ont été codées. Finalement, certaines données du tableau de suivi ont été fusionnées à la grille d'analyse afin de disposer de variables additionnelles, telles

que le caractère interministériel du programme évalué, le budget annuel moyen estimé, le type d'obligation et le type de MO (ministère ou organisme).

3.1.3 Les méthodes d'analyse des données

Concernant les méthodes d'analyse des données, c'est principalement l'analyse de contenu qui a été utilisée dans la réalisation du bilan. En effet, « l'analyse de contenu est une technique de recherche servant à la description objective, systématique et quantitative du contenu manifeste des communications » (Leray et Bourgeois, 2016, p. 428). Elle a été choisie étant donné qu'elle est particulièrement adaptée pour la nature du mandat, qui est de quantifier des informations de nature qualitative, et qu'en ce sens, les pratiques qui lui sont associées peuvent contribuer à renforcer la validité interne des résultats.

Des variables spécifiques ont donc été sélectionnées et des ensembles catégoriels ont été élaborés. Des définitions opérationnelles des catégories et des procédures de codage ont été établies. Pour les ensembles catégoriels, certains critères ont été pris en compte, tels que leur exhaustivité et le caractère mutuellement exclusif des catégories. L'annexe 4 décrit de manière exhaustive les principales variables sélectionnées dans le cadre du bilan ainsi que la méthodologie du processus de codage de celles-ci. Pour certaines d'entre elles, il définit spécifiquement chacun des éléments de l'ensemble catégoriel ainsi que les critères d'inclusion qui ont été appliqués. Aussi, le processus du codage des questions traitées y est expliqué ainsi que la segmentation des sections des rapports qui y sont associées pour le codage des méthodes par question.

Il est aussi à noter que la réalisation du bilan s'est appuyée sur les travaux antérieurs. La base de données a été reprise en partie et regroupée en une seule grille d'analyse. Notamment, certaines typologies élaborées de manière inductive dans les analyses précédentes ont été reprises afin d'effectuer un codage déductif pour la collecte des données. La base de données a été aussi utile afin d'effectuer des tests préliminaires sur une dizaine de rapports afin

d'assurer la validité interjuge du codage (fiabilité du codage entre plusieurs codeurs différents) pour certaines variables. Pour les variables de nature plus subjective, le pourcentage d'accords a été calculé et lorsque celui-ci n'était pas satisfaisant (en dessous de 75%), certains éclaircissements ont été apportés afin de renforcer la validité interne du bilan.

Finalement, à la suite de la réalisation du codage, la réalisation d'une multitude de tableaux de statistiques descriptives aura permis d'analyser les données, de repérer les principales tendances et d'établir de principaux constats.

3.2 Limites méthodologiques et mesures d'atténuation adoptées

Le bilan comporte certaines limites sur le plan méthodologique qui ont été prises en compte dans la présentation des résultats. Des mesures d'atténuation ont été adoptées pour les limites les plus importantes afin de réduire le risque de résultats erronés et de renforcer la fiabilité du codage, assurant ainsi une validité raisonnable des résultats. Le tableau situé dans l'annexe 5 présente les limites initiales, les mesures d'atténuation adoptées ainsi que les limites résiduelles. Si l'on résume brièvement, les principales mesures d'atténuation appliquées sont :

- la réalisation de tests préliminaires et l'appel à un regard extérieur par l'entremise d'évaluateurs expérimentés ;
- l'élaboration conjointe de définitions et des critères d'inclusion pour les variables les plus sensibles ;
- l'adoption d'une compréhension large des questions de l'*Outil* (inclusion des questions apparentées) et l'élaboration d'une grille de détermination des questions traitées visant à définir les critères de décision permettant de considérer si les questions de l'*Outil* ont été traitées ou non, et comment (par exemple, le libellé de la question ou des sections spécifiques du rapport).

Malgré les mesures employées visant une analyse la plus robuste et rigoureuse possible, il demeure qu'on ne peut garantir avec certitude la validité interne pour l'ensemble des variables étant donné qu'une seule personne a participé au codage des rapports.

3.3 Matrice d'analyse

Le tableau 1 ci-dessous reprend les questions spécifiques posées dans la sous-section liée à la nature du mandat et présente les indicateurs et les sources d'information qui ont été utilisés pour répondre à celles-ci.

Tableau 1 : Matrice d'analyse

Volets	Questions	Indicateurs	Sources d'information
1. Portrait général des évaluations	1.1 Quels sont les principaux éléments généraux et contextuels qui caractérisent les rapports d'évaluation déposés par les MO en 2019 et 2020 ?	<p>Nb et % de ministères et organismes ayant déposé au moins un rapport d'évaluation en 2019 et 2020</p> <p>Nb et % de rapports en fonction...</p> <ul style="list-style-type: none"> • du type de MO • des MO les plus actifs • de l'année du dépôt à la DGEAIGR • du dépôt préalable d'un CSEP/d'un CE • de la présence et le type de comité d'évaluation • de l'équipe d'exécution • de l'origine du besoin d'évaluation 	Rapports d'évaluation Cadres d'évaluation Tableau de suivi de la DGEAIGR
2. Traitement des enjeux de la pertinence, de l'efficacité et de l'efficience	<p>2.1 Combien de rapports ont abordé, d'une manière ou d'une autre, chacun des enjeux de la pertinence, de l'efficacité et de l'efficience ?</p> <p>2.2 Les questions normalisées liées aux enjeux de pertinence, d'efficacité et d'efficience de l'<i>Outil d'évaluation des programmes</i> ont-elles été traitées dans les rapports d'évaluation des MO ?</p>	<p>Nb et % des rapports traitant chacun des enjeux de pertinence, d'efficacité et d'efficience</p> <p>Nb et % de rapports traitant chacune des questions de l'<i>Outil</i> (pertinence, efficacité, efficience)</p>	Rapports d'évaluation
3. Portrait des méthodes et sources utilisées dans les rapports	3.1 Quelles méthodes de collecte de données et sources d'information les MO ont-ils utilisées dans leurs rapports d'évaluation ?	<p>Nb et % de rapports en fonction...</p> <ul style="list-style-type: none"> • de la présence de sondage, d'entrevues, de groupe de discussion et d'autres méthodes • des parties prenantes consultées pour chacune méthode 	Rapports d'évaluation
4. Méthodologie employée pour les questions normalisées	<p>4.1 Quelles ont été les méthodes et sources utilisées pour chacune des questions normalisées traitant la pertinence, l'efficacité et l'efficience ?</p> <p>4.2 Y a-t-il eu une triangulation des méthodes pour chacune des questions de l'<i>Outil</i> ?</p>	<p>Pour chacune des questions : nb et % de rapports en fonction...</p> <ul style="list-style-type: none"> • de l'utilisation des méthodes suivantes pour répondre à la question : la revue documentaire, l'analyse des données administratives, le sondage, l'entrevue, le groupe de discussion et autres méthodes • des parties prenantes consultées pour les sondages, les entrevues et les groupes de discussion pour répondre à la question <p>Taux de triangulation des méthodes de collecte de données pour répondre à chacune des questions</p>	Rapports d'évaluation

4. RÉSULTATS

Cette section vise à présenter les résultats du bilan des pratiques évaluatives. Elle consiste en une version abrégée et modifiée partiellement de la même section issue du rapport rédigé à l'interne. Seules les principales interprétations seront présentées afin de synthétiser la présentation des résultats. Elle a été divisée en cinq sous-sections, en lien avec les questions spécifiques du mandat.

Premièrement, un portrait général des rapports d'évaluation déposés au SCT sera dressé. Dans un second temps, nous présenterons des données sur le traitement des questions de l'*Outil* relatives aux enjeux de la pertinence, de l'efficacité et de l'efficience. Ensuite, les résultats issus du recensement des méthodes de collecte de données et des sources d'information utilisées par les MO seront présentés et analysés. Pour la quatrième sous-section, les méthodes et sources utilisées dans les rapports d'évaluation pour répondre à chacune des questions de l'*Outil* seront abordées. Finalement, une synthèse des principaux constats sera présentée dans un tableau.

4.1 Portrait général des évaluations

Afin de répondre à la question concernant les caractéristiques générales et contextuelles des évaluations, nous avons procédé de la manière suivante. Tout d'abord, la distribution des rapports a été présentée dans l'objectif de bien comprendre le profil des MO réalisant des évaluations. Ensuite, nous avons analysé les résultats de manière globale pour chacune des variables descriptives en question.

4.1.1 Distribution des rapports d'évaluation

Le tableau 2 présente les résultats de la distribution des 127 rapports. Nous pouvons en déduire les constats suivants :

- Parmi les 20 ministères, 17 ont déposé au moins un rapport d'évaluation en 2019 et 2020. Concernant les 42 organismes, seulement 9 ont fait de même.
- La majorité des rapports déposés (85,8 %) provient des ministères. Seulement 18 rapports (14,2 %) ont été réalisés par les organismes.
- Un seul ministère a déposé à lui seul 40 des 127 rapports, soit 31,5 % de l'ensemble des rapports.
- Les 5 MO les plus actifs ont déposé 78 rapports sur l'ensemble des 127 reçus par le SCT (61,4 %).

Variable	Catégories	Nb	%
Ministères	Ministères ayant déposé au moins un rapport	17	85,0 %
	Ministères n'ayant pas déposé de rapport	3	15,0 %
	Rapports déposés par les ministères	109	85,8 %
Organismes	Organismes ayant déposé au moins un rapport	9	21,4 %
	Organismes n'ayant pas déposé de rapport	33	78,6 %
	Rapports déposés par les organismes	18	14,2 %
Rapports déposés — Principaux MO	Rapports déposés par le premier MO	40	31,5 %
	Rapports déposés par les 3 premiers MO	60	47,2 %
	Rapports déposés par les 5 premiers MO	77	60,6 %
	Rapports déposés par les 10 premiers MO	99	78,0 %
Ensemble des rapports		127	100,0 %

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.1.2 Portrait général des évaluations

Le tableau 3 présente un portrait général de l'ensemble des variables descriptives et pour chacune des catégories qui y sont associées. On peut constater notamment que :

- Davantage de rapports ont été déposés au SCT en 2019 (73) qu'en 2020 (54).

- 20 rapports sur 127 (ou 15,7 %) ont été précédés d'un CSEP.
- Au moins 85,8% des rapports ont été précédés d'un cadre d'évaluation (CE). Pour les 18 rapports restants (14,2%), le non-dépôt des CE ne se traduit pas nécessairement par leur non-existence. En effet, il est possible que certains CE aient été réalisés avant la création de la DGEAIGR en 2017, et donc qu'ils ne soient pas présents dans la base de données pour cette raison.
- Une grande majorité des rapports, soit au moins 106 sur 127 (83,5 %), ont été réalisés en s'appuyant sur un comité d'évaluation. La plupart des comités étaient composés exclusivement de membres provenant de l'interne (MO, direction du programme ou bénéficiaire) (53,5 %). Une partie des comités comportait des membres provenant de l'externe (partenaires non-bénéficiaires, représentants des usagers, autres MO impliqués, etc.) (26 %). De leur côté, pour 5 rapports (3,9 %), les informations fournies ne permettaient pas de préciser le type de comité (interne ou mixte), même si son existence était mentionnée. Enfin, 21 évaluations (16,5 %) ne mentionnaient aucun comité d'évaluation dans le rapport ainsi que le cadre le cas échéant.
- Seulement 7 rapports ont été exécutés par une entité externe au MO (milieu universitaire, des firmes spécialisées en évaluation de programme, autres organismes publics). Il est à noter que cela n'inclut pas le recours à des ressources externes (consultants ou sondeurs) afin de réaliser seulement une partie de l'évaluation.
- La majorité des évaluations répond d'abord et avant tout à des besoins internes au MO (55,9 %). De l'autre côté, environ le quart des évaluations sont requises par une décision du CT et 7,9 % sont celles de programmes autorisés par le CT, mais sans obligation de réalisation du rapport. L'origine du besoin d'évaluation pour 14 rapports (11 %) concerne d'autres exigences telles qu'un décret du Conseil des ministres, une recommandation du VGQ ou une loi constitutive.

Tableau 3 : Statistiques descriptives — Rapports d'évaluation déposés au SCT en 2019 et 2020

Variable	Catégorie	Nb	%
----------	-----------	----	---

Années du dépôt	2019	73	57,5 %
	2020	54	42,5 %
Dépôts préalables	Dépôt préalable CSEP	20	15,7 %
	Dépôt préalable CE	109	85,8 %
Comité d'évaluation	Présence d'un comité...	106	83,5 %
	... interne	68	53,5 %
	... mixte	33	26,0 %
	... avec composition non précisée	5	3,9 %
	Pas de comité spécifié dans le CE et/ou le RE	21	16,5 %
Type d'exécution	Exécution interne	120	94,5 %
	Exécution externe	7	5,5 %
Origine du besoin d'évaluation	Évaluations requises par une décision du CT	32	25,2 %
	Programmes autorisés par le CT — Évaluations non obligatoires	10	7,9 %
	Besoin d'initiative	71	55,9 %
	Autre exigence	14	11,0 %
Ensemble des rapports		127	100,0 %

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.2 Traitement des enjeux de la pertinence, de l'efficacité et de l'efficience

Maintenant que nous avons brossé un portrait général des évaluations déposées au SCT en 2019 et 2020, analysons les enjeux et les questions traitées dans les rapports.

Le tableau 4 présente le taux de traitement des enjeux et des questions de pertinence, d'efficacité et d'efficience pour l'ensemble des rapports déposés en 2019 et 2020. Il permet d'abord de constater qu'une partie importante des évaluations (71,7 %) a traité l'enjeu de la pertinence et que la quasi-totalité (96 %) a abordé l'efficacité. Il indique aussi qu'uniquement 33 rapports (26 %) ont abordé l'enjeu de l'efficience. Ce résultat n'est pas tant surprenant étant donné que contrairement aux deux autres enjeux, aucune question d'efficience n'est obligatoire dans l'*Outil*.

Pour l'enjeu de la pertinence, parmi les quatre questions liées à celui-ci, celle de l'analyse des besoins et du contexte (Q1) a été la plus abordée avec un taux de traitement de 63 %. La

question de la pertinence de la nature de l'intervention (Q4) a été davantage traitée que celles de la cohérence avec le MO et les orientations gouvernementales (Q2) et de l'analyse des chevauchements (Q3). Pourtant, cette question n'est pas obligatoire pour les évaluations dont le cycle de vie du programme est la mise en œuvre, alors que c'est le cas pour les deux dernières. En réalité, ce haut traitement pour l'enjeu de pertinence s'explique en grande partie par le fait que certains MO traitent systématiquement la question indirectement, par exemple en mesurant l'utilité perçue du programme par les bénéficiaires et la clientèle.

Pour les questions d'efficacité de la gouvernance, elles ont été relativement peu abordées par les interventions interministérielles, seulement 3 des 12 rapports les concernant (25 %) ont traité celle de la définition des rôles et responsabilités des parties prenantes (Q7) et 4 rapports (33,3 %) ont traité celle du rôle de leader de l'organisation responsable (Q8). Tous ces rapports ont été réalisés en 2020, ce qui implique qu'aucun rapport de 2019 n'a abordé les deux questions de l'*Outil*. Il faut d'ailleurs se rappeler que le document *Évaluation des interventions interministérielles : Guide d'accompagnement* (Secrétariat du Conseil du trésor, 2019a) a été publié seulement en juillet 2019 par la DGEAIGR.

Les questions de l'efficacité opérationnelle (Q9) et celles de l'atteinte des objectifs et des effets visés (Q10/Q11) ont été traitées dans une grande majorité des rapports, avec des taux de traitement de 90,6 % et 75,4 % respectivement. Pour l'efficacité opérationnelle, le taux élevé s'explique en partie par le nombre important de questions apparentées abordant l'enjeu sous différents angles : questions de mise en œuvre, sur la qualité des extraits, sur le bilan des extraits, sur les délais opérationnels ou sur l'atteinte de cibles liées aux extraits. Aussi, la fusion des questions 10 (atteinte des objectifs) et 11 (effets visés) a contribué à faire augmenter le taux de traitement. Concernant les effets non intentionnels (Q12), seulement 11 des 127 rapports (8,7 %) l'ont traité.

Les quatre questions normalisées d'efficience relative de l'intervention de l'*Outil d'évaluation des programmes* sont très peu utilisées par les MO, le nombre de rapports

traitant chacune d'entre elles ne dépassant pas 2. Cependant, 29 rapports (22,8 %) ont abordé l'efficacité sous un autre angle (questions se limitant à la mesure de l'efficacité du programme évalué, questions liées à l'efficacité opérationnelle, etc.). On peut se rappeler que l'efficacité relative se définit comme étant le « résultat de la comparaison entre les efficacités de deux ou plusieurs interventions poursuivant les mêmes objectifs ou ayant des raisons d'être ou des natures d'intervention similaires ou proches » (Leblanc-Constant et Bard, 2013, p. 7). En bref, alors que les questions normalisées de l'*Outil* concernent l'efficacité relative de l'intervention, la plupart des rapports abordant l'enjeu de l'efficacité l'ont plutôt traité en se limitant à la mesure de l'efficacité du programme évalué.

Tableau 4 : Traitement des enjeux et questions de pertinence, d'efficacité et d'efficacité				
Enjeux	Questions	Traitement – Questions	Rapports analysés	Traitement – Enjeux
Pertinence	Q1 – Besoins et contexte	63,0 %	127	71,7 %
	Q2 – Cohérence MO et orientations gouvernementales	49,6 %	127	
	Q3 – Chevauchements des programmes	46,5 %	127	

	Q4 – Pertinence de la nature de l'intervention	51,2 %	127	
Efficacité de la gouvernance	Q7 – Définition adéquate des rôles et responsabilités des parties prenantes	25,0 %	12	96,0 %
	Q8 – Exercice du rôle de leader de l'organisation responsable	33,3 %	12	
	Autres questions d'efficacité de la gouvernance (interMO)	25,0 %	12	
Efficacité	Q9 – Efficacité opérationnelle	90,6 %	117	
	Q10/Q11 – Atteinte des objectifs/effets visés	75,4 %	126	
	Q12 – Effets non intentionnels	8,7 %	127	
Efficience	Q13 – Coûts des extrants/autres programmes	1,6 %	127	26,0 %
	Q14 – Coûts des effets/autres programmes	1,6 %	127	
	Q15 – Mêmes résultats à moindre coût	0,8 %	127	
	Q16 – Meilleurs résultats sans coûts supplémentaires	0,8 %	127	
	Autres questions d'efficience	22,8 %	127	

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.3 Portrait des méthodes et sources utilisées dans les rapports

Cette sous-section vise à décrire les méthodes de collecte de données et sources d'information qui ont été utilisées dans les rapports d'évaluation déposés en 2019 et 2020. Les données les concernant sont présentées dans le tableau 5 situé ci-dessous. Concernant la typologie des parties prenantes, on peut rappeler que le tableau A4.4 situé en annexe 4 définit chacune de celles-ci.

Parmi les trois méthodes de collecte de données, le **sondage** (terme générique qui regroupe aussi les questionnaires ou les enquêtes et dont le mode d'administration peut être soit par téléphone, sur format papier ou sur le web) a été la principale utilisée dans les rapports d'évaluation, alors que 96 rapports sur 127 (75,6 %) en ont fait l'usage. Ils ont été davantage utilisés auprès des intervenants externes/partenaires (43,3 %), des bénéficiaires (32,3 %) et de la clientèle-entreprise (26 %). Très peu de rapports ont comporté un sondage interrogeant les intervenants internes (7,9 %). Cependant, cette faible proportion peut se justifier étant

donné que de nombreux rapports évaluaient de petits programmes avec peu d'intervenants (par exemple, les programmes d'aide financière).

Pour **les entrevues**, 75 évaluations sur 127 (59,1 %) les ont utilisées afin d'interviewer au moins une partie prenante. Elles ont été davantage utilisées auprès des intervenants internes et externes (44,9 % et 29,9 % respectivement), mais beaucoup moins auprès des bénéficiaires et de la clientèle (15 % et 7,9 % respectivement).

De son côté, par rapport aux deux autres méthodes, **le groupe de discussion** a été relativement peu utilisé, soit dans seulement 18 rapports (14,2 %). Les principaux groupes formés par les évaluateurs étaient auprès d'intervenants internes et externes (6,3 % et 7,1 % respectivement).

Sinon, par rapport à la « clientèle – population », on peut remarquer que peu importe la méthode de collecte de données, elle a été peu consultée de manière générale (15,7% des rapports pour les sondages).

Finalement, si l'on fait exception de la revue documentaire et l'analyse des données administratives (exclues de ce tableau), très peu d'autres méthodes de collecte de données ont été utilisées dans les rapports d'évaluation. Par exemple, pour les observations directes, seulement 2 évaluations sur les 127 (1,6 %) en ont fait l'usage.

Tableau 5 : Méthodes de collecte et sources utilisées dans les rapports d'évaluation			
Méthodes et sources	Sondage(s)	Entrevue(s)	Groupe(s) de discussion
Total — Méthode de collecte	96 75,6 %	75 59,1 %	18 14,2 %
Intervenants internes	10 7,9 %	57 44,9 %	8 6,3 %

Intervenants externes/partenaires	55 43,3 %	38 29,9 %	9 7,1 %
Bénéficiaires	41 32,3 %	19 15,0 %	2 1,6 %
Clientèle — entreprises/organismes	33 26,0 %	1 0,8 %	0 0,0 %
Clientèle — population	20 15,7 %	9 7,1 %	4 3,1 %

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.4 Méthodologie utilisée pour traiter les questions normalisées de l’*Outil*

Cette sous-section vise à approfondir la réflexion en y répertoriant les méthodes de collecte de données et les sources d’information utilisées par les MO pour répondre à chacune des questions. Elle vise aussi à examiner le niveau de triangulation des méthodes selon la question. Pour y parvenir, nous avons procédé de la manière suivante. Premièrement, nous avons présenté les résultats des méthodes et des sources d’information utilisées pour chacun des trois enjeux de pertinence, d’efficacité et d’efficience. Cela inclut les parties prenantes consultées en lien avec les trois méthodes abordées précédemment, mais on inclut aussi, cette fois-ci, la revue documentaire et l’analyse de données administratives. Ensuite, nous avons analysé la triangulation des méthodes pour chacune des questions. Pour l’ensemble des résultats, seules les questions ayant été traitées dans au moins une dizaine de rapports ont été abordées.

4.4.1 Les méthodes et sources d’information utilisées pour les questions de pertinence

Le tableau 6 présente les méthodes et sources d’information utilisées pour répondre aux quatre questions normalisées liées à l’enjeu de la pertinence. Plusieurs constats pour l’enjeu de la pertinence peuvent en ressortir.

Pour répondre à la question de l’analyse des besoins et du contexte (Q1), la revue documentaire et l’analyse des données administratives ont été les principales méthodes

employées (50 % des rapports traitant la question pour chacune), suivi des sondages (42,5 %) et des entrevues (35 %). Diverses publications ont été utilisées pour documenter la question :

- cadre légal et réglementaire
- documentation du programme (stratégies, politiques, plans d'action)
- revue de littérature,
- études, mémoires et statistiques contextuelles.

Concernant les sources d'information, la question a été principalement traitée en sondant les bénéficiaires du programme (33,8 %) ou en interviewant les intervenants internes (22,5 %) et externes (12,5 %). Cependant, la clientèle a été sondée dans seulement 6 rapports (7,5 %) et interviewée dans seulement 3 rapports (3,8 %). Cela peut être surprenant dans la mesure où une partie de la question concerne les besoins des populations cibles du programme, qui sont souvent les clientèles.

Pour la question traitant la cohérence avec le MO et les orientations gouvernementales (Q2), la quasi-totalité des rapports (98,4 %) a utilisé la revue documentaire comme principale source d'information. Cela s'explique par la nature de la question, qui exige de comparer le programme avec les orientations gouvernementales ainsi qu'avec la mission, les mandats et les responsabilités avec le MO. Divers documents ont été utilisés, notamment :

- des lois et plans stratégiques;
- des plans d'action ou plans d'intervention;
- des stratégies et des politiques;
- des discours sur le budget et des documents d'étude des crédits;
- d'autres documents de l'organisation responsable de la gestion du programme.

Cependant, très peu d'autres méthodes et sources d'information ont été utilisées. La seule autre combinaison de méthodes et de sources principalement utilisée l'a été dans une petite proportion des rapports (seulement 14,3 %), soit les entrevues auprès des intervenants internes. Quelle que soit la méthode de collecte, la clientèle n'a jamais été interrogée pour

répondre à cette question (ce qui peut s'expliquer par le fait qu'elle n'est pas la partie prenante étant le plus en mesure d'apprécier la cohérence du programme avec les orientations gouvernementales).

Pour la question traitant l'analyse des chevauchements (Q3), c'est la revue documentaire qui a été principalement utilisée pour y répondre (78 %). Cela s'explique aussi par la nature de la question, qui vise à examiner les documents de programmes similaires afin de s'assurer qu'il n'y ait pas de duplication avec le programme évalué. Souvent, une comparaison était effectuée à l'aide des documents du programme évalué et les cadres normatifs d'autres programmes. Plusieurs MO ont aussi interviewé les intervenants internes (25,4 % des rapports traitant la question), souvent pour venir confirmer les constats établis par la comparaison des programmes issue de la revue documentaire ou tout simplement pour connaître leurs points de vue sur le sujet. Les intervenants externes, les bénéficiaires et les entreprises clientes ont aussi été interrogés à l'aide d'un sondage à de nombreuses reprises (25,4 % ; 11,9 % et 23,7 % respectivement).

Pour la question traitant la pertinence de la nature de l'intervention (Q4), le sondage a été la principale méthode de collecte de données utilisée (83,1 %), surtout auprès des intervenants externes (33,8%), des bénéficiaires (44,6%) et des entreprises/organismes client(e)s (32,3%). Plusieurs MO ont aussi interviewé les intervenants internes afin de répondre à la question (24,6 %). Aussi, les données administratives ont été souvent utilisées (46,2 %) et la revue documentaire l'a été modérément (26,2 %). La faible utilisation d'une revue documentaire pour analyser la pertinence de la nature de l'intervention semble relativement surprenante puisqu'elle inclut l'étalonnage. En effet, la plupart des sources de renseignements possibles mentionnées dans l'*Outil* en lien avec cette question sont liées à la revue documentaire, telle que la documentation colligée sur les moyens et les mécanismes du programme, des études et autres productions scientifiques, des documents de veille et des évaluations, audits ou examens ayant déjà analysé le programme en question ou un programme similaire (Secrétariat du Conseil du trésor, 2018b, p. 18).

Tableau 6 : Méthodes de collecte et sources utilisées par question normalisée — Enjeu de la pertinence								
Méthodes et sources	Q1 — Besoins et contexte		Q2 — Cohérence MO		Q3 — Chevauchements		Q4 — Pertinence de la nature	
Revue documentaire	40	50,0 %	62	98,4 %	46	78,0 %	17	26,2 %
Données administratives	40	50,0 %	3	4,8 %	8	13,6 %	30	46,2 %
Sondage(s)	34	42,5 %	4	6,3 %	25	42,4 %	54	83,1 %
Intervenants internes	2	2,5 %	0	0,0 %	1	1,7 %	2	3,1 %
Intervenants externes/partenaires	5	6,3 %	2	3,2 %	15	25,4 %	22	33,8 %
Bénéficiaires	27	33,8 %	3	4,8 %	7	11,9 %	29	44,6 %
Clientèle — entreprises/organismes	2	2,5 %	0	0,0 %	14	23,7 %	21	32,3 %
Clientèle — population	4	5,0 %	0	0,0 %	0	0,0 %	4	6,2 %
Entrevue(s)	28	35,0 %	10	15,9 %	18	30,5 %	25	38,5 %
Intervenants internes	18	22,5 %	9	14,3 %	15	25,4 %	16	24,6 %
Intervenants externes/partenaires	10	12,5 %	3	4,8 %	6	10,2 %	5	7,7 %
Bénéficiaires	3	3,8 %	0	0,0 %	3	5,1 %	3	4,6 %
Clientèle — entreprises/organismes	1	1,3 %	0	0,0 %	0	0,0 %	1	1,5 %
Clientèle — population	2	2,5 %	0	0,0 %	0	0,0 %	0	0,0 %
Groupe(s) de discussion	5	6,3 %	1	1,6 %	2	3,4 %	4	6,2 %
Total	80		63		59		65	

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.4.2 Les méthodes et sources d'information utilisées pour les questions d'efficacité

Le tableau 7 présente les méthodes et les sources utilisées pour les questions liées à l'enjeu de l'efficacité.

Pour répondre à la question traitant l'efficacité opérationnelle (Q9), les données administratives ont été particulièrement exploitées pour cette question (87,7 % des rapports traitant l'efficacité opérationnelle). Cela s'explique en grande partie par la nécessité de mesurer les résultats opérationnels obtenus ou les extrants. Les MO ont aussi parfois utilisé des entrevues auprès des intervenants internes et externes (26,4 % et 22,6 % des rapports respectivement). Ces cas sont souvent liés à des éléments de réponse qui concernent des questions sur la mise en œuvre du programme, tels que les facteurs pouvant contribuer à faciliter ou entraver la réalisation des activités et des extrants. Finalement, les sondages ont aussi été parfois utilisés pour interroger les bénéficiaires et la « clientèle — population »

(26,4 % et 13,2 % respectivement). Ces cas sont souvent liés à des sondages sur la satisfaction par rapport à la qualité des extraits.

Pour les questions traitant l'atteinte des objectifs et des effets visés (**Q10/Q11**), la plupart des rapports les traitant (75,8 %) ont utilisé les données administratives et une partie importante (41,1 %) a employé la revue documentaire pour répondre à celle-ci. Cette dernière est souvent utilisée en accompagnement de l'interprétation des données administratives. Par exemple, souvent, des liens étaient faits entre l'atteinte des objectifs et le cadre normatif du programme, divers documents internes, des politiques, des statistiques de l'ISQ, des plans stratégiques, etc. Aussi, le sondage a été particulièrement utilisé (75,8 %) pour répondre à la question, et ce auprès de l'ensemble des parties prenantes, à l'exception des intervenants internes (seulement 6,3 %).

Pour répondre à la question traitant les effets non visés (**Q12**), les MO ont principalement employé des entrevues avec les intervenants internes (63,6 %), les intervenants externes (36,4 %) et les bénéficiaires (27,3 %). Très peu de rapports traitant la question ont utilisé la revue documentaire et les données administratives (9,1 % pour chacun). Ces résultats ne sont pas très surprenants dans la mesure où l'objectif est de repérer des effets non prévus dans le programme et que la mesure de ceux-ci n'est pas toujours évidente à prévoir.

Tableau 7 : Méthodes de collecte et sources utilisées par question normalisée						
— Enjeu de l'efficacité						
Méthodes et sources	Q9 — Efficacité opérationnelle		Q10/Q11 – Objectifs/effets visés		Q12 — Effets non intentionnels	
Revue documentaire	30	28,3 %	39	41,1 %	1	9,1 %
Données administratives	93	87,7 %	72	75,8 %	1	9,1 %

Sondage(s)	52	49,1 %	72	75,8 %	6	54,5 %
Intervenants internes	4	3,8 %	6	6,3 %	1	9,1 %
Intervenants externes/partenaires	13	12,3 %	30	31,6 %	0	0,0 %
Bénéficiaires	28	26,4 %	37	38,9 %	3	27,3 %
Clientèle — entreprises/organismes	2	1,9 %	24	25,3 %	0	0,0 %
Clientèle — population	14	13,2 %	18	18,9 %	2	18,2 %
Entrevue(s)	41	38,7 %	38	40,0 %	9	81,8 %
Intervenants internes	28	26,4 %	22	23,2 %	7	63,6 %
Intervenants externes/partenaires	24	22,6 %	17	17,9 %	4	36,4 %
Bénéficiaires	9	8,5 %	10	10,5 %	0	0,0 %
Clientèle — entreprises/organismes	1	0,9 %	1	1,1 %	0	0,0 %
Clientèle — population	6	5,7 %	3	3,2 %	0	0,0 %
Groupe(s) de discussion	10	9,4 %	6	6,3 %	1	9,1 %
Total		106		95		11

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.4.3 Les méthodes et sources d'information utilisées pour les questions d'efficience

Le tableau 8 présente les données pour les questions d'efficience autres que celles présentes dans l'*Outil*. La quasi-totalité des rapports traitant une autre question d'efficience (96,6 %) a exploité les données administratives. Il s'agit de la principale méthode utilisée pour répondre à ce type de questions. Souvent, les données administratives étaient utilisées afin de réaliser une analyse coûts-avantages, par exemple. Très peu d'autres méthodes de collecte de données et sources d'information ont été utilisées pour répondre à ces questions d'efficience.

Tableau 8 : Méthodes de collecte et sources utilisées par question normalisée — Enjeu de l'efficience		
Méthodes et sources	Autres questions — Efficience	
Revue documentaire	4	13,8 %
Données administratives	28	96,6 %
Sondage(s)	4	13,8 %

Intervenants internes	0	0,0 %
Intervenants externes/partenaires	1	3,4 %
Bénéficiaires	1	3,4 %
Clientèle — entreprises/organismes	2	6,9 %
Clientèle — population	0	0,0 %
Entrevue(s)	3	10,3 %
Intervenants internes	2	6,9 %
Intervenants externes/partenaires	1	3,4 %
Bénéficiaires	0	0,0 %
Clientèle — entreprises/organismes	0	0,0 %
Clientèle — population	1	3,4 %
Groupe(s) de discussion	1	3,4 %
Total	29	

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.4.4 Triangulation des méthodes pour répondre aux questions de l'*Outil*

À quel point y a-t-il eu une triangulation des méthodes dans les rapports d'évaluation pour répondre à chacune des questions ? Le tableau 9 présente les pourcentages de rapports ayant triangulé les méthodes ou non pour répondre aux principales questions de l'*Outil*. Il est considéré qu'il y a eu triangulation des méthodes lorsque plus d'une méthode parmi les suivantes a été utilisée pour répondre à la question : la revue documentaire, l'analyse des données administratives, le sondage, l'entrevue, le groupe de discussion et l'observation. Il faut tenir compte du fait que les sources d'information n'ont pas été prises en compte dans la comptabilisation des méthodes et qu'il s'agit d'une limite dans l'interprétation des données.

Nous pouvons constater que la triangulation des méthodes est relativement importante pour la plupart des questions, à l'exception de celle de la cohérence avec le MO et les orientations gouvernementales (Q2) et des autres questions d'efficacité. Les questions de l'atteinte des objectifs et des effets visés (Q10/Q11) sont celles dont la proportion d'utilisation de plus d'une méthode est la plus forte (86,3 %), suivies de la pertinence de la nature de l'intervention (Q4) (64,6 %) et de l'efficacité opérationnelle (Q9) (63,2 %). Malgré tout, il demeure qu'une partie importante des rapports ne triangule pas les méthodes si l'on exclut

l'atteinte des objectifs et des effets visés (Q10/Q11) (plus d'un tiers pour l'ensemble des autres questions).

Tableau 9 : Taux de triangulation des méthodes selon la question normalisée				
	Questions	Pas de triangulation	Triangulation (plus d'une méthode)	Rapports
Pertinence	Q1 — Besoins et contexte	46,3 %	53,8 %	80
	Q2 — Cohérence MO	76,2 %	23,8 %	63
	Q3 — Chevauchements	44,1 %	55,9 %	59
	Q4 — Pertinence de la nature	35,4 %	64,6 %	65
Efficacité	Q9 — Efficacité opérationnelle	36,8 %	63,2 %	106
	Q10/Q11 — Objectifs/effets visés	13,7 %	86,3 %	95
	Q12 — Effets non intentionnels	45,5 %	54,5 %	11
Efficience	Autres questions — Efficience	75,9 %	24,1 %	29

Source : Compilation effectuée par la DGEAIGR, avril 2021.

4.5 Principaux constats du bilan

Le tableau 10 reprend les questions qui avaient été définies dans la description du bilan et énumère les principaux constats pour chacun des volets.

Tableau 10 : Questions spécifiques du bilan et principaux constats établis

Volets	Questions	Principaux constats
1. Portrait général des évaluations	1.1 Quels sont les principaux éléments généraux et contextuels qui caractérisent les rapports d'évaluation déposés par les MO en 2019 et 2020 ?	<p>Constat 1 : La plupart des rapports d'évaluation déposés en 2019 et 2020 ont été réalisés par seulement quelques ministères.</p> <p>Constat 2 : Une très grande majorité des évaluations ont été réalisées en s'appuyant sur un comité d'évaluation. Cependant, seulement 33 évaluations (26 %) disposaient d'un comité d'évaluation avec au moins un membre provenant de l'externe et non bénéficiaire (partenaires, représentants des usagers, autres MO impliqués, etc.).</p> <p>Constat 3 : Les MO délèguent rarement l'exécution de leurs évaluations à des entités externes (seulement 7 rapports sur 127).</p>

		<p>Constat 4 : La majorité des rapports d'évaluation déposés répondaient d'abord et avant tout à des besoins internes aux MO (55,9 %).</p>
<p>2. Traitement des enjeux de la pertinence, de l'efficacité et de l'efficience</p>	<p>2.1 Combien de rapports ont abordé, d'une manière ou d'une autre, chacun des enjeux de la pertinence, de l'efficacité et de l'efficience ?</p> <p>2.2 Les questions normalisées liées aux enjeux de pertinence, d'efficacité et d'efficience de l'<i>Outil d'évaluation des programmes</i> ont-elles été traitées dans les rapports d'évaluation des MO ?</p>	<p>Constat 5 : Les enjeux de pertinence et d'efficacité ont été bien couverts, mais celui de l'efficience l'a été beaucoup moins.</p> <p>Constat 6 : Les quatre questions normalisées qui portent sur l'efficience relative de l'intervention sont très peu reprises par les MO.</p> <p>Constat 7 : La question concernant les effets non intentionnels (Q12) est abordée dans une minorité des rapports (9%).</p>
<p>3. Portrait des méthodes et sources utilisées dans les rapports</p>	<p>3.1 Quelles méthodes de collecte de données et sources d'information les MO ont-ils utilisées dans leurs rapports d'évaluation ?</p>	<p>Constat 8 : À l'exception de la revue documentaire et de l'analyse des données administratives, les sondages ont été la principale méthode utilisée dans les évaluations de 2019 et 2020 (75,6 % des rapports), suivi des entrevues (59,1 %).</p> <p>Constat 9 : Les groupes de discussion ont été relativement peu utilisés comparativement aux sondages et aux entrevues (18 rapports seulement soit 14,2 %).</p> <p>Constat 10 : Très peu de méthodes de collecte de données autres que celles abordées dans le bilan ont été utilisées dans les rapports d'évaluation. Par exemple, pour les observations directes, seulement 2 évaluations sur les 127 (1,6 %) en ont fait l'usage.</p> <p>Constat 11 : Les sondages ont été principalement utilisés pour interroger les intervenants externes, les bénéficiaires et la clientèle-entreprise, mais très peu pour les intervenants internes (7,9 % des rapports).</p> <p>Constat 12 : Les entrevues ont été principalement utilisées pour interroger les intervenants internes et externes, mais peu auprès des bénéficiaires et de la clientèle (15 % et 7,9 % des rapports respectivement).</p> <p>Constat 13 : Peu importe la méthode de collecte de données, la « clientèle population » a été peu consultée (15,7 % de sondage, 7,1 % d'entrevues et 3,1 % de groupes de discussion).</p>
<p>4. Méthodologie employée pour les questions normalisées</p>	<p>4.1 Quelles ont été les méthodes et sources utilisées pour chacune des questions normalisées traitant la pertinence, l'efficacité et l'efficience ?</p> <p>4.2 Y a-t-il eu une triangulation des méthodes pour chacune des questions de l'Outil ?</p>	<p>Constat 14 : La question de l'analyse des besoins et du contexte (Q1) a été principalement documentée par les données administratives et la revue documentaire (50 %), ainsi qu'en sondant les bénéficiaires (33,8 %). Cependant, la clientèle a été très peu consultée, et ce même si elle est souvent la population cible des programmes.</p> <p>Constat 15 : La quasi-totalité des rapports traitant la question de la cohérence avec le MO et les orientations gouvernementales (Q2) ont utilisé la revue documentaire pour y répondre (98,4 %). Très peu de rapports ont utilisé une autre combinaison de méthode et de source.</p> <p>Constat 16 : La question de l'analyse des chevauchements (Q3) a été principalement traitée à l'aide d'une revue documentaire (78 %) visant spécifiquement la documentation de programmes similaires.</p> <p>Constat 17 : La question de la pertinence de la nature de l'intervention (Q4) a été répondue de différentes manières, tant avec les données administratives (46,2 %) qu'avec des sondages auprès des intervenants externes (33,8 %), des bénéficiaires (44,6 %) et la clientèle entreprise (32,3 %). Seulement 26,2 % des rapports ont utilisé la revue documentaire (incluant l'étalonnage).</p> <p>Constat 18 : Les questions d'efficacité opérationnelle (Q9) et de l'atteinte des objectifs et des effets visés (Q10/Q11) ont été principalement traitées à l'aide de</p>

		<p>données administratives (87,7 % et 75,8 % respectivement) et aussi par l'entremise de sondages auprès de différentes parties prenantes et des entrevues auprès des intervenants internes et externes. Les citoyens clients ont été sondés dans une quantité non négligeable de rapports pour répondre à ces questions (13,2 % et 18,9 % respectivement).</p> <p>Constat 19 : Pour répondre à la question traitant les effets non intentionnels (Q12), les MO ont principalement employé des entrevues avec les intervenants internes (63,6 %), les intervenants externes (36,4 %) et des sondages auprès des bénéficiaires (27,3 %) et des citoyens clients (18,2 %). Très peu de rapports traitant la question ont utilisé la revue documentaire et les données administratives (9,1 % pour chacun).</p> <p>Constat 20 : La triangulation des méthodes (au moins deux méthodes) est relativement importante pour la plupart des questions, à l'exception de Q2 et des autres questions d'efficience (entre 52,5 % et 85,3 % des rapports ayant traité la question concernée). Cependant, une partie non négligeable des rapports ne triangulent pas les méthodes pour répondre à plusieurs questions.</p>
--	--	---

5. RECENSION DES ÉCRITS POST-STAGE

Après avoir présenté les résultats du bilan, nous avons cru opportun de réaliser une recension des écrits afin de mieux interpréter les principaux constats établis. Pour ce faire, nous avons sélectionné, parmi les 20 constats, certaines thématiques qui nous semblaient plus importantes que d'autres. Le choix fut effectué dans le but d'approfondir la réflexion sur les résultats du bilan et, éventuellement, d'émettre des recommandations ancrées dans la littérature. Nous avons donc retenu trois thématiques : le traitement des enjeux de pertinence, d'efficacité et d'efficience ; l'utilisation des groupes de discussion en évaluation et la triangulation des méthodes en évaluation. Pour chacune de ces thématiques, de brèves recensions non exhaustives ont été réalisées et seront présentées dans chacune des sous-sections les concernant.

5.1 Le traitement des enjeux de pertinence, d'efficacité et d'efficience en évaluation

Nous avons préféré détailler davantage cette sous-section. Elle sera divisée en cinq parties : l'enjeu de la pertinence, la question de l'efficacité opérationnelle (Q9), celle des effets de l'intervention (Q10/Q11), celle des effets non intentionnels (Q12) ainsi que l'enjeu de l'efficience. Pour chacune des parties, nous avons opté pour la présentation d'un ou deux articles, chapitres de livre ou mémoires qui concernent l'enjeu ou la question.

Pour l'enjeu de pertinence, nous nous sommes principalement intéressés au chapitre de Champagne et al. (2011), qui aborde la question de l'utilisation de l'analyse stratégique (AS) en évaluation dont l'objectif est de déterminer la pertinence d'une intervention ou d'un programme. Bien que l'ouvrage cible spécifiquement l'évaluation de la pertinence dans le secteur de la santé, il peut très bien s'appliquer aussi aux autres secteurs. Les auteurs spécifient que la « littérature met peu l'accent sur la pertinence comme critère d'évaluation (...) (malgré le fait qu'elle peut s'avérer très importante) et même cruciale dans la prise de

décision » (Ibid, p. 108). Les auteurs proposent donc un cadre comportant trois principales étapes : l'analyse de la pertinence du problème visé (AS 1), l'analyse de la pertinence des objectifs de l'intervention (AS 2) et l'analyse de la pertinence du partenariat stratégique (AS 3) (Ibid, pp. 105-106). La première, visant notamment à identifier les problèmes et analyser leur priorité relative, concerne directement la question 1 de l'*Outil* qui traite l'analyse des besoins et du contexte. Les auteurs affirment que l'identification des problèmes peut se faire à l'aide de trois principales approches, soit par l'utilisation des données existantes, soit à l'aide d'une enquête provenant d'une collecte de données, soit par l'utilisation des techniques qualitatives de recherche de consensus (entrevues auprès des informateurs clés ou d'autres techniques de consultation telles que des groupes de réflexion ou la technique de Delphi) (Ibid, pp. 109-110). Cependant, les auteurs déconseillent de se fier uniquement à des données sur l'utilisation des services, qui reflètent uniquement les besoins qui ont déjà été satisfaits et qui ne fournissent aucune information sur les besoins des personnes qui n'utilisent pas les services (Ibid, p.109). Or, nous avons vu dans le bilan que les MO ont beaucoup utilisé les données administratives (50 % des rapports) et les sondages auprès des bénéficiaires (33,8 % des rapports) pour traiter la question de l'analyse des besoins et du contexte (Q1). Cependant, nous n'avons pas établi de distinction entre les bénéficiaires du programme et les bénéficiaires potentiels, ce qui aurait permis de mieux déterminer la qualité des sources utilisées pour cette question. Concernant l'AS 2, qui concerne à la fois la pertinence du choix des causes visées et celle du choix de la population ciblée, elle vise notamment à « construire le modèle causal du problème visé (...) (et à) comparer la cause visée avec toutes les autres causes à l'origine du problème à résoudre » (Ibid, p. 111). Malgré les atouts considérables que peut apporter cette seconde étape, il est très rare que les MO aient approfondi la réflexion sur les besoins et le contexte à ce niveau pour déterminer si le programme est toujours pertinent ou non. Concernant l'analyse de la pertinence du partenariat stratégique (AS 3), aucune question de l'*Outil* ne semble la concerner directement.

Concernant la question concernant l'atteinte des objectifs opérationnels (Q9), la large compréhension que nous avons eue, impliquant l'inclusion de questions apparentées, nous

incite à nous intéresser à un concept plus englobant et qui est directement associé à celle-ci, soit l'analyse de la production. Celle-ci consiste à une étude des « relations entre le volume et la qualité des services produits et les ressources utilisées pour leur production (et) (...) comporte deux volets : l'analyse de la productivité et l'analyse de la qualité » (Farand, 2011, p. 127). L'analyse de production s'intéresse notamment aux caractéristiques des processus de production (besoins, ressources, services) et à sa productivité (totale, du capital, technique, économique). Elle vise aussi à étudier divers éléments liés à la qualité globale du processus tels que les différents types d'accessibilité (temporelle, géographique, financière, culturelle et informationnelle), l'intensité du processus, sa pertinence, sa globalité, sa continuité et la qualité des services. Enfin, elle a aussi pour objectif d'analyser les divers éléments d'influence pouvant affecter le processus de production (Ibid, p. 136). Par l'entremise d'une étude de cas en lien avec le système de santé, Farand propose diverses méthodes de collecte de données pour analyser différentes composantes liées à l'analyse de la production et, ce faisant, qui pourraient être pertinentes dans le cadre de la mesure de l'efficacité opérationnelle. Notamment, pour la description des ressources (ou intrants), il mentionne qu'on utilise en général des données administratives, mais que les questionnaires, l'observation directe et des entrevues semi-dirigées peuvent aussi y contribuer (Ibid, pp.139-140). Pour l'analyse d'accessibilité, les méthodes et sources d'information qu'il propose varient en fonction du type d'accessibilité : notamment, les données administratives pour les accessibilités temporelle et géographique (délais d'attente, distances), les questionnaires aux clients pour les accessibilités financière et culturelle et auprès des intervenants pour l'accessibilité informationnelle (Ibid, p. 153). Ces diverses méthodes et sources d'information se reflètent dans les résultats de notre bilan par rapport à la question de l'efficacité opérationnelle, qui ont montré que les MO pouvaient utiliser autant les sondages et les entrevues auprès de différentes parties prenantes pour la traiter, malgré le fait que l'analyse des données administratives était la plus utilisée. Cependant, les données générées par le bilan sont relativement limitées quant à leur portée sur la manière dont les MO ont traité la question.

Pour l'analyse des effets (en lien avec la question 11 de l'*Outil*), nous nous sommes limités à présenter les méthodes et sources utilisées, mais généralement, des devis sont spécifiquement élaborés pour répondre à cet enjeu. En effet, l'une des problématiques souvent rencontrées dans la mesure des effets est la question de l'attribution (ou la causalité), qui permet de lier une variable dépendante à une variable indépendante, mais qui nécessite l'élaboration d'un devis expérimental ou quasi expérimental permettant de faire des inférences causales (Haccoun et McDuff, 2018, p. 111). En évaluation de programmes, six principaux devis (trois étant non expérimentaux et trois étant soit expérimentaux ou quasi expérimentaux) sont généralement utilisés pour y attribuer des effets : le devis pré-post à groupe unique, la stratégie par référence interne, la série chronologique à groupe unique, le devis post avec comparaison indépendante, le devis pré-post avec comparaison indépendante ainsi que la série chronologique avec groupe comparaison (Ibid, pp. 112-113). Lors du prochain bilan, s'il y a bien lieu, il serait intéressant d'identifier quels types de devis ont été utilisés à partir de cette typologie (ou d'une autre semblable) pour mesurer les effets du programme.

Quant à la question des effets non intentionnels (Q12), Champagne et al. (2011a) mentionnent l'importance de les prendre en compte étant donné qu'il arrive « qu'une intervention provoque des effets pervers qui en neutralisent largement les effets positifs (et que) dans ces cas-là, le fait de mesurer uniquement les effets positifs ne permet pas d'obtenir une image valide des effets » (p.174). Pourtant, le faible traitement constaté de la question pourrait alimenter le débat quant aux facteurs autres que celui du caractère obligatoire qui peuvent expliquer un tel résultat. Jabeen (2016), en réalisant une recension des écrits sur la question, a repéré trois raisons majeures contribuant à la non-prise en compte des effets non prévus en évaluation du développement et qui pourraient s'appliquer aussi dans ce cas-ci. Les trois causes identifiées sont le manque de compréhension du concept, la nature politique de l'évaluation (la connotation négative qu'il peut avoir) et les approches méthodologiques abordant les effets non intentionnels étant très limitées à l'heure actuelle (pp. 152-153). Sur ce dernier point, le même auteur, ayant réalisé sa thèse de doctorat sur le sujet, a développé

deux ans plus tard une approche évaluative des effets imprévus pour les programmes à dimension sociale dont pourrait s'inspirer la DGEAIGR dans le cadre de la réalisation d'un guide ou d'une fiche (Jabeen, 2018). De son côté, Champagne et coll. (2011a) affirment qu'il est nécessaire de tenir compte de l'expérience des parties prenantes au programme et de ne pas se limiter à une simple revue de littérature spécialisée, car « les effets nuls ou indésirables sont sous-représentés en raison d'un biais de publication » (p. 174). Cela se reflète dans les résultats obtenus dans le bilan, alors que nous avons vu que les intervenants internes et partenaires ont été particulièrement interviewés pour traiter cette question.

Enfin, concernant les questions d'efficacité, Leblanc (2019) a réalisé son mémoire spécifiquement sur le traitement de cet enjeu par les MO québécois depuis la tenue de la Commission de révision permanente des programmes publics. Par l'entremise notamment d'une analyse de contenu de 54 rapports du ministère de l'Éducation et du ministère de l'Économie et de l'Innovation, d'un questionnaire auprès de 13 évaluateurs travaillant dans un MO québécois et 4 entretiens semi-dirigés, il est arrivé au constat que « l'efficacité n'est pas systématiquement évaluée dans les évaluations de programme au gouvernement du Québec (et que) lorsque celle-ci l'est, l'évaluation économique est peu utilisée » (Leblanc, 2019, p. 78). En effet, dans les rapports analysés, l'auteur a pu constater que les évaluateurs utilisaient d'autres indicateurs pour mesurer l'efficacité tels que le coût de revient, l'effet de levier, le ratio des charges nettes sur les ventes et les revenus générés sur le nombre d'équivalents à temps complet (Ibid, pp.57-58). Il a aussi confirmé certaines hypothèses mentionnées dans la littérature quant aux principaux obstacles à l'utilisation des évaluations économiques, notamment la difficulté de convertir en dollars les coûts et les bénéfices ou de pondérer le niveau d'utilité, le manque de connaissances ou de compétences en la matière ainsi que le manque de ressources financières (Ibid, pp. 55-56). La DGEAIGR, si elle a pour objectif d'accroître le taux de couverture de l'enjeu de l'efficacité, devrait donc cibler principalement ces trois obstacles.

5.2 L'utilisation des groupes de discussion en évaluation

Nous avons vu dans le bilan que les groupes de discussion étaient beaucoup moins utilisés que les entrevues et les sondages dans les évaluations réalisées par les MO. Pourtant, cette méthode de collecte de données dispose de plusieurs atouts qui pourraient être pris davantage en compte par les évaluateurs. Notamment, l'un de ses principaux avantages est que les interactions entre les différents membres du groupe peuvent produire de l'information et des données qui ne seraient pas autant accessibles d'une autre manière (Morgan (1988), cité dans (Race et al., 1994, p. 731)). La méthode peut être utilisée à différentes fins, notamment pour obtenir des informations générales sur un programme, générer des idées qui pourront être testées ultérieurement, diagnostiquer des problèmes liés à l'intervention et collecter des données sur les perceptions de la clientèle sur le programme (Stewart et Shamdasani [1990], cité dans [Race et al., 1994, p. 731]). Aussi, elle peut être utilisée dans plusieurs types d'évaluations : évaluation des besoins, développement de la théorie du programme, évaluation d'implantation et évaluation des résultats (Ryan et al., 2014, p. 328). Si la DGEAIGR voulait appuyer les évaluateurs des MO afin qu'ils utilisent davantage le groupe de discussion, elle pourrait agir là aussi sur les principaux obstacles. Notamment, elle pourrait considérer que d'une manière générale, « la conduite d'un groupe de discussion requiert un temps de préparation plus long qu'une entrevue individuelle et nécessite des habiletés de communication pour faciliter la discussion de groupe » (Fortin et Gagnon, 2016, p. 323). Aussi, au-delà de l'objectif d'une augmentation du taux d'utilisation de la méthode, elle pourrait jouer un rôle quant à la maximisation de l'utilité de celle-ci dans les évaluations en s'inspirant notamment des bonnes pratiques qui sont déjà en œuvre dans certains ministères ou en organisant des formations sous forme de partenariats avec des évaluateurs expérimentés des MO concernés. Des fiches pourraient être aussi élaborées en s'inspirant de la littérature en la matière, notamment les questions proposées par Ryan et al. (2014) visant notamment à renforcer la crédibilité des données collectées à partir des groupes de discussion (p.341).

5.3 La triangulation des méthodes en évaluation

Nous avons vu dans le bilan que la triangulation des méthodes de collecte de données n'était pas toujours au rendez-vous pour répondre aux différentes questions de l'*Outil*. Pourtant, les bonnes pratiques spécifient que la triangulation des méthodes est importante en évaluation de programme. Selon Pluye (2018), les trois principales raisons pour combiner des méthodes qualitatives et quantitatives sont que :

Les évaluateurs peuvent avoir besoin de méthodes qualitatives pour interpréter des résultats quantitatifs. Ils peuvent avoir besoin de méthodes quantitatives pour généraliser des résultats qualitatifs. Enfin, ils peuvent avoir besoin d'explorer ou de mieux comprendre un phénomène nouveau (méthodes qualitatives) et d'en mesurer l'ampleur, l'évolution, les causes et les effets (méthodes quantitatives) (p.125).

Autrement dit, chacune des méthodes qualitatives et quantitatives apporte des avantages et des inconvénients que ne comporte pas nécessairement l'autre. Les méthodes quantitatives, d'un côté, sont souvent reconnues comme étant utiles en évaluation de programme pour permettre des généralisations à l'aide de petits échantillons représentatifs, pour prouver la causalité des impacts d'un programme et pour assurer une bonne validité externe de l'évaluation (Rao et Woolcock, 2003). Cependant, elles ne permettent pas à elles seules de bien comprendre le processus et les mécanismes sous-jacents qui entourent la chaîne causale, ce que peuvent permettre les méthodes qualitatives. Néanmoins, ces dernières comportent aussi des limites, notamment le fait que l'échantillon est souvent petit (rendant difficile la généralisation), que la sélection de l'échantillon est souvent basée sur le jugement de l'évaluateur ou sur la recommandation d'autres participants (rendant difficile sa répliquabilité) et que l'analyse des données qualitatives est souvent réalisée à l'aide d'interprétations subjectives (Ibid). En bref, étant donné que les avantages de l'une viennent compenser les désavantages de l'autre, la combinaison de méthodes qualitatives et quantitatives devrait idéalement être utilisée dans les évaluations.

6. DISCUSSION ET RECOMMANDATIONS

Le bilan des pratiques évaluatives visait à examiner les pratiques, les méthodes et les sources utilisées dans les 127 rapports d'évaluation déposés en 2019 et 2020 par les MO québécois. Les objectifs étaient à la fois de décrire les caractéristiques et éléments contextuels des évaluations, de mesurer le traitement des enjeux et des questions de pertinence, d'efficacité et d'efficience, de répertorier les méthodes et sources utilisées globalement ainsi que par question normalisée. De nombreux constats ont pu être établis dans la section des résultats. Notamment, nous avons pu observer qu'une faible partie des évaluations s'appuyait sur un comité d'évaluation mixte, que les MO délèguent rarement leurs évaluations à une entité externe, que l'enjeu de l'efficience est beaucoup moins traité que ceux de la pertinence et de l'efficacité et que les effets non intentionnels ne sont pas souvent abordés. Aussi, nous avons vu que les groupes de discussion ont été moins utilisés que les autres méthodes, que la « clientèle – population » a été peu consultée pour l'ensemble des méthodes, que le sondage auprès des bénéficiaires a été beaucoup utilisé pour répondre à certaines questions de pertinence et d'efficacité et enfin, que la triangulation des méthodes n'est pas toujours au rendez-vous. De leur côté, les deux recensions des écrits réalisées avant et après le stage auront permis de contextualiser les résultats. La recension sur l'historique du processus d'institutionnalisation de l'évaluation de programme au Québec a pu contribuer à une meilleure compréhension du rôle du SCT, qui, désormais, ne se limite plus tout simplement à de la reddition de comptes, mais qui agit en tant qu'acteur de renforcement des capacités. La recension post-stage aura contribué à renforcer l'analyse des principaux constats établis du bilan des pratiques évaluatives. De manière globale, la réalisation du bilan aura illustré l'hétérogénéité des pratiques évaluatives dans les MO.

Pour la DGEAIGR, le bilan sera utile dans la mesure où les principaux constats seront présentés au Réseau des répondants en évaluation de programme (RPEP). De cette présentation découlera aussi la création de tables de réflexion composées d'évaluateurs expérimentés de différents MO et qui réfléchiront à des recommandations et des pistes de solution. Aussi, les données collectées seront utilisées afin d'établir des portraits par MO dans le but d'élaborer des constats personnalisés et qui seront des intrants à des rencontres

dans un but d'amélioration continue. De notre côté, la réalisation du bilan nous aura permis, notamment, de passer de la théorie, largement abordée durant nos cours à l'ENAP, à la réalité en plongeant directement dans une multitude de réels exercices évaluatifs par l'entremise des rapports.

Il est important aussi de tenir compte des limites concernant la portée du bilan réalisé. Le présent bilan est particulièrement axé sur les méthodes, mais pas sur l'utilité des évaluations, élément n'ayant pas été analysé dans le cadre de cet exercice. Ainsi, si l'on adoptait une approche évaluative axée sur l'utilisation, qui « est basée sur le principe selon lequel les évaluations devraient être jugées en fonction de leur utilité » (Patton et LaBossière, 2018, p. 145), on pourrait demeurer perplexe par rapport au degré d'importance des constats établis. Cependant, cette dimension n'était pas l'objet central de la réflexion de ce bilan, qui a été abordée plutôt dans un bilan des recommandations réalisé en parallèle par une autre stagiaire. Aussi, l'un des trois objectifs de la Directive concerne l'amélioration de la qualité des évaluations. Cependant, les résultats présentés dans le bilan, ne sont pas de très bons indicateurs de la qualité des évaluations, même s'ils peuvent fournir certains indices concernant celle-ci. Nous avons vu dans la recension post-stage que les méthodes et les sources d'information répertoriées pour répondre à chacune des questions ne permettaient pas toujours de fournir des données très approfondies sur la qualité du traitement de celles-ci. Certaines recommandations en ce sens ont même été suggérées dans le cas d'une mise à jour ultérieure du bilan, telle que la prise en compte des différents types de devis pour mesurer les effets de l'intervention. Aussi, même si nous avons mesuré le traitement effectif des enjeux et des questions, le bilan ne produit aucune donnée importante sur la qualité de la réponse apportée à l'exception du taux de triangulation.

Si le SCT voulait pousser la réflexion sur le sujet, il serait nécessaire de réaliser une méta-évaluation de la qualité des rapports d'évaluation, exercice similaire déjà réalisé au gouvernement fédéral (Secrétariat du Conseil du Trésor du Canada, 2004) (un gabarit de l'examen est disponible à partir de la page 49 du document). Cette méta-évaluation a été

considérée par Daigneault (2010) comme étant pertinente et d'un niveau de crédibilité acceptable malgré certaines lacunes importantes. L'auteur propose d'ailleurs aux « gardiens de la qualité » des administrations publiques qui voudraient répéter un tel exercice de tenir compte de ces quatre recommandations :

1. Bien définir la conception théorique de la qualité de l'évaluation retenue et en déduire des indicateurs de mesure. Appliquer ces critères de manière appropriée et flexible, c'est-à-dire en prenant en compte le type, l'objet, et le contexte d'évaluation. (...)
2. Mettre en œuvre une stratégie de maximisation de la validité et de la fiabilité de la mesure de la qualité des évaluations ; éviter le recours aux rapports d'évaluation comme source exclusive de données. (...)
3. Le devis doit être approprié aux objectifs de la méta-évaluation ; les analyses doivent être suffisamment approfondies pour supporter les conclusions de la méta-évaluation. (...)
4. L'implication des différentes parties prenantes durant l'ensemble de la démarche peut contribuer à améliorer la crédibilité et l'acceptabilité des conclusions de la méta-évaluation (Ibid, pp. 215-217).

Étant donné les importantes ressources que peut nécessiter la réalisation d'une méta-évaluation, il pourrait être pertinent d'envisager la refonte de la grille d'analyse et d'appréciation des livrables afin que les rétroactions réalisées par les conseillers en évaluation de programme puissent générer en même temps des données probantes sur la qualité des rapports. Des critères de nature plus objective (par exemple, sous forme de questions précises, comme c'est le cas dans la grille d'analyse critique du cadre d'évaluation proposée par Mediell et Dionne (2018, p. 102-109)) pourraient faciliter la production de données probantes. De cette manière, celles-ci pourraient être plus facilement partagées aux MO et ainsi, pourraient contribuer à alimenter la réflexion sur la qualité des travaux évaluatifs.

CONCLUSION

Le présent rapport avait pour objectif de présenter les tâches réalisées dans le cadre d'un stage réalisé à la DGEAIGR et dont le mandat était de réaliser un bilan des pratiques évaluatives. Les principaux objectifs du bilan étaient de dresser un portrait de la distribution et de caractéristiques générales des rapports d'évaluation, d'examiner le traitement des enjeux et des questions de l'*Outil* concernant la pertinence, l'efficacité et l'efficience ainsi que de recenser les méthodes et les sources utilisées dans les rapports et pour chacune des questions.

Pour ce faire, une analyse de contenu de 127 rapports d'évaluation a été effectuée. Des variables ont été sélectionnées, des ensembles catégoriels et des définitions opératoires pour certaines catégories ont été élaborés et des procédures de codage ont été précisées afin de renforcer la validité interne de l'étude. Plusieurs mesures d'atténuation des limites ont été aussi employées afin de réduire les risques de résultats erronés, malgré la persistance d'un risque résiduel pour le codage de certains rapports et de certaines variables.

Les principaux constats établis sont que les MO ne s'appuient pas toujours sur des comités d'évaluation mixtes, qu'ils délèguent rarement la réalisation des évaluations à une entité externe, qu'ils abordent beaucoup moins l'efficience par rapport aux enjeux de pertinence et d'efficacité, qu'ils utilisent beaucoup moins les groupes de discussion par rapport à d'autres méthodes de collecte de données et qu'ils ne triangulent pas toujours leurs méthodes pour répondre aux différentes questions de l'*Outil*.

De ces constats pourront découler des pistes de solution potentielles qui émaneront des discussions prévues dans les tables rondes composées en partie d'évaluateurs expérimentés provenant des MO. En attendant, le rapport ouvre la porte à la réalisation d'une méta-évaluation qui disposerait de données plus globales sur la qualité des rapports.

BIBLIOGRAPHIE

- Arbour, G. (2016). Chapitre 2 : L'organisation de la fonction d'évaluation dans les politiques ministérielles au Québec : un état des lieux. Dans *Praxis de l'évaluation et de la révision des programmes publics : Approches, compétences et défis* (p. 33-58). Presses de l'Université du Québec.
- Champagne, F., Brousselle, A., Contandriopoulos, A.-P. et Hartz, Z. (2011a). L'analyse des effets. Dans *L'évaluation : concepts et méthodes 2e édition* (p. 173-198). Les Presses de l'Université de Montréal.
- Champagne, F., Brousselle, A., Contandriopoulos, A.-P. et Hartz, Z. (2011b). L'analyse stratégique. Dans *L'évaluation : concepts et méthodes 2e édition* (p. 105-116). Les Presses de l'Université de Montréal.
- Daigneault, P.-M. (2010). L'examen de la qualité des évaluations fédérales : une méta-évaluation réussie ? *The Canadian Journal of Program Evaluation*, 34.
- Farand, L. (2011). L'analyse de la production. Dans *L'évaluation : concepts et méthodes 2e édition* (p. 127-172). Les Presses de l'Université de Montréal.
- Fortin, M.-F. et Gagnon, J. (2016). *Fondements et étapes du processus de recherche : méthodes quantitatives et qualitatives*. Chenelière éducation.
- Haccoun, R. R. et McDuff, P. (2018). Attribution et causalité des effets. Dans C. Dagenais et V. Ridde (dir.), *Approches et pratiques en évaluation de programmes* (p. 109-124). Presses de l'Université de Montréal.
- Jabeen, S. (2016). Do we really care about unintended outcomes? An analysis of evaluation theory and practice. *Evaluation and Program Planning*, 55, 144-154. <https://doi.org/10.1016/j.evalprogplan.2015.12.010>
- Jabeen, S. (2018). Unintended outcomes evaluation approach: A plausible way to evaluate unintended outcomes of social development programmes. *Evaluation and Program Planning*, 68, 262-274. <https://doi.org/10.1016/j.evalprogplan.2017.09.005>
- Jacob, J. L., Lamari, M., Poulin-Larivière, L. et Beaucage, J. (2016). Chapitre 3 : Évaluation et révision des dépenses publiques : portrait analytique de l'expérience québécoise. Dans *Praxis de l'évaluation et de la révision des programmes publics : Approches, compétences et défis* (p. 59-84). Presses de l'Université du Québec.

- Leblanc, L. (2019). *La mesure de l'efficience dans les évaluations de programme de la fonction publique québécoise depuis la tenue de la Commission de révision permanente des programmes publics* [mémoire, École nationale d'administration publique]. Repéré à <http://espace.enaq.ca/id/eprint/188/1/032277420.pdf>
- Leblanc-Constant, M. et Bard, C. (2013). *Glossaire des termes usuels en mesure de performance et en évaluation : pour une gestion saine et performante*. https://www.tresor.gouv.qc.ca/fileadmin/PDF/publications/glossaire_termes_usuels.pdf
- Leray, C. et Bourgeois, I. (2016). L'analyse de contenu. Dans *Recherche sociale : De la problématique à la collecte des données* (6e édition, p. 427-453). PUQ.
- Loi sur l'administration publique. RLRQ c., A-6.01. <http://legisquebec.gouv.qc.ca/fr/ShowDoc/cs/A-6.01>
- Marceau, R., Simard, P. et Otis, D. (1992). La gestion de l'évaluation de programme au gouvernement du Québec. Dans *Management public : Comprendre et gérer les institutions de l'État*. Les Presses de l'Université du Québec.
- Mediell, M. et Dionne, E. (2018). Un outil à visée pédagogique pour discuter de méthodologie. *Canadian Journal of Program Evaluation*, 33(1). <https://doi.org/10.3138/cjpe.31127>
- Patton, M. Q. et LaBossière, F. (2018). L'évaluation axée sur l'utilisation. Dans C. Dagenais et V. Ridde (dir.), *Approches et pratiques en évaluation de programmes* (p. 145-160). Presses de l'Université de Montréal.
- Pluye, P. (2018). Les méthodes mixtes. Dans C. Dagenais et V. Ridde (dir.), *Approches et pratiques en évaluation de programmes* (p. 125-143). Presses de l'Université de Montréal.
- Race, K. E. H., Hotch, D. F. et Packer, T. (1994). Rehabilitation Program Evaluation: Use of Focus Groups to Empower Clients. *Evaluation Review*, 18(6), 730-740. <https://doi.org/10.1177/0193841X9401800605>
- Rao, V. et Woolcock, M. (2003). Integrating Qualitative and Quantitative Approaches in Program Evaluation. Dans *The impact of economic policies on poverty and income distribution: Evaluation techniques and tools* (p. 165-190).
- Ryan, K. E., Gandha, T., Culbertson, M. J. et Carlson, C. (2014). Focus Group Evidence: Implications for Design and Analysis. *American Journal of Evaluation*, 35(3), 328-345. <https://doi.org/10.1177/1098214013508300>

- Secrétariat du Conseil du trésor. (2002). *L'évaluation de programme : document destiné aux dirigeants et dirigeantes de ministères et d'organismes*. Secrétariat, Conseil du trésor. Repéré à <http://collections.banq.qc.ca/ark:/52327/44281>
- Secrétariat du Conseil du trésor. (2014). Directive concernant l'évaluation de programme dans les ministères et les organismes. Repéré à <http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=1&file=61131.pdf>
- Secrétariat du Conseil du trésor. (2017). Cadre de suivi et d'évaluation préliminaire — Instructions destinées aux ministères et organismes. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/CSEP.pdf
- Secrétariat du Conseil du trésor. (2018a). Cadre d'évaluation : instructions destinées aux ministères et organismes. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/cadre_evaluation.pdf
- Secrétariat du Conseil du trésor. (2018b). Outil d'évaluation des programmes. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/outil_evaluation_programme.pdf
- Secrétariat du Conseil du trésor. (2018c). Plan pluriannuel d'évaluation — Instructions destinées aux ministères et organismes. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/PPE_Instructions_2018.pdf
- Secrétariat du Conseil du trésor. (2018d). Rapport d'évaluation — Instructions destinées aux ministères et organismes. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/rapport_evaluation_MO.pdf
- Secrétariat du Conseil du trésor. (2019a). *Évaluation des interventions interministérielles : guide d'accompagnement*. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/guide_evaluation_interventions_interministerielles.pdf
- Secrétariat du Conseil du trésor. (2019b). *Guide d'élaboration des politiques publiques*. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/guide_elaboration_politiques_publiques.pdf
- Secrétariat du Conseil du trésor. (2019c). Plan stratégique 2019-2023 du Secrétariat du Conseil du trésor. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/publications/planstrat_1923.pdf

- Secrétariat du Conseil du trésor. (2019d). Profil de compétences en évaluation de programme au gouvernement du Québec. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/depliant_profil_competence_web.pdf
- Secrétariat du Conseil du trésor. (2020a). Rapport annuel de gestion 2019-2020 du Secrétariat du Conseil du trésor. Repéré à https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/secretariat-du-conseil-du-tresor/publications-adm/rapport-annuel-de-gestion/rag_1920.pdf?1601488485
- Secrétariat du Conseil du trésor. (2020b). Rédaction et suivi des recommandations en évaluation — Guide d’accompagnement. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/Guide_redaction_suivi_recommandations.pdf
- Secrétariat du Conseil du trésor. (2020c). *Stratégie de gestion des dépenses — Renseignements supplémentaires. Budget de dépenses 2020-2021*. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/budget_depenses/20-21/1-Strategie_de_gestion_des_depenses_et_reenseignements_supplementaires.pdf
- Secrétariat du Conseil du trésor. (s. d.). Plan pluriannuel d’évaluation — Trousse d’accompagnement. Repéré à https://www.tresor.gouv.qc.ca/fileadmin/PDF/cadre_gestion/Trousse_SCT_plan_pluriannuel_evaluation.pdf
- Secrétariat du Conseil du Trésor du Canada. (2004). Examen de la qualité des évaluations dans les ministères et organismes. Repéré à <https://www.tbs-sct.gc.ca/cee/pubs/review-examen2004-fra.pdf>
- Smits, P. et Jacob, S. (2014). La fonction d’évaluation dans l’administration publique québécoise : analyse de la cohérence du système d’actions. *Canadian Public Administration*, 57(1), 71-96. <https://doi.org/10.1111/capa.12059>

ANNEXE 2 : LES QUESTIONS NORMALISÉES DE L'OUTIL D'ÉVALUATION DES PROGRAMMES ANALYSÉES LORS DE LA RÉALISATION DU BILAN

Tableau 11 : Les questions normalisées de l'*Outil d'évaluation des programmes* analysées lors de la réalisation du bilan

TABLEAU I - CARACTÈRE OBLIGATOIRE DES QUESTIONS SELON LE CYCLE DE VIE D'UN PROGRAMME^a

ÉLÉMENT DOCUMENTÉ	QUESTION	MISE EN ŒUVRE	RÉVISION
Pertinence de l'intervention	1 Les analyses réalisées sur les besoins et le contexte justifient-elles que le gouvernement intervienne ou continue à intervenir?	Obligatoire	Obligatoire
	2 Le programme est-il cohérent avec la mission, les mandats et les responsabilités de l'organisation et avec les orientations gouvernementales?	Obligatoire	Obligatoire
	3 Existe-t-il des chevauchements avec d'autres programmes qui visent des besoins comparables?	Obligatoire	Obligatoire
Pertinence de la nature de l'intervention	4 La nature de l'intervention est-elle la plus appropriée pour atteindre les objectifs de l'intervention?	Facultative	Obligatoire
Cohérence interne	5 Les composantes du programme et ses résultats s'articulent-ils logiquement?	Obligatoire	Obligatoire
Conformité opérationnelle	6 Le programme est-il géré adéquatement et conformément aux règles établies et conditions prévues?	Obligatoire	Obligatoire
Efficacité de la gouvernance	7 Les rôles et responsabilités des parties prenantes aux processus de planification, de coordination et de déploiement de l'intervention interministérielle sont-ils définis adéquatement?	Obligatoire interministériel	Obligatoire interministériel
	8 L'organisation responsable de l'intervention interministérielle exerce-t-elle son rôle de leader conformément aux attentes dans la planification, la coordination et le déploiement de l'intervention?	Obligatoire interministériel	Obligatoire interministériel
Efficacité opérationnelle	9 Les objectifs opérationnels sont-ils atteints?	Obligatoire	Obligatoire
Efficacité de l'intervention	10 Les objectifs de l'intervention sont-ils atteints?	Facultative	Obligatoire
	11 Les effets visés par l'intervention sont-ils obtenus?	Facultative	Obligatoire
	12 Des effets non intentionnels ont-ils été constatés?	Facultative	Facultative
Efficience relative de l'intervention	13 Comment les coûts des extrants se comparent-ils, notamment, à ceux d'autres programmes?	Facultative	Facultative
	14 Comment les coûts associés aux effets se comparent-ils, notamment, à ceux d'autres programmes?	Facultative	Facultative
	15 Existe-t-il des opportunités d'obtenir les mêmes résultats à moindre coût?	Facultative	Facultative
	16 Existe-t-il des opportunités d'obtenir de meilleurs résultats sans coûts supplémentaires?	Facultative	Facultative

Source : Secrétariat du Conseil du trésor (2018), p.9.

ANNEXE 3 : ORGANIGRAMME DES TÂCHES TEL QUE SPÉCIFIÉ DANS L'OFFRE DE SERVICE ET RÉALISATION EFFECTIVE RÉELLE

Tableau 12 : Organigramme des tâches de la réalisation du bilan des pratiques évaluatives

Éléments du projet	N°	Lots de travail et tâches prévus	Réalisation effective
	0	Réaliser un bilan des pratiques évaluatives	Réalisé
Lot 1	1	Réaliser les étapes préliminaires au lancement du mandat	Réalisé
	1.1	Lecture des documents du SCT	Réalisé
	1.2	Appropriation des différents outils et plateformes du SCT	Réalisé
	1.3	Lecture et exploration des documents déjà réalisés en lien avec le bilan	Réalisé
	1.4	Rédaction d'une ébauche du plan d'analyse	Réalisé
Lot 2	2	Officialiser la sélection des variables, des tableaux et des graphiques	Réalisé
	2.1	Dresser une liste préliminaire de variables	Réalisé
	2.2	Développer les ensembles catégoriels pour chaque variable	Réalisé
	2.3	Dresser une liste préliminaire de tableaux et de graphiques	Réalisé
	2.4	Terminer la rédaction du plan d'analyse final	Réalisé
Lot 3	3	Élaboration du tableau des données brutes du fichier Excel	Réalisé
	3.1	Apparier les données et fusionner les bases de données existantes	Réalisé
	3.1	Effectuer des tests de codage préliminaires	Réalisé
	3.2	Recoder pour les rapports déjà déposés, avec les nouvelles variables (incluant l'efficacité)	Réalisé partiellement (exclusion 2018)
	3.3	Intégrer les rapports manquants de 2018 et 2019 dans la base de données	Réalisé partiellement (exclusion 2018)
	3.4	Intégrer les rapports reçus en 2020 dans la base de données	Réalisé
Lot 4	3.5	Adapter les données finales pour la création des tableaux et des graphiques	Réalisé
	4	Élaboration des tableaux et des graphiques pour les enjeux de pertinence, d'efficacité et d'efficience	Réalisé
	4.1	Effectuer une formation rapide préliminaire Excel (intermédiaire et avancé)	Réalisé
	4.2	Explorer les données et ajuster la liste des tableaux et des graphiques	Réalisé
	4.3	Élaborer les tableaux	Réalisé
Lot 5	4.4	Élaborer les graphiques	Non réalisé (exclusion des graphiques)
	5	Rédaction du bilan (document d'accompagnement)	Réalisé
	5.1	Élaborer une ébauche du bilan	Réalisé
	5.2	Présenter l'ébauche aux membres de l'équipe d'évaluation de la DGEAIGR	Réalisé
	5.3	Terminer le document final du bilan	Réalisé
Lot 7	6	Élaborer le fichier PowerPoint	Réalisation post-stage
	6.1	Effectuer une formation rapide préliminaire PowerPoint (intermédiaire et avancé)	Réalisation post-stage
	6.2	Élaborer une ébauche du document PowerPoint	Réalisation post-stage
	6.3	Présenter l'ébauche du PowerPoint aux membres de l'équipe (voir la tâche 5.2.2)	Réalisation post-stage
	6.4	À partir des recommandations mentionnées à la tâche 5.2.3, apporter les corrections requises	Réalisation post-stage

ANNEXE 4 : MÉTHODOLOGIE DÉTAILLÉE DU PROCESSUS DE CODAGE DES VARIABLES

Définition des variables descriptives sélectionnées

Afin de documenter les caractéristiques générales des rapports et le contexte de réalisation de l'évaluation, six variables ont été retenues, soit l'année du dépôt de la dernière version, le dépôt préalable d'un CSEP, le dépôt préalable d'un CE, la présence d'un comité d'évaluation et son type le cas échéant, l'équipe d'exécution (internalisation ou externalisation de l'évaluation) et l'origine du besoin d'évaluation. Il est à noter que pour le type de comité d'évaluation, des critères furent établis afin de bien distinguer un comité interne d'un comité mixte. Plus précisément, il a été décidé d'utiliser ces critères d'inclusion :

Tableau 13 : Critères d'inclusion de la variable de la présence d'un comité d'évaluation et son type	
Comité interne	<ul style="list-style-type: none"> • Comité composé uniquement de membres provenant du MO, de la direction du programme et/ou de bénéficiaires du programme ;
Comité mixte	<ul style="list-style-type: none"> • Comité composé d'au moins un membre externe du MO, de la direction du programme et aussi non bénéficiaire (intervenants externes, partenaires, représentants des usagers, représentant d'un autre MO, etc.).
Comité avec composition incertaine ou non précisée	<ul style="list-style-type: none"> • Un comité est mentionné dans le rapport et/ou le cadre, mais les informations ne permettent pas de déterminer avec certitude si sa composition est interne ou mixte. Pour la plupart du temps, soit la composition n'est aucunement précisée, soit l'évaluation présente une liste de noms sans spécifier les postes ou les organismes de provenance de chacun des membres.
Pas de comité spécifié dans le CE et/ou le RE	<ul style="list-style-type: none"> • Aucun comité d'évaluation n'est mentionné dans le cadre ou le rapport d'évaluation. Il est à noter que cela ne signifie pas avec certitude qu'il y avait absence de comité pour l'évaluation en question.

Pour l'équipe d'exécution, elle a été codée dans une perspective globale de l'évaluation, c'est-à-dire qu'elle ne prend pas en compte des tâches spécifiques ayant été effectuées à l'externe. Par exemple, le fait que certains MO délèguent la réalisation du sondage par une firme externe n'est pas considéré comme étant une exécution externe si le MO réalise lui-même l'évaluation. Concernant l'origine du besoin d'évaluation, trois catégories ont été établies, soit :

Tableau 14 : Critères d'inclusion de la variable de l'origine du besoin d'évaluation	
Programmes autorisés par le CT sans évaluation requise	<ul style="list-style-type: none"> • Évaluation d'un programme autorisé par le Conseil du trésor. <ul style="list-style-type: none"> ○ Le programme est autorisé par le Conseil du trésor, mais l'évaluation n'est pas requise par la décision du CT.
Évaluations requises par le CT	<ul style="list-style-type: none"> • Programme autorisé ou reconduit par le Conseil du trésor dont un indiqué précise la nécessité que le programme soit évalué dans une échéance spécifique.
Besoin d'initiative	<ul style="list-style-type: none"> • Évaluation ayant été réalisée en réponse à des besoins d'information au MO
Autre exigence	<ul style="list-style-type: none"> • Évaluation ayant été réalisée en réponse à une autre exigence. Exemples d'autres exigences : <ul style="list-style-type: none"> ○ un décret du Conseil des ministres, une entente administrative, une convention de subvention ou un plan d'action d'application des recommandations du Vérificateur général du Québec (VGQ).

Il est à noter que cette variable était déjà présente dans le tableau de suivi et qu'elle a été reprise pour chacun des rapports. Cependant, des vérifications ont été apportées. Certains rapports ayant été codés « aucune obligation » dans le tableau de suivi n'ont été ainsi recodés « autre exigence » dans le cadre de ce bilan.

Deux types d'information ont été pris en compte pour le codage de la méthodologie employée dans les rapports : les méthodes utilisées et les parties prenantes consultées pour chacune d'entre elles. D'abord, pour les méthodes, trois catégories ont été utilisées : les entrevues, les sondages et les autres méthodes de collecte de données. Pour chacune d'entre elles, si la méthode n'était pas décrite dans la section dédiée à la méthodologie du rapport ou si celle-ci n'était pas clairement indiquée dans une autre section, le rapport était codé « Non ». Si le rapport faisait la mention de l'une des méthodes, la ou les parties prenantes consultées étaient codées à l'aide d'une liste déroulante distinguant les différentes combinaisons possibles :

Tableau 15 : Combinaisons des sources d'information pour les sondages, les entrevues et les groupes de discussion
<ul style="list-style-type: none">• Intervenants internes• Intervenants externes• Bénéficiaires• Clientèles• Intervenants internes + externes• Intervenants internes + bénéficiaires• Intervenants internes + clientèle• Intervenants externes + bénéficiaires• Intervenants externes + clientèle• Bénéficiaires + clientèle• Intervenants internes et externes + bénéficiaires• Intervenants internes et externes + clientèle• Intervenants internes + bénéficiaires + clientèle• Intervenants externes + bénéficiaires + clientèle• Intervenants internes + externes + bénéficiaires + clientèle• Partie(s) prenante(s) non spécifiée(s)

La spécification du type de clientèle (entreprises/organismes ou population) s'est faite à l'aide d'une colonne à part étant donné qu'elle a été décidée en cours de route (mais appliquée de manière rétroactive pour les rapports déjà codés).

Les entrevues incluent celles réalisées individuellement ou en petit groupe (deux ou trois personnes ou lorsqu'il semble clair qu'il ne s'agit pas d'un groupe de discussion). Elles peuvent inclure aussi les entrevues dirigées, semi-dirigées ou non structurées. Les consultations informelles n'ont pas été prises en compte, à moins que des citations ou des propos émis par des parties prenantes fussent clairement présentés dans la section dédiée aux résultats. Par exemple, certains rapports mentionnaient le fait que des parties prenantes ont été consultées, sans spécifier s'il s'agissait bien d'entrevues ou non. Ces consultations n'ont donc pas été codées comme étant des entrevues. Cependant, pour répondre à certaines questions dans certains rapports, il est parfois arrivé que des citations ou des propos aient été présentés en lien avec l'une des parties prenantes consultées. Dans ce cas-ci, ces consultations ont été codées comme étant des entrevues.

Pour les sondages, ils ont été pris en compte lorsqu'ils ont été réalisés dans le cadre de celle-ci, excluant ainsi les sondages réalisés dans les années précédentes à l'extérieur du cadre de l'évaluation (qui ont été intégrés à la revue documentaire dans ces derniers cas). Les sondages téléphoniques, électroniques ou en personne ont été pris en compte. De plus, les requêtes d'information, lorsque mentionnées sous forme de formulaires ou des questionnaires, ont été jugées comme étant des sondages.

Pour les autres méthodes, un codage brut a été réalisé, c'est-à-dire qu'aucune typologie ne fut préétablie d'avance. Les groupes de discussion y ont été codés. Cependant, elles incluent aussi d'autres méthodes, telles que les observations. Les études de cas, quant à elles, ont été intégrées aux autres méthodes de collecte de données étant donné les chevauchements qu'elles auraient pu générer avec les autres méthodes de collecte de données (par exemple, pour des entrevues réalisées dans le cadre d'une étude de cas, il aurait fallu les coder à deux reprises).

Concernant la revue documentaire et l'analyse des données administratives, il n'a pas été jugé utile de présenter les résultats pour la méthodologie globale employée puisque la quasi-totalité des rapports inclut à la fois des données administratives et de la documentation. Cependant, il a été jugé pertinent d'examiner pour chacune des questions normalisées si cette méthode avait bien été utilisée afin de poser un jugement évaluatif sur celles-ci.

Pour le codage des parties prenantes, une typologie à cinq catégories a été utilisée pour les trois colonnes. Les définitions ont été élaborées conjointement avec un évaluateur expérimenté. Voici la liste des cinq catégories ainsi que les critères d'inclusion :

Tableau 16 : Critères d'inclusion des parties prenantes consultées	
Intervenants internes	<ul style="list-style-type: none"> • Intervenants du MO qui occupent un poste (cadres, professionnels, autres employés) à l'intérieur de l'une des directions responsables du programme ou qui y sont impliqués de manière importante ; <ul style="list-style-type: none"> ○ Chargés de programme, analystes du ministère, gestionnaires et professionnels du programme, direction du MO, intervenants concernés par la gestion, le suivi et l'opérationnalisation du programme, etc.
Intervenants externes	<ul style="list-style-type: none"> • Intervenants du MO qui occupent un poste à l'extérieur des directions responsables du programme, mais qui y sont impliqués de manière importante (par exemple, le personnel d'une direction régionale d'un ministère) ; • Intervenants qui occupent un poste dans un autre MO ou autre acteur non bénéficiaire ; • Partenaires externes du programme qui ne semblent pas bénéficier financièrement du programme. <ul style="list-style-type: none"> ○ Représentants d'organismes partenaires, répondants régionaux qui accompagnent les bénéficiaires et qui contribuent à la mise en œuvre du programme, agents d'intervention directement impliqués dans l'implantation du modèle, etc.
Bénéficiaires	<ul style="list-style-type: none"> • Organismes/entreprises soutenus financièrement directement par le MO ; • Organismes/entreprises recevant des fonds provenant du MO et qui doivent les redistribuer à la clientèle par la suite. <ul style="list-style-type: none"> ○ Des fonds sont versés aux bénéficiaires (unique ou multiples) par le ministère afin de financer son fonctionnement et/ou les projets qu'elle soutient.
Clientèles — Population	<ul style="list-style-type: none"> • Citoyens, contribuables, clients, participants, patients ou tout autre groupe d'individus qui ne sont pas des entreprises ou des organismes et qui bénéficient de biens et/ou des services offerts par le programme.
Clientèles — Entreprises/organismes	<ul style="list-style-type: none"> • Organismes/entreprises soutenus financièrement indirectement par le MO par l'entreprise d'un bénéficiaire intermédiaire. <ul style="list-style-type: none"> ○ Des fonds sont versés au bénéficiaire par le ministère afin de financer son fonctionnement et/ou les projets qu'elle soutient. Cet organisme bénéficiaire est responsable de l'évaluation des projets et de l'octroi des subventions à des entreprises ou des organismes. Dans ce cas-ci, les entreprises et organismes sont considérés comme étant une clientèle plutôt que des bénéficiaires.

Codage du traitement des enjeux

Tout simplement, l'enjeu est considéré comme étant traité si au moins une question de l'*Outil* ou sous une autre forme associée à celui-ci est traitée dans le rapport.

Codage du traitement des questions

Le codage des questions s'est effectué à l'aide d'une grille de détermination des questions traitées. Comme il a été mentionné précédemment dans la stratégie d'atténuation des limites, une compréhension large des questions de l'*Outil*, incluant les questions apparentées, a été adoptée. Les questions de l'*Outil* ont donc été considérées comme étant des modèles génériques couvrant des thématiques particulières auxquelles il était possible d'ajuster le libellé en fonction de la réalité du programme plutôt que des

questions rigides. Lorsque des questions ou des éléments d'information correspondaient ou étaient similaires aux questions de l'Outil, il a été considéré que la question était traitée.

Il est à noter que le traitement des questions a été analysé dans une perspective effective, c'est-à-dire que le fait de poser tout simplement la question dans le rapport ou la matrice n'était pas suffisant pour considérer la question comme étant traitée. Autrement dit, c'est le traitement effectif et non l'intention du traitement qui a été considéré. Il était donc nécessaire de présenter des résultats et de répondre en partie à la question, lorsque considérée comme étant traitée.

Des critères spécifiques ont été établis en cours de route pour le traitement de l'efficacité opérationnelle (Q9) étant donné l'importance du degré d'extensibilité adopté pour considérer la question. Le tableau A1.5 comporte les critères d'inclusion et les critères d'exclusion pour cette question.

Tableau 17 : Critères d'inclusion et d'exclusion pour la question de l'efficacité opérationnelle (Q9)	
Critères d'inclusion	<ul style="list-style-type: none"> • Lorsque le libellé d'une question est identique à celui de <i>l'Outil d'évaluation des programmes</i> (« les objectifs opérationnels sont-ils atteints ? ») ; • Lorsque le libellé de la question n'est pas celui de <i>l'Outil</i>, mais que le rapport l'identifie comme relevant de l'enjeu de l'efficacité opérationnelle ; • Lorsqu'il y a une ou plusieurs questions, des indicateurs/cibles ou des sections qui visent à présenter de manière détaillée les extraits (bilan des extraits) ou à y poser un jugement (évaluation des extraits) : <ul style="list-style-type: none"> ○ Sauf si les éléments de réponse se retrouvent dans la section dédiée à la conformité opérationnelle. • Lorsqu'il s'agit d'une question en lien avec la mise en œuvre du programme, mais que celle-ci puisse avoir un impact soit sur le volume de biens ou de services produits, des délais ou de la qualité des extraits. Autrement dit, c'est la finalité de la question de la mise en œuvre qui doit viser l'efficacité opérationnelle ou l'efficacité d'une intervention. <ul style="list-style-type: none"> ○ Exemple : « Quels sont les facteurs qui facilitent ou entravent le fonctionnement du programme au regard de l'atteinte de ses objectifs ? » • Lorsque les questions concernent la satisfaction des clientèles ou des intervenants envers la qualité ou la quantité des extraits.
Critères d'exclusion	<ul style="list-style-type: none"> • Si les éléments de réponse abordent l'efficacité de l'intervention ou se réfèrent aux effets de l'intervention ; • Les questions liées à l'efficacité de la gouvernance • Si la question de la mise en œuvre touche davantage la nature de l'intervention ou le mécanisme d'intervention plutôt que l'efficacité de ce dernier. <ul style="list-style-type: none"> ○ Exemple : « Dans quelle mesure les modalités de mise en œuvre doivent-elles se poursuivre ou être améliorées pour répondre aux besoins des ONL ? »

10 rapports n'ont pas été pris en compte pour l'efficacité opérationnelle étant donné la difficulté de trancher la question du traitement. Ces difficultés rencontrées concernent souvent des questions de mise en œuvre qui mélangent à la fois l'efficacité et la nature de l'intervention. Aussi, l'absence de question ou l'éparpillement des éléments d'information dans différentes sections concernant les activités et les extraits ont été parmi les causes de la difficulté du codage du traitement de certains rapports. Aussi, un rapport fut exclu pour le traitement de Q10/Q11 en raison de la dernière cause mentionnée.

Codage des méthodes et des sources par question

Le codage des méthodes et des sources d'information reprend intégralement les catégories qui ont été présentées dans la section 2 du bilan concernant la méthodologie globale employée. À cela s'ajoute le codage de la revue documentaire et de l'analyse de données administratives, qui seront incluses dans les tableaux cette fois-ci.

Initialement, l'étalonnage avait été considéré comme étant une méthode à part entière. À la suite des tests préliminaires, nous avons décidé de le retirer étant donné une certaine confusion que cela générerait avec la revue documentaire. La revue documentaire peut inclure à la fois des documents provenant de

l'externe du MO (littérature scientifique, cadre légal et administratif, documents d'autres programmes, documents de suivi, etc.) que ceux de l'interne (documents du programme, rapports annuels de gestion, etc.). Les données administratives, quant à elles, incluaient l'ensemble des bases de données internes au MO. En résumé, lorsque de l'information de nature quantitative était présentée, qu'elle ne provenait pas de documents externes au MO ou à l'organisme responsable du programme (par exemple, des données de Statistique Canada) et que celle-ci ne provenait pas des autres méthodes de collecte de données (entrevues, sondages, groupes de discussion), généralement, elles ont été généralement codées comme étant des données administratives.

Comme le traitement des questions de l'*Outil*, c'est le caractère effectif de l'utilisation de méthode et des sources d'information qui a été codé pour chacune des questions et non le caractère intentionnel. Autrement dit, il ne s'agit pas des méthodes annoncées ou prévues pour chaque question qui a été analysée, mais la présentation de données probantes en lien avec cette méthode lors de l'interprétation des données. Pour ce faire, pour chaque question normalisée, l'analyse a consisté à parcourir les sections la traitant et à repérer, par une lecture en diagonale, les méthodes et les sources d'information. À la suite du passage en diagonale, pour chacun des rapports, des recherches par mots-clés ont été effectuées afin de vérifier que des éléments manquants n'ont pas été oubliés. Pour économiser du temps dans la collecte des données, lorsque l'ensemble des méthodes employées présentées dans la section méthodologique avait déjà été repéré dans le traitement effectif pour une question en particulière, on passait immédiatement à la suivante sans s'y attarder, la saturation étant atteinte.

Évidemment, comme il a été mentionné dans les limites, la détermination des sections faisant partie du traitement des questions n'a pas été facile à effectuer. Aussi, il est possible que certains oublis aient eu lieu durant la collecte des données. Cependant, il demeure que chacun des rapports a été annoté et le codeur a pris son temps afin de s'assurer que les données soient les plus fiables possible. Finalement, dès qu'une correction était apportée à une méthode et une source d'information pour un rapport en particulier, le codage de ces variables pour chacune des questions était automatiquement changé afin d'éviter les incohérences dans les résultats.

ANNEXE 5 : LIMITES DU BILAN ET MESURES D'ATTÉNUATION ADOPTÉES

Tableau 18 : Principales limites du bilan et mesures d'atténuation adoptées		
Limites initiales	Mesures d'atténuation adoptées	Limites résiduelles
<p>1. La collecte des données a été réalisée par une seule personne, ce qui accentue le risque d'un manque de validité interne du codage.</p>	<ul style="list-style-type: none"> • Les tests préliminaires ont permis au codeur de bien s'approprier les données et de repérer les désaccords avec les analystes précédents, en plus des désaccords entre ces derniers. En identifiant clairement les variables les plus sensibles, il a pu s'assurer d'établir des critères clairs qui permettent une standardisation du codage. • Pour certaines variables plus difficiles à coder, le codeur a fait appel à un regard extérieur par l'entremise d'évaluateurs expérimentés de la DGEAIGR afin d'établir avec eux des critères de codage. • Après avoir terminé la collecte de données, le codeur a repassé les variables les plus sensibles afin de réviser le codage. Lorsque des incohérences à travers le temps furent constatées, des vérifications ont été entreprises et des corrections ont été apportées. 	<ul style="list-style-type: none"> • Les stratégies d'atténuation n'éliminent pas entièrement les risques d'un manque de validité interne pour certaines variables, telles que les sources d'information.
<p>2. Des difficultés ont été rencontrées par rapport au codage des parties prenantes. Plus particulièrement, la distinction entre les bénéficiaires et la clientèle pouvait s'avérer difficile à faire dans les programmes d'aide financière avec un organisme intermédiaire recevant les fonds du ministère et redistribuant les sommes à d'autres par la suite.</p>	<ul style="list-style-type: none"> • Une définition précise des différentes catégories de parties prenantes a été élaborée. Afin de prendre en compte de la nature de l'intervention, une distinction entre les clients-entreprises et la « clientèle – population » a été ajoutée afin de mieux refléter la réalité dans les données présentées. • Des croisements ont été effectués entre les parties prenantes consultées et la nature de l'intervention (variable non présentée dans le bilan) afin d'assurer une cohérence avec les ajustements qui ont été apportés au cours de l'analyse. 	<ul style="list-style-type: none"> • Les catégories des parties prenantes ainsi que les critères sélectionnés auraient pu être différents si un autre analyste avait effectué le bilan. • Il n'est pas impossible que d'autres analystes aient codé différemment certaines parties prenantes malgré les critères.
<p>3. L'hétérogénéité des rapports d'évaluation selon les MO et le fait que plusieurs d'entre eux ne suivaient pas l'<i>Outil</i> a rendu particulièrement difficile la détermination des questions traitées, plus particulièrement pour les rapports provenant d'un besoin d'initiative.</p>	<ul style="list-style-type: none"> • Les questions 10 et 11 (objectifs et effets visés de l'intervention) ont été fusionnées afin de faciliter la détermination des questions traitées. De cette manière, il a été beaucoup plus facile non seulement de ne pas avoir à les distinguer entre elles, mais aussi de distinguer plus clairement l'efficacité opérationnelle de l'efficacité de l'intervention. • Pour la plupart des questions, une compréhension large des questions normalisées a été adoptée, c'est-à-dire que celles-ci pouvaient être considérées comme étant traitées, et ce même si la ou les questions étaient posées différemment. Pour la question 9 de l'<i>Outil</i> (efficacité opérationnelle), l'extension a été plus large étant donné que plusieurs questions/sections abordant la mise en œuvre traitaient indirectement la question. Des critères d'inclusion et d'exclusion ont été établis spécifiquement pour cette question. • Pour les questions liées aux enjeux de l'efficacité de la gouvernance (interventions interministérielles) et de l'efficience, celles qui étaient différentes de l'<i>Outil</i> ont été considérées à part entière et séparément des questions normalisées. Cette décision s'est justifiée par 	<ul style="list-style-type: none"> • Les stratégies d'atténuation n'éliminent pas entièrement les risques d'un manque de validité interne pour certains rapports.

	<p>le fait que les questions de l'<i>Outil</i>, pour ces enjeux, sont beaucoup plus précises et spécifiques que d'autres questions beaucoup plus générales et couvrant des thèmes plus vastes (analyse des besoins et du contexte, cohérence externe, chevauchements, pertinence de la nature de l'intervention, efficacité opérationnelle, efficacité de l'intervention).</p> <ul style="list-style-type: none"> • Une grille de détermination des questions traitées, comportant les libellés des questions, des objectifs et/ou des sections permettant de considérer chaque question comme étant traitée ou non, a été élaborée. Elle comportait aussi des colonnes afin d'y inclure les questions d'évaluation incertaines et des commentaires. Cette grille a été utilisée afin de réviser le codage en collaboration avec un évaluateur expérimenté. De cette manière, certaines corrections ont pu être apportées. 	
<p>4. La segmentation des sections dédiées au traitement de chacune des questions a été parfois très difficile à réaliser, rendant ainsi très compliqué le codage des méthodes et des sources par question traitée. Par exemple, certains rapports ne disposaient pas de questions évaluatives, d'autres éparpillaient différents éléments de réponses dans différentes sections, ou d'autres mélangeaient différents concepts dans différentes parties.</p>	<ul style="list-style-type: none"> • Une lecture rapide des sections difficiles à segmenter était effectuée afin d'analyser l'impact de son inclusion ou exclusion du corpus faisant partie du traitement de la question. • Pour certaines questions (Q9 et Q10/Q11), lorsque la segmentation des sections a été trop difficile à réaliser, le traitement de celles-ci a été codé « NA ». 	<ul style="list-style-type: none"> • Les mesures d'atténuation ne garantissent pas la validité interne des résultats.
<p>5. Le nombre de rapports déposés varie considérablement d'un MO à un autre. Un seul MO a déposé à lui seul 31,5% de l'ensemble des rapports d'évaluation en 2019 et 2020. Les résultats présentés seront teintés par cette forte représentation, ce qui pourrait rendre plus difficile l'interprétation des données.</p>	<ul style="list-style-type: none"> • Des tableaux ventilant les données par MO (plusieurs tableaux situés dans un document complémentaire) ont été élaborés afin de prendre en compte la pondération des MO dans les résultats finaux. De plus, des constats personnalisés par MO ont été établis. 	<ul style="list-style-type: none"> • Les tableaux ventilant les données par MO ne pourront pas être présentés au Réseau des répondants en évaluation de programme pour des raisons de confidentialité. • Les risques de résultats erronés s'accroissent lorsque des ventilations spécifiques sont effectuées (une seule erreur a plus d'impact sur un n faible qu'un n élevé). Si les données ventilées par MO ne sont pas interprétées avec parcimonie ou sont communiquées aux MO concernés sans vérifications préalables ou sans exemples concrets, ces derniers pourraient être sceptiques.