

Vingt ans d'évaluations d'impact en France et à l'étranger

Analyse comparée des pratiques dans six pays

Adam Baïz
Pierre-Henri Bono
Sandra Debu
Rozenn Desplatz
Clément Lacouette-Fougère
Alain Trannoy



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

Table des matières

Résumé	3
Introduction	4
Liste des bonnes pratiques identifiées	7
PARTIE I – CONTEXTE ET ACTEURS INSTITUTIONNELS	10
1. Un développement récent des évaluations d'impact porté principalement par les exécutifs nationaux	10
2. Une production concentrée sur quelques acteurs spécialisés	13
3. Une commande émanant majoritairement des administrations centrales à destination des milieux académiques	16
4. Une faible structuration des milieux professionnels sur le champ de l'évaluation d'impact	18
PARTIE II – L'OUVERTURE DU MILIEU ADMINISTRATIF À LA RECHERCHE	19
1. Des universitaires et des chercheurs très présents dans l'administration aux États-Unis, au Royaume-Uni, en Allemagne et en Suède	19
2. Des compétences et des formations aux méthodes d'évaluation d'impact inégales au sein de l'administration dans la quasi-totalité des pays.....	20
3. Des collaborations entre l'administration et la recherche souvent bien exploitées et parfois originales.....	22
PARTIE III – LES DISPOSITIFS FORMELS FAVORISANT LA CONDUITE ET L'UTILISATION DES ÉVALUATIONS D'IMPACT	25
1. Un cadre de l'évaluation structuré et coordonné au Canada et aux États-Unis, plus souple dans les autres pays.....	25
2. Des corpus et standards méthodologiques favorisant la montée en qualité, surtout aux États-Unis, au Royaume-Uni et en Allemagne	28
PARTIE IV – LE DEGRÉ DE DIFFUSION ET D'INFLUENCE DES ÉVALUATIONS	35
1. Un principe général de publication dans tous les pays, mais de portée inégale	35
2. Des outils de capitalisation et de diffusion des résultats particulièrement innovants au Royaume-Uni et aux États-Unis	36
3. Des incitations à l'utilisation des résultats en vigueur aux États-Unis et au Royaume-Uni, mais encore peu employées	39
Conclusion	43

Références	45
Annexe 1 – Grille d'analyse par pays	48

Résumé

Ce document de travail dresse un panorama de l'évaluation d'impact dans six pays : il compare la situation en France avec celle de cinq pays parmi les plus avancés en la matière, à savoir les États-Unis, le Royaume-Uni, la Suède, le Canada et l'Allemagne. Il s'agit d'appréhender les similitudes et les différences s'agissant de la commande et de la production des évaluations d'impact et de repérer les bonnes pratiques en matière d'utilisation et d'influence de ces travaux sur la décision publique. Cette étude fait suite à une [analyse bibliométrique publiée en décembre 2018](#), qui avait permis de quantifier la production d'évaluations d'impact en France par rapport aux autres pays. L'étude proposée ici se fonde sur [six monographies](#) qui examinent pour chaque pays la place des évaluations d'impact dans le processus de décision, sur la base de 15 questions regroupées en trois axes d'interrogation : quel est le degré d'ouverture du milieu administratif aux chercheurs ? Existe-t-il des dispositifs formels favorisant la conduite d'évaluations d'impact ? Enfin quelle est la diffusion et l'influence des évaluations sur la décision publique ?

La comparaison internationale montre qu'il n'y a pas de modèle « clé en main » pour encourager l'évaluation d'impact. Ces dernières années, chaque pays a développé son modèle de « politiques publiques fondées sur la preuve » en s'appuyant sur ses propres leviers d'action, au premier rang desquels la vitalité académique, l'impulsion politique et l'accès aux données administratives. Avec l'ouverture du milieu administratif aux acteurs de la recherche aux États-Unis, la formation des agents publics aux méthodes d'évaluation d'impact au Royaume-Uni ou encore la place des agences exécutives d'évaluation des politiques publiques en Suède, le panorama réalisé souligne qu'il existe des déterminants différents de réussite dans chaque pays.

Loin donc de plaider pour l'importation d'un modèle institutionnel unique, les bonnes pratiques repérées à l'étranger dessinent plutôt un ensemble de pistes à envisager et à expérimenter pour soutenir l'élan de l'évaluation d'impact en France. Elles mettent l'accent sur trois enjeux majeurs : la nécessité d'une meilleure articulation entre la demande et la production d'évaluation d'impact ; la définition de principes communs garantissant notamment l'indépendance, la crédibilité et la transparence des évaluations ainsi que le partage des enjeux, des pratiques et des résultats de l'évaluation d'impact au sein d'une alliance réunissant la communauté la plus large possible d'acteurs intéressés.

Mots clés : évaluation d'impact, politiques publiques, *ex ante*, *ex post*, administration, comparaison internationale

Vingt ans d'évaluations d'impact en France et à l'étranger

Analyse comparée des pratiques dans six pays

Adam Baïz, Sandra Debu, Rozenn Desplatz,
France Stratégie

Pierre-Henri Bono, conseiller scientifique
Clément Lacouette-Fougère, Sciences Po
et Alain Trannoy, conseiller scientifique¹

Introduction

Depuis vingt ans, l'évaluation d'impact des politiques publiques connaît un développement croissant en France. Reposant sur le principe qu'il est possible de mesurer, au moyen de méthodes statistiques et économétriques rigoureuses, « l'effet propre » d'une intervention publique, l'évaluation d'impact est souvent considérée comme la branche quantitative, voire scientifique, de l'évaluation des politiques publiques.

Plus encore que les autres formes d'évaluation – contrôle, audit, suivi de la mise en œuvre, etc. –, l'évaluation d'impact représente une exigence démocratique et un enjeu crucial pour aider les décideurs publics à améliorer l'efficacité de leurs politiques, et permettre aux citoyens de se former un avis éclairé sur la raison d'être de ces politiques.

De par la complexité des méthodes qu'elle mobilise, l'évaluation d'impact est principalement réalisée par des chercheurs en économie. Elle donne lieu le plus souvent à des publications

¹ Remerciements à Vincent Aussilloux pour son implication active et aux collègues de France Stratégie, ainsi qu'aux administrations et aux experts interrogés dans le cadre de ce projet.

scientifiques et ses résultats sont généralement discutés lors de séminaires et de journées d'études. Pour autant, son objet dépasse les enjeux purement académiques : l'évaluation d'impact vise à nourrir le débat public et les considérations sur les choix politiques. Elle implique aussi de changer la manière de concevoir les politiques publiques, en prenant le soin d'interroger les objectifs poursuivis, de collecter les données nécessaires à l'objectivation des résultats et d'expérimenter, lorsque c'est possible, les interventions publiques, avant de les généraliser.

Récemment, en France, plusieurs programmes ont fait l'objet d'une évaluation d'impact, comme le Crédit d'impôt pour la compétitivité et l'emploi (CICE), la Garantie Jeunes ou encore le dédoublement des classes primaires (CP) dans les zones d'éducation prioritaires, ce qui traduit un engouement renouvelé pour ces méthodes. L'Assemblée nationale a quant à elle lancé en 2018 le « Printemps de l'évaluation », une initiative visant à renforcer les missions de contrôle et d'expertise des parlementaires sur les textes de lois. Enfin, le prix Nobel d'économie décerné en octobre 2019 à trois chercheurs du laboratoire J-PAL, dont la franco-américaine Esther Duflo, a consacré l'intérêt pour ces méthodes d'évaluation statistiques.

Toutefois, le développement de l'évaluation d'impact reste modeste en France comparé à d'autres pays, en particulier anglo-saxons (Royaume-Uni, États-Unis), où cet exercice est depuis quelques années une pratique bien établie. Plusieurs pays comme la Suède, l'Allemagne ou le Canada ont également mis en place des dispositifs institutionnels (structures *ad hoc*, financements, réglementation) permettant d'encourager les évaluations d'impact et leur prise en compte dans la gouvernance publique.

C'est pourquoi cette étude vise à dresser un panorama complet des enjeux et des pratiques d'évaluation d'impact dans les pays les plus avancés. Elle propose une comparaison de la France avec cinq pays proches en termes de niveau de vie, de structures académiques et de système politico-administratif, mais considérés comme plus matures dans le domaine de l'évaluation d'impact, à savoir les États-Unis, le Royaume-Uni, la Suède, le Canada et l'Allemagne². Cette approche vise à appréhender les similitudes et les différences en matière de commande et de production des évaluations d'impact dans les différents pays et à repérer les bonnes pratiques en matière d'utilisation et d'influence des évaluations d'impact sur la décision publique.

Cette étude fait suite à une analyse bibliométrique publiée en décembre 2018³, qui avait permis de quantifier la production d'évaluations d'impact en France par rapport aux autres

² Cette sélection s'est appuyée sur une méta-analyse d'études portant sur les systèmes d'évaluation nationaux de divers pays (Bachtler *et al.* 2000 ; Furubo, Rist, Sandahl 2002 ; Jacob, Varone 2004 ; Polverari, Bachtler 2004 ; Sharma *et al.* 2013 ; Lázaro 2015 ; OCDE 2015 ; SGI network 2017). Pour chacune de ces études, nous avons recensé les pays considérés comme les plus avancés en matière d'évaluation selon les critères propres à l'étude (« culture » d'évaluation, institutionnalisation, maturité de la pratique, mise en œuvre effective, etc.). Nous avons ensuite retenu, dans l'ordre, les pays les plus cités dans l'ensemble des études. À noter que les Pays-Bas qui figuraient dans l'échantillon initial n'ont pas pu faire l'objet d'une fiche finalisée du fait des contraintes liées à la langue et à la disponibilité de notre équipe pour ce deuxième volet. Il convient également de souligner que l'évaluation au Canada relève davantage du suivi de la performance et des dépenses, mais présente des qualités et des mécanismes qui pourraient être transposés pour des évaluations *ex post* d'impact.

³ Bono P.-H., Debu S., Desplatz R., Hayet M., Lacouette-Fougère C. et Trannoy A. (2018), « Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique », *Document de travail*, n° 2018-06, France Stratégie, décembre.

pays. Ce travail avait souligné, entre autres, que si la France était en train de réduire son retard par rapport aux pays européens les plus prolifiques (Royaume-Uni, Allemagne, pays du Nord) elle présentait une production relativement modeste et fortement influencée par le poids de ses particularités politico-administratives et académiques (forte présence des services d'inspection et d'études au sein des ministères, faible ouverture de la sphère administrative aux milieux académiques, etc.).

Prolongeant ce travail, l'étude proposée ici se fonde sur six monographies⁴ qui examinent pour chaque pays la place des évaluations d'impact dans le processus de décision, sur la base de 15 questions regroupées en trois axes d'interrogation : quel est le degré d'ouverture du milieu administratif aux chercheurs ? Existe-t-il des dispositifs formels favorisant la conduite d'évaluations d'impact ? Enfin quelle est la diffusion et l'influence des évaluations sur la décision publique ? Chaque monographie s'appuie sur une revue de la littérature scientifique et de la documentation institutionnelle ainsi que sur une série d'entretiens auprès des acteurs nationaux de l'évaluation d'impact⁵.

Au-delà des réponses à ces questions, l'étude des différents pays permet d'identifier les bonnes pratiques dont la France pourrait s'inspirer afin de développer les évaluations d'impact, consolider leurs usages dans les politiques publiques et favoriser leur appropriation par les décideurs publics et les citoyens. Ces bonnes pratiques sont listées dans le tableau ci-après et détaillées tout au long du document, en lien avec leur contexte de conception et de mise en œuvre dans les pays concernés⁶.

⁴ Les [six monographies par pays](#) sont disponibles sur le site de France Stratégie : Baïz A. et Bono P.-H. (2019), « L'évaluation des politiques publiques en Allemagne », *Document de travail*, n° 2019-10, France Stratégie, décembre ; Lacouette-Fougère C. (2019), « L'évaluation des politiques publiques au Canada », *Document de travail*, n° 2019-11, France Stratégie, décembre ; Desplatz R. (2019), « L'évaluation des politiques publiques aux États-Unis », *Document de travail*, n° 2019-12, France Stratégie, décembre ; Desplatz R. et Lacouette-Fougère C. (2019), « L'évaluation des politiques publiques en France », *Document de travail*, n° 2019-13, France Stratégie, décembre ; Debu S. (2019), « L'évaluation des politiques publiques au Royaume-Uni », *Document de travail*, n° 2019-14, France Stratégie, décembre ; Debu S. (2019), « L'évaluation des politiques publiques en Suède », *Document de travail*, n° 2019-15, France Stratégie, décembre.

⁵ Cette structure et le [questionnaire qualitatif](#) sont rappelés en annexe 1 de chaque étude par pays.

⁶ Signalons enfin la note qui synthétise les enseignements de ces travaux : Baïz A., Debu S., Desplatz R. Bono P.-H., Lacouette-Fougère C. et Trannoy A. (2019), « [Évaluation d'impact des politiques publiques : que peut attendre la France des pays les plus avancés ?](#) », *Note de synthèse*, France Stratégie, décembre.

Liste des bonnes pratiques identifiées

BONNE PRATIQUE N° 1

Le programme de mobilité Intergovernmental Personnel Act (IPA), pour faciliter la présence de chercheurs au sein de l'administration aux États-Unis

Le programme américain IPA favorise les passerelles entre les milieux administratifs et académiques par l'échange de personnel qualifié. Les agences fédérales peuvent ainsi accueillir des universitaires ou des chercheurs pour une période déterminée (jusqu'à quatre ans) avec l'assurance pour ces derniers de pouvoir retourner dans leur université ou leur centre de recherche d'origine.

BONNE PRATIQUE N° 2

Les Policy Profession Standards au Royaume-Uni

Créés en 2010, les Policy Profession Standards sont des standards de compétences que doivent maîtriser les fonctionnaires britanniques impliqués dans la conception des politiques publiques (la Policy Profession). Ces standards comprennent des compétences techniques liées à l'évaluation. Ils sont utilisés à la fois pour le recrutement, l'évaluation individuelle des fonctionnaires et l'évolution de carrière.

BONNE PRATIQUE N° 3

Le développement de *Nudge units* dans tous les pays

Les *Nudge units* sont chargées d'appliquer les apports des sciences comportementales aux politiques publiques. Ces unités apportent généralement leur soutien aux administrations pour tester de nouvelles solutions et en évaluer rigoureusement les résultats à l'aide de méthodes expérimentales (RCT). Il en existe dans la plupart des pays étudiés.

BONNE PRATIQUE N° 4

Le Cross-Government Trial Advice Panel (CG-TAP) au Royaume-Uni

Créé en 2015 au Royaume-Uni, le CG-TAP apporte un accompagnement technique aux ministères souhaitant mettre en œuvre des évaluations d'impact randomisées. Il est composé d'une vingtaine d'experts de l'administration et d'une trentaine d'experts du monde académique. Une vingtaine de ministères et d'agences publiques au Royaume-Uni ont eu recours à ses services depuis sa création.

BONNE PRATIQUE N° 5

Les chefs évaluateurs pour animer, piloter et coordonner les évaluations dans les ministères aux États-Unis

La fonction de chef évaluateur est courante dans la plupart des grands départements ministériels américains, dont celui de l'emploi depuis 2010 (Department of Labor, DOL). Le chef évaluateur a généralement plusieurs missions dont celles de conduire et de superviser un portefeuille d'évaluations sur les politiques, programmes ou dispositifs du ministère ; d'accroître la qualité des évaluations et la connaissance des méthodes d'évaluation ; d'améliorer l'utilisation et la diffusion des résultats des évaluations ; enfin d'améliorer l'accès aux données.

BONNE PRATIQUE N° 6

Les guides méthodologiques édités par le Trésor britannique

Les publications éditées par le Trésor britannique depuis plus de quinze ans offrent la collection la plus complète de méthodologies d'évaluation à destination des administrations. Le *Green Book*, le document *Appraisal and Evaluation in Central Government*, le *Magenta Book* et son complément *Quality in policy impact evaluation* couvrent l'ensemble des dimensions de l'évaluation, de la mise en œuvre à l'impact.

BONNE PRATIQUE N° 7

Les standards de preuves

Les standards de preuve permettent de déterminer la qualité méthodologique des évaluations en fonction de la validité des méthodes utilisées. Les plus utilisés sont les « standards of evidence » édités par Nesta depuis 2013 au Royaume-Uni et l'échelle de Maryland (Maryland Scientific Methods Scale). Dans la pratique, ils sont le plus souvent mobilisés pour réaliser des revues systématiques (synthèse des enseignements tirés des évaluations répondant à un standard de qualité élevé) ou pour construire des bases de données de résultats issus des évaluations, appelées Clearinghouses.

BONNE PRATIQUE N° 8

Le titre d'évaluateur accrédité au Canada

Mis en place en 2009 par la Société canadienne d'évaluation, le titre d'évaluateur accrédité est une démarche de certification professionnelle qui garantit le respect de principes déontologiques et de normes professionnelles ainsi que d'un ensemble de 36 compétences théoriques et pratiques. En 2018, le Canada comptait 395 évaluateurs accrédités, parmi lesquels cependant une minorité de chercheurs en économie spécialisés dans les méthodes d'évaluation d'impact.

BONNE PRATIQUE N° 9

Les What Works Centres et les Clearinghouses, deux vecteurs de diffusion des résultats des évaluations d'impact

Les What Works Centres au Royaume-Uni et les Clearinghouses aux États-Unis ont pour but de centraliser les résultats des évaluations et de les rendre accessibles au plus grand nombre. Ils permettent de classer et de comparer des dispositifs selon notamment leur performance et la robustesse des méthodes utilisées. Ils adoptent généralement des

présentations visuelles pour faciliter leur compréhension par un public non expert des méthodes statistiques. Les études recensées reposent généralement sur des méthodes expérimentales et quasi expérimentales, et sont publiées dans des revues scientifiques ou sous la forme de documents de travail, de rapports d'évaluation, etc.

BONNE PRATIQUE N° 10

Les *Social Impact Bonds* au Royaume-Uni et aux États-Unis

Apparu pour la première fois en 2007 au Royaume-Uni, les *Social Impact Bonds* (ICB) permettent à des investisseurs privés de financer des interventions délivrées par des organisations non gouvernementales ou à but non lucratif qui, en cas de succès, seront ensuite remboursées par le gouvernement. Les interventions financées par SIB font l'objet d'une évaluation qui permet d'établir si le dispositif a ou non atteint ses objectifs, donc si l'investisseur privé sera remboursé par le décideur public. Connu sous le nom de *Pay for Success* (PFS) aux États-Unis, ce mécanisme est apparu pour la première fois en 2010. La plupart des projets de type PFS utilisent des méthodes rigoureuses d'évaluation, comme les méthodes expérimentales ou quasi expérimentales.

BONNE PRATIQUE N° 11

Les *Tiered-Evidence Grants* (TEG) aux États-Unis

Ce mécanisme financier proposé depuis 2010 par le gouvernement fédéral aux États pour la mise en œuvre de programmes sociaux repose sur des subventions dont les montants sont d'autant plus importants que les dispositifs proposés sont appuyés par des preuves solides de leur efficacité. La force de la preuve est fonction notamment de la méthode utilisée (méthodes expérimentales *versus* méthodes quasi expérimentales) et du nombre d'études à l'appui. Ce mécanisme incite ainsi les États à choisir des dispositifs efficaces pour obtenir des montants de subvention plus élevés.

PARTIE I – CONTEXTE ET ACTEURS INSTITUTIONNELS

1. Un développement récent des évaluations d'impact porté principalement par les exécutifs nationaux

L'évaluation d'impact des politiques publiques est une activité plutôt récente. Les pratiques les plus anciennes remontent aux années 1960 aux États-Unis à l'occasion de l'évaluation des grands programmes sociaux de lutte contre la pauvreté mis en place par l'administration Johnson. Ce développement précoce, combiné au dynamisme du milieu universitaire américain, explique en grande partie la maturité de la pratique de l'évaluation d'impact des politiques publiques aux États-Unis.

En Europe, le Royaume-Uni fait figure de précurseur dans le recours aux évaluations d'impact, avec le développement des premiers travaux dans les années 1980 et un élan de grande ampleur à la fin des années 1990. Les autres pays européens (France, Suède, Allemagne) ont développé les évaluations d'impact plus tardivement, à partir des années 2000. En Suède cependant, ces évaluations ont bénéficié d'un contexte favorable (ouverture des données administratives, intérêt politique, création d'agences spécialisées dans l'évaluation des politiques publiques) qui a permis au pays de développer rapidement une pratique et une culture solides de l'évaluation d'impact. L'Allemagne s'est intéressée dès les années 1960 à l'évaluation de l'action publique, mais il a fallu attendre les années 2000 avec de grandes réformes du marché du travail pour voir apparaître de plus en plus fréquemment l'évaluation d'impact. La France a connu, pour sa part, un rattrapage plus étalé dans la dernière décennie, et sans doute moins abouti à ce jour. Enfin, le Canada connaît une situation particulière : si la pratique de l'évaluation des politiques publiques a été développée et formalisée dès les années 1970-1980 au sein de l'appareil politico-administratif, avec la mise en place en 2001 d'un cadre réglementaire rendant obligatoire l'évaluation dans les ministères et organismes publics, la pratique de l'évaluation ne couvre aujourd'hui que très marginalement les évaluations d'impact et porte davantage sur le suivi de la bonne exécution des programmes.

Le développement de l'évaluation des politiques publiques apparaît le plus souvent comme la résultante d'un processus continu de rationalisation croissante des politiques publiques. Ce *continuum* débute par la mise en place, dans les différents pays, d'activités de contrôle des dépenses puis est prolongé par le suivi de la performance avant de se traduire, dans la période récente, par le développement de méthodes d'évaluation des résultats et des impacts des programmes publics. Ce processus, fortement marqué par l'application des théories du nouveau management public diffusées à partir des années 1980, a été suivi par l'ensemble des pays de notre échantillon. Dans plusieurs pays – États-Unis, Royaume-Uni et dans une moindre mesure France et Allemagne –, **il se double d'un mouvement plus récent en faveur de l'evidence-based policy**, c'est-à-dire de politiques publiques davantage fondées sur les preuves et notamment sur la démonstration, grâce à l'évaluation scientifique, de leur impact sur les publics ou sur les grandeurs ciblées. Ce second mouvement a facilité dans les pays concernés le développement des évaluations d'impact et leur autonomisation vis-à-vis des autres types d'évaluations et démarches voisines (contrôle, audit, suivi,

etc.) avec lesquels elles sont encore parfois confondues dans la plupart des autres pays (Canada, Suède, France).

Dans les pays précurseurs (États-Unis et Royaume-Uni), **des impulsions politiques fortes émanant de l'exécutif** ont joué un rôle décisif dans le développement et la généralisation des évaluations d'impact dans tous les domaines de l'action publique. Récemment aux États-Unis, l'administration Obama a été à l'origine de l'adoption en 2018 de la loi [Foundations for Evidence-Based Policymaking Act \(FEBPA\)](#) qui améliore l'accès aux données sensibles et renforce les protections entourant la confidentialité des données ainsi que la capacité du gouvernement fédéral à produire et à utiliser les éléments de preuve. Au Royaume-Uni, le choix du gouvernement Blair de créer, à la fin des années 1990, le premier des What Works Centres⁷ dans le domaine de la santé (le National Institute for Health and Care Excellence, NICE) a eu une influence déterminante dans le développement des politiques publiques fondées sur la preuve.

Les autres pays, par contraste, n'ont pas connu d'impulsion politique majeure en faveur de l'évaluation d'impact. En France, l'élan donné par Martin Hirsch lors de la création du Fonds d'expérimentation pour la jeunesse (FEJ) en 2009 ne s'est pas traduit par un développement massif de la pratique d'évaluation d'impact. Plus récemment, les propositions émanant de l'Assemblée nationale en faveur des évaluations quantitatives et l'instauration d'un « printemps de l'évaluation »⁸ n'ont pas encore trouvé de concrétisation institutionnelle.

En Allemagne, les réformes du marché du travail de 2004, dites réformes Hartz, ont établi les évaluations d'impact *ex post* comme l'outil fondamental de la mesure de l'efficacité d'une politique publique. La mise en œuvre d'évaluations d'impact *ex post* dans ce cadre a bénéficié de plusieurs événements favorables – un tournant néolibéral du gouvernement allemand, mais aussi une grande disponibilité des données concernant le marché du travail et l'existence de grands panels comme le [German Socio-Economic Panel \(GSOEP\)](#) produit par l'Institut pour la recherche économique ([Deutsches Institut für Wirtschaftsforschung, DIW](#)) – sans pour autant créer un systématisme quant à la production et l'utilisation d'évaluations d'impact dans d'autres secteurs de politique publique.

Dans tous les pays étudiés, **le Parlement a mission d'évaluer les politiques publiques**. Cette mission est parfois inscrite formellement dans les textes constitutionnels – comme en France ou en Suède – et réalisée par des commissions ou missions parlementaires. La contribution effective du Parlement à l'évaluation des politiques publiques varie cependant selon les pays. **Si, dans aucun des pays étudiés, le Parlement ne produit d'évaluations d'impact *ex post*, son rôle dans le suivi des politiques publiques et dans leur évaluation *ex ante* est parfois important, notamment aux États-Unis et en Suède (voir encadré 1).**

⁷ Allard & Rickey, ANSA, 2017

⁸ Barrot, groupe de travail sur les moyens de contrôle et d'évaluation de l'Assemblée National, 2018 ; CEC, Morel-à-L'Huissier & Petit, 2018

Encadré 1 – Le rôle du Parlement dans l'évaluation des politiques publiques

Le Congrès américain s'appuie sur trois instances d'expertise : le [Government Accountability Office \(GAO\)](#), le [Congressional Research Service \(CRS\)](#) et le [Congressional Budget Office \(CBO\)](#). Ces agences ne réalisent pas d'évaluation d'impact à proprement parler mais elles peuvent s'appuyer sur celles réalisées ou commandées par l'exécutif. Le GAO est la plus importante d'entre elles avec un budget annuel de plus de 600 millions de dollars en 2018. Ses évaluations de programmes consistent en des audits de performance réalisés à l'aide de méthodes qualitatives (entretiens, questionnaires, observations, etc.) ou quantitatives (exploitation de données). De son côté, le CBO élabore principalement, pour le compte du Congrès, des projections économiques et budgétaires sur dix ans et trente ans et produit des estimations de coût des projets de loi et des options législatives envisageables.

En Suède, le Parlement est impliqué dans l'évaluation *ex ante* grâce à la pratique des [comités d'enquête](#) nommés par le gouvernement lorsqu'une réforme majeure est envisagée. Ces comités, qui existent depuis le XVI^e siècle, se composent de 8 à 10 personnes, incluant des experts, des parlementaires – bien qu'ils soient juridiquement des organes de l'exécutif – et des parties prenantes (associations, syndicats, etc.). La mission de ces comités est de réaliser un diagnostic des problèmes publics à traiter, de définir les enjeux principaux et de formuler des options de réformes sous la forme de projets de textes de loi. Les comités s'appuient sur une revue des connaissances existantes dans le domaine concerné : historique des réformes antérieures, données sur leur mise en œuvre, leurs résultats et leurs impacts, comparaisons internationales, etc. Le gouvernement mandate environ 140 nouveaux comités par an, et quelque 200 comités travaillent simultanément en moyenne. Ils sont installés le plus souvent pour une durée d'un an ou deux, tandis que certains comités sont permanents.

En Allemagne, le Parlement fédéral a le pouvoir de contrôler chaque euro dépensé par un organisme public ou pour toute action publique, dans une logique d'efficacité des dépenses publiques : à cet égard, il commande près de 200 rapports par an, dont 80 environ concernent une action publique évaluée de manière annuelle. À tous les niveaux de gouvernance, des audits sont par ailleurs menés pour vérifier la régularité des procédures fédérales de construction des budgets et des dépenses.

Dans les autres pays, le rôle du Parlement est plus limité à la fois pour l'évaluation *ex post* et *ex ante*. En France, les principales fonctions exercées par les parlementaires restent le vote de la loi et le contrôle de l'exécutif. La [fonction d'évaluation des politiques publiques](#), consacrée depuis la réforme constitutionnelle de 2008, est concentrée au sein de deux entités : la mission d'évaluation et de contrôle (MEC) chargée de contrôler l'utilisation des dépenses publiques et la mission d'évaluation et de contrôle des lois de financement de la sécurité sociale (MECSS) chargée de vérifier l'application des lois de financement de la sécurité sociale et de procéder à l'évaluation de toute question relative aux finances de la sécurité sociale. À ce dispositif est venu s'ajouter, en 2009, le comité d'évaluation et de contrôle (CEC), composé de 36 députés, qui coordonne chaque année en moyenne 6 rapports d'évaluation, avec l'appui de la Cour des Comptes. Au Canada, les organes du Parlement ne pratiquent pas non plus directement l'évaluation. Les commissions parlementaires mobilisent cependant les rapports du [Bureau du Contrôleur général du Canada \(BCG\)](#), qui, afin de contrôler les activités gouvernementales, recueille lors d'audits une information objective et factuelle et des conseils experts sur les activités du gouvernement.

Enfin une tendance récente dans la pratique de l'évaluation d'impact est commune à tous les pays de l'échantillon : le **recours plus important aux données administratives**, bien que la situation soit variable selon les pays. L'utilisation de ces données engendre également des enjeux de disponibilité, à la fois en termes d'accès par les chercheurs et en termes de délai ; dans tous les pays, les administrations effectuent aujourd'hui un effort de disponibilité de plus en plus conséquent et rapide de leurs données, ce qui permet la réalisation d'évaluations d'impact dans des délais raccourcis (Bozio et Geoffard, 2017).

2. Une production concentrée sur quelques acteurs spécialisés

Du fait de la complexité des méthodes mobilisées, dans tous les pays de l'échantillon à l'exception de l'Allemagne, **les principaux producteurs d'évaluations d'impact sont les grandes universités nationales en économie**. On peut citer l'École d'économie de Paris pour la France (voir encadré 2), les universités de Stockholm et d'Uppsala en Suède, l'université de Laval au Canada, la London School of Economics et l'University College London au Royaume-Uni, les universités d'Harvard, de Chicago, du Michigan et de Pennsylvanie ainsi que le Massachusetts Institute of Technology (MIT) aux États-Unis.

Encadré 2 – Les acteurs universitaires en France

En France, plusieurs structures spécialisées dans l'évaluation d'impact des politiques publiques et rattachées aux universités ou à des grandes écoles ont été créées il y a une dizaine d'années : outre J-PAL Europe, on peut citer l'IPP, le LIEPP et le TEPP.

Créé en 2007 au sein de l'École d'économie de Paris, **J-PAL Europe** est l'un des six bureaux régionaux de J-PAL, le laboratoire d'action contre la pauvreté, fondé en 2003 au département d'économie du Massachusetts Institute of Technology (MIT) aux États-Unis. La spécificité de J-PAL repose sur l'utilisation exclusive d'expériences contrôlées pour mesurer les effets des programmes sociaux et de lutte contre la pauvreté. J-PAL Europe regroupe 44 chercheurs affiliés dans 16 universités en Europe et travaillant sur des projets d'évaluation divers en France, en Europe et en Afrique francophone. Depuis sa création, J-PAL Europe a conduit ou engagé 53 projets d'évaluations randomisées sur l'Europe, dont 26 sur la France. Ces projets couvrent différents domaines de la politique publique comme la criminalité et la violence, la finance, le genre, le marché du travail, la santé, l'économie politique et la gouvernance ainsi que l'éducation.

L'**Institut des politiques publiques (IPP)**, créé en 2012 sur le modèle de l'*Institute for Fiscal Studies* (IFS) au Royaume-Uni, est un institut d'évaluation généraliste couvrant un large éventail de politiques publiques (fiscalité, retraites, éducation, emploi, environnement, santé, etc.). Doté d'un budget d'environ un million d'euros, il est financé par une aide de l'État gérée par l'Agence nationale de la recherche (ANR) au titre du programme d'investissements d'avenir ainsi que par des contrats de recherche et d'expertise. Il est constitué d'une équipe permanente (24 économistes et doctorants) et d'une soixantaine de chercheurs issus de l'École d'économie de Paris et du CREST. L'IPP promeut une approche quantitative de l'évaluation des politiques publiques en s'appuyant sur des méthodes variées (expériences contrôlées, expériences naturelles, économie structurelle, micro-simulation, analyse coût-bénéfice, économie expérimentale) et sur des données individuelles issues d'enquêtes ou de sources administratives.

De son côté, le **Laboratoire interdisciplinaire d'évaluation des politiques publiques (LIEPP)**, créé en 2011, est un laboratoire spécialisé dans l'évaluation des politiques publiques au sein de Sciences Po dont l'approche est fondée sur l'interdisciplinarité. Également financé par l'ANR, il regroupe plus de 180 chercheurs issus de différentes disciplines (économie, sociologie, sciences politiques, droit et histoire). La recherche au LIEPP est structurée autour de quatre axes principaux : les discriminations et inégalités sociales, l'évaluation de la démocratie, les politiques éducatives et les politiques socio-fiscales.

Enfin, la fédération de recherche **Théorie et évaluation des politiques publiques (TEPP)**, créée en 2008, regroupe dix laboratoires de recherche de dix universités. Elle mobilise les méthodes quantitatives empiriques et théoriques en économie, ainsi que les méthodes qualitatives en sociologie, en gestion, en droit et en sciences politiques. Elle n'a pas de ressources budgétaires récurrentes, ses ressources propres étant issues des contrats de recherche.

Dans plusieurs pays, d'autres **producteurs indépendants des universités** interviennent dans la production de travaux d'évaluation d'impact, qu'il s'agisse d'organismes de recherche autonomes, de cabinets de conseil privés ou de *think tanks* :

- **En Allemagne, les évaluations sont majoritairement réalisées par des instituts de recherche non universitaires** : plus ou moins liés aux structures gouvernementales et au monde universitaire, ces instituts sont très présents dans le débat conjoncturel économique et sont également producteurs de données statistiques. Ces instituts sont régulièrement commissionnés par les ministères pour mettre en œuvre l'évaluation de programmes ou de politiques publiques. Certains sont généralistes comme le Centre de sciences sociales de Berlin ([Wissenschaftszentrum Berlin für Sozialforschung, WZB](#)) ou l'Institut allemand pour la recherche économique ([Deutsches Institut für Wirtschaftsforschung, DIW](#)), tandis que d'autres sont spécialisés sur une thématique à l'instar de l'Institut d'économie du travail ([Forschungsinstitut zur Zukunft der Arbeit, IZA](#)) ou l'Institut pour la recherche sur l'emploi ([Institut für Arbeitsmarkt und Berufsforschung, IAB](#)) dans le domaine de l'emploi. Le statut légal de ces instituts est variable. Certains sont des entreprises privées, le plus souvent à but non lucratif et à financement partiellement public. D'autres sont statutairement liés à des autorités administratives, comme l'[IAB](#), créé par le ministère du Travail et des Affaires sociales. Les instituts de recherche disposent de financements et de personnels relativement importants, au moins depuis les années 2010.
- **Au Royaume-Uni, de nombreuses structures de petite taille interviennent aux côtés des universités.** Il s'agit tout d'abord d'organismes indépendants de recherche parmi lesquels l'[Institute for Fiscal Studies \(IFS\)](#) tient un rôle de référence avec ses analyses budgétaires qui évaluent *ex ante* les projets de loi de finances ainsi que les propositions des principaux partis politiques dans les périodes pré-électorales. L'IFS est ainsi spécialisée sur les questions fiscales et budgétaires, mais ses domaines de recherche sont variés (éducation, marché du travail, santé, retraites, inégalités, innovation, Brexit, etc.). Il existe également de nombreux *think tanks* et cabinets de conseil qui complètent le paysage britannique de l'évaluation (Kantar Public, Ipsos Mori, etc.). En outre, des structures innovantes proches de l'administration ont été développées récemment, en particulier la [Behavioural Insights Team](#), placée auprès du Premier ministre, qui conduit des expérimentations reposant sur les sciences comportementales, ou encore l'[Education Endowment Foundation \(EEF\)](#), un What Works Centre consacré à l'éducation, qui réalise des évaluations d'impact randomisées concernant aujourd'hui une école sur trois en Angleterre.
- **Aux États-Unis, plusieurs grands organismes de recherche indépendants sont très souvent sollicités par les ministères et les agences fédérales pour évaluer leurs programmes.** [MDRC](#) (connu sous le nom de Manpower Demonstration Research Corporation avant 2003) conduit des évaluations dans le domaine des politiques sociales à l'aide de méthodes à la fois qualitatives et quantitatives et contribue également à développer et à faire connaître les méthodologies d'évaluation. Le [Mathematica Policy Research Group](#) mène également des évaluations dans le domaine des politiques sociales pour le gouvernement et le secteur privé. En outre, il définit et met régulièrement à jour différentes bases de données utilisées par des agences fédérales pour recenser les dispositifs les plus efficaces ([What Works Clearinghouse](#) du Department of Education ou [CLEAR](#) du Department of Labor).

Dans ces trois pays, si les administrations produisent aussi parfois des évaluations d'impact, la majorité des besoins en évaluation sont couverts par l'externalisation à d'autres acteurs (universités, organismes indépendants, etc.), afin de garantir l'indépendance des travaux. À l'inverse, la France, le Canada et dans une moindre mesure la Suède fonctionnent différemment en raison du **poids important de structures administratives** dans la production d'évaluations d'impact.

- **En France, l'Institut national de la statistique et des études économiques (Insee) et les services statistiques ministériels (SSM) sont des producteurs majeurs d'évaluations d'impact.** Cette singularité tient à la fois à la place importante dévolue à l'économie quantitative à l'Insee dès sa création en 1946, et à un accès quasi exclusif aux données par les administrations pendant plusieurs décennies avant l'ouverture progressive des micro-données aux chercheurs. L'Insee a ainsi comme mission, outre la production statistique, la réalisation d'études économiques et de recherches qui incluent les évaluations quantitatives de politique publique. De leur côté, les ministères disposent pour la plupart d'un service de statistiques propre qui, au-delà de leur mission originelle consistant à collecter des données, à produire et diffuser des statistiques publiques, réalisent également des études dans leurs domaines de compétences, qui peuvent aller jusqu'à « la participation à des évaluations de politiques publiques » (par exemple les ministères du Travail, de la Santé ou de l'Éducation nationale)⁹.
- **Au Canada, beaucoup d'évaluations sont produites par les administrations mais relèvent rarement de l'évaluation d'impact.** [Statistics Canada](#), l'organisme national de statistique canadien, est un acteur important dans la production des données nécessaires à l'évaluation mais également dans la réalisation de travaux d'évaluation d'impact.
- **En Suède, les agences exécutives d'évaluation des politiques publiques**, indépendantes des ministères, produisent des évaluations des programmes correspondant à leur spécialisation sectorielle. Certaines de ces agences sont de véritables organismes de recherche produisant de grands volumes d'évaluations d'impact, en particulier l'[Institute for Evaluation of Labour Market and Education Policy \(IFAU\)](#), consacré aux politiques d'emploi et d'éducation, dont la proximité avec l'université d'Uppsala permet d'alimenter un vivier de chercheurs spécialisés dans l'évaluation.

On peut retenir de ce panorama, enfin, **le fort cloisonnement thématique qui caractérise les producteurs d'évaluation d'impact**. En Suède, au Royaume-Uni comme au Canada, la majorité des producteurs sont des structures d'études spécialisées sur un champ particulier de politiques publiques (travail, santé, éducation, etc.). Il n'y a qu'en France, en Allemagne et aux États-Unis que l'on trouve quelques **acteurs transversaux de l'évaluation d'impact des politiques publiques** ; ces derniers ont émergé récemment et cohabitent généralement avec les acteurs sectoriels. En Allemagne, deux instituts, le DIW et le WZB, couvrent aujourd'hui l'ensemble des domaines économiques.

⁹ [Charte des services statistiques ministériels, 2017.](#)

3. Une commande émanant majoritairement des administrations centrales à destination des milieux académiques

Dans tous les pays de l'échantillon, **la commande d'évaluations d'impact de politiques publiques émane principalement des ministères**, en particulier dans les domaines du marché du travail, de l'éducation et de la santé. Les moyens humains et financiers consacrés à l'évaluation d'impact sont souvent très inégaux selon les ministères. Aux États-Unis, certains ministères possèdent des unités d'études et d'évaluation particulièrement développées, ayant un rôle avancé de pilotage et de coordination de ces activités pour les différentes agences du ministère. Par exemple, le [Chief Evaluation Office \(CEO\)](#) du ministère du Travail (US Department of Labor, DOL), qui comprend 15 personnes et dispose d'un budget de 21,4 millions de dollars en 2018, a notamment un rôle d'appui et d'assistance pour l'ensemble des évaluations conduites ou commandées au sein du ministère. En France, les services statistiques ministériels (SSM) consacrent une part de leurs missions à la réalisation ou au pilotage d'évaluations de politiques publiques à partir notamment des données qu'ils produisent. C'est le cas en particulier de la Direction de l'animation de la recherche, des études et des statistiques (DARES) du ministère du Travail, de la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) du ministère de la Santé et de la Direction de l'évaluation, de la prospective et de la performance (DEPP) du ministère de l'Éducation nationale. Dans de nombreux cas également, les chercheurs se saisissent eux-mêmes des sujets, indépendamment de la demande d'évaluations émanant des commanditaires.

Concernant l'Allemagne, les ministères bénéficient d'une très forte autonomie par rapport à la chancellerie. Certains ministères possèdent des cellules spécialisées, d'autres s'appuient sur des structures indépendantes qu'ils financent en grande partie. Par exemple, l'agence du ministère de l'Économie et de l'Énergie en charge des exportations, le *Bundesamt für Wirtschaft und Ausfuhrkontrolle* (BAFA), dispose depuis 2008 d'une [unité d'évaluation](#) interne qui réalise des évaluations de programmes pour le BAFA ou le ministère et assure le suivi des évaluations externalisées. Sur un modèle différent, l'[Institut pour la recherche sur l'emploi](#) (*Institut für Arbeitsmarkt- und Berufsforschung, IAB*), est rattaché au Service fédéral pour l'Emploi (BA pour *Bundesagentur für Arbeit*). En 2017, l'IAB employait 274 personnes, dont 187 académiques, avec un budget estimé entre 30 et 40 millions d'euros.

Dans certains pays (Royaume-Uni, États-Unis, Canada), **l'administration en charge du budget** joue un rôle particulier dans l'évaluation, notamment pour favoriser leur développement ou pour produire des lignes directrices sur la conduite des évaluations. Ainsi, au Royaume-Uni, le Trésor a publié de nombreux guides et recommandations à destination des administrations. Aux États-Unis, l'[Office of Management and Budget \(OMB\)](#), administration en charge du budget à la Maison blanche, comprend une petite équipe dédiée (cinq personnes environ), l'Evidence Team, qui réfléchit aux moyens de développer les évaluations et leur utilisation au sein des agences fédérales. De même, le [Secrétariat du Conseil du Trésor \(SCT\)](#) du Canada est chargé dans le domaine de l'évaluation de vérifier les compétences et les capacités des unités internes d'évaluation dans les départements ministériels ainsi que la qualité des travaux d'évaluation recueillis. Il s'est doté en 1985 d'un [Centre d'excellence pour l'évaluation \(CEE\)](#) pour accompagner la montée en compétences des unités d'évaluation au sein des administrations et des organismes publics (un réseau d'environ 300 évaluateurs au sein du gouvernement fédéral).

Dans les deux pays plus centralisés, la France et le Royaume-Uni, **des structures interministérielles ou ad hoc** ont pour rôle d'impulser des évaluations au sein de l'administration. Au Royaume-Uni, la What Works Team, dirigée par David Halpern et hébergée par le Cabinet Office, promeut la pratique des évaluations d'impact dans les ministères et coordonne le réseau des What Works Centres. Elle a notamment créé en 2015 un [Cross-Government Trial Advice Panel \(CG-TAP\)](#), composé d'experts de l'administration (24) et du monde académique (27), qui forme et aide techniquement les ministères souhaitant mettre en œuvre des évaluations d'impact randomisées. En France, [France Stratégie](#) a développé une expertise reconnue en matière d'évaluation d'impact. France Stratégie porte plusieurs évaluations d'impact de politiques publiques *ex post*, confiées à des commissions d'évaluation regroupant les parties prenantes (rassemblant selon les cas des partenaires sociaux, des parlementaires, des administrations, des chercheurs et des experts). Les travaux d'évaluation eux-mêmes sont la plupart du temps confiés à des chercheurs et aux services statistiques ministériels. L'organisme mène également des réflexions sur les méthodes d'évaluation via l'élaboration de guides méthodologiques ou l'organisation d'événements et séminaires dédiés, par exemple lors du [cycle de séminaires](#) organisé par France Stratégie en 2017-2018.

En outre, dans les pays anglo-saxons, les **fondations** interviennent pour financer des évaluations d'impact et promouvoir leur utilisation par les décideurs publics. C'est le cas par exemple de [Nesta](#) au Royaume-Uni, qui finance des projets innovants dans les domaines économiques et sociaux, et organise des conférences pour favoriser l'utilisation des preuves scientifiques dans la décision publique. Aux États-Unis, par exemple, la [Mac Arthur Foundation](#) et la [Laura & John Arnold Foundation \(LJAF\)](#) financent et promeuvent presque exclusivement les méthodes expérimentales (RCT). La fondation LJAF soutient une coalition de chercheurs, [Evidence-Based Policymaking Collaborative](#), issus de différents organismes et instituts, dont le but est de développer la conception de politiques publiques fondées sur la preuve, notamment au niveau fédéral. La fondation Mac Arthur soutient pour sa part la [Pew-Mac Arthur Results First Initiative](#) pour aider les États à adopter des politiques publiques fondées sur la preuve (à partir d'analyses coûts-bénéfices de leurs politiques).

Comme décrit plus haut, **les commanditaires d'évaluations peuvent recourir à des tiers variés pour réaliser les travaux** : selon les pays, universités, organismes de recherche indépendants, cabinets de conseil privés. Dans les pays où de nombreuses structures indépendantes existent (États-Unis, Royaume-Uni, Allemagne), l'externalisation est très largement pratiquée auprès d'acteurs multiples qui se réunissent en consortium pour répondre aux commandes. Ce modèle constitue pour ces pays une forme de garantie d'indépendance des évaluateurs vis-à-vis des commanditaires. En Suède et en France, l'externalisation est également pratiquée (par exemple en France via les comités d'évaluation pilotés par France Stratégie), mais une proportion importante d'évaluations d'impact sont également **réalisées « en propre » par l'administration**. Dans le cas de la Suède, il s'agit d'évaluations financées et mises en œuvre par les agences d'évaluation (telles que l'IFAU pour les politiques d'éducation et du marché du travail) ou les unités d'évaluations des agences exécutives sectorielles (comme Sida dans le domaine de l'aide au développement). Dans le cas français, l'Insee, les services statistiques ministériels mais aussi les établissements publics de l'État (Pôle emploi par exemple) peuvent être la fois commanditaires et producteurs d'évaluations d'impact.

4. Une faible structuration des milieux professionnels sur le champ de l'évaluation d'impact

Dans tous les pays de notre échantillon, il existe une **société professionnelle de l'évaluation**, c'est-à-dire une organisation regroupant les praticiens de la sphère administrative, académique et du privé, engagés dans l'évaluation. Si aux États-Unis, la société américaine d'évaluation ([American Evaluation Association, AEA](#)) a établi des principes directeurs pour les travaux d'évaluation et formulé des propositions pour renforcer la place de l'évaluation d'impact au sein du gouvernement fédéral, d'autres sociétés professionnelles sont moins impliquées dans la promotion des méthodes d'évaluation d'impact, comme c'est le cas de la [Société canadienne d'évaluation \(SCE\)](#), la [Société française d'évaluation \(SFE\)](#), ou encore la société d'évaluation du Royaume-Uni ([United Kingdom Evaluation Society, UKES](#)). L'Allemagne s'est pour sa part inspirée de l'AEA lorsqu'elle a créé en 1997 sa propre société d'évaluation ([Deutsche Gesellschaft für Evaluation, DeGEval](#)). Son rôle a été primordial dans le développement des standards d'évaluation pour les évaluateurs et les commanditaires d'évaluations.

D'autres acteurs internationaux sont spécialisés dans la **production de revues systématiques**, synthétisant par domaine les enseignements des évaluations d'impact disponibles. C'est le cas de [Cochrane](#), collaboration née en 1993 pour synthétiser les connaissances scientifiques sur les interventions médicales. Dans le domaine social, il existe également la [Campbell Collaboration](#), créée en 1999 et basée en Norvège, qui réalise des synthèses de la recherche en éducation, en criminalité, en protection sociale ou encore en développement international.

Au Royaume-Uni, l'[Alliance for Useful Evidence \(A4UE\)](#), fondée en 2011, promeut l'évaluation au sein d'un réseau de plus de 3 500 personnes grâce à des activités de recherche, de formation et de lobbying. En 2012, après sa création, l'A4UE a par exemple été au cœur de la campagne en faveur de la création d'un « *NICE for social policy* »¹⁰, proposition qui s'est matérialisée par le lancement du réseau des [What Works Centres](#). Indépendants du gouvernement, ces derniers synthétisent la recherche existante dans leur domaine de compétence et diffusent les résultats des évaluations d'impact auprès des professionnels et des décideurs publics. Chacun des What Works Centres est spécialisé sur un domaine ou un territoire et joue le rôle de « guichet unique » d'accès aux évaluations. Par leur visibilité, les What Works Centres sont un vecteur puissant de diffusion et de discussion des résultats de la recherche auprès des décideurs, des professionnels du domaine ainsi que du grand public.

¹⁰ Le [National Institute for Health and Care Excellence \(NICE\)](#) a été fondé en 1999 pour évaluer les médicaments et les pratiques médicales à partir de méthodes scientifiques rigoureuses et a été le premier centre du réseau des What Works Centres.

PARTIE II – L'OUVERTURE DU MILIEU ADMINISTRATIF À LA RECHERCHE

1. Des universitaires et chercheurs très présents dans l'administration aux États-Unis, au Royaume-Uni, en Allemagne et en Suède

Aux États-Unis, au Royaume-Uni, en Allemagne et en Suède, il est fréquent que des universitaires ou des chercheurs soient recrutés au sein de l'administration centrale pour des postes qui requièrent une expertise scientifique. Outre l'administration centrale, les chercheurs sont particulièrement présents au sein des What Works Centres au Royaume-Uni et des agences d'évaluation en Suède. Le recrutement dans la fonction publique s'opère par candidature classique en fonction du niveau d'expertise des candidats dans le domaine concerné. Un master ou un doctorat sont le plus souvent requis. Toutefois, le recrutement d'universitaires et de chercheurs venant d'organismes de recherche est rendu difficile par les salaires proposés et par la question de la progression de carrière (un passage dans l'administration n'étant pas toujours valorisé dans la suite des parcours professionnels). Ainsi, le choix entre une carrière académique et administrative se fait souvent en début de carrière. À l'instar du **programme de mobilité IPA** aux États-Unis (voir la **BONNE PRATIQUE N° 1**), des passerelles sont parfois organisées entre les milieux. L'exemple américain montre qu'elles sont cependant inégalement empruntées par les agences gouvernementales.

BONNE PRATIQUE N° 1

Le programme IPA aux États-Unis pour faciliter la présence de chercheurs dans l'administration

Au niveau fédéral, les agences ont la possibilité de recourir à un programme, **Intergovernmental Personnel Act (IPA) Mobility Program**, qui leur permet d'accueillir des universitaires ou des chercheurs pour une période déterminée (jusqu'à quatre ans) avec l'assurance pour ces derniers de pouvoir retourner dans leur université ou centre de recherche d'origine. De la même manière, les agences fédérales peuvent envoyer temporairement certains de leurs fonctionnaires dans des universités ou centres de recherche. Plus généralement, le programme IPA est destiné à favoriser la coopération entre le gouvernement fédéral et les entités non fédérales (les États et les collectivités locales, les universités, les centres de recherche, etc.) par l'échange temporaire de personnel qualifié. Selon les cas, les agents en mobilité peuvent continuer à être rémunérés par leur organisation d'origine ou bien être payés par l'entité qui les accueille. Ils conservent les droits et devoirs attachés à leur statut, et doivent retrouver à l'issue du programme un emploi équivalent à celui qu'ils occupaient avant leur mobilité (en termes de salaire et de grade). Par exemple, l'*Office of Evaluation Sciences* (OES) hébergé au sein du ministère *General Service Administration* (GSA) et spécialisé dans l'application des résultats des sciences comportementales aux programmes et dispositifs fédéraux, est composé d'une équipe interdisciplinaire de chercheurs issus notamment de l'université (*Harvard University, Reed College, Washington University*) et recrutés pour des périodes courtes (d'une durée allant de 1 à 4 ans) à l'aide du programme IPA.

À l'inverse, les universitaires et chercheurs sont peu présents en France et au Canada.

En France, les postes d'encadrement supérieur dans la fonction publique d'État sont occupés par les sortants des écoles nationales (École nationale d'administration, École polytechnique et Écoles normales supérieures) appartenant aux grands corps administratifs (Conseil d'État, Cour des Comptes, Inspection générale des finances, Inspection générale des affaires sociales, etc.) et techniques (ingénieurs des Mines, Ponts, eaux et forêts, administrateurs de l'Insee, corps de l'Armement). En particulier, les métiers de la statistique et de l'analyse économique dans l'administration sont occupés par les membres du corps de l'Insee. La présence des docteurs, enseignants-chercheurs et chercheurs dans l'administration publique est possible lorsqu'une expertise scientifique est nécessaire et n'est pas disponible parmi les corps de la fonction publique. Au Canada, la politique de recrutement de la fonction publique au niveau fédéral relève d'un dispositif centralisé piloté par la [Commission de la fonction publique](#), ce qui explique la faible présence de chercheurs au sein de l'administration. En outre, les compétences requises pour les agents publics au sein des unités d'évaluation obligatoirement présents dans chaque ministère privilégient des profils généralistes et une connaissance du secteur concerné à la maîtrise de compétences techniques et scientifiques.

2. Des compétences et des formations aux méthodes d'évaluation inégales au sein de l'administration dans la quasi-totalité des pays

La sensibilisation et la formation des hauts fonctionnaires est un autre levier de développement des pratiques. **Le Royaume-Uni est le seul des pays étudiés à avoir développé des standards de compétences prenant en compte les preuves scientifiques** (« *evidence* »). Utilisés à la fois pour le recrutement, pour l'évaluation individuelle des fonctionnaires et pour leur évolution professionnelle, le cadre général, appelé [Civil Service Competency Framework](#), définit trois groupes de compétences, dont le premier, « *setting direction* », inclut la capacité à utiliser les preuves dans la conception des politiques publiques. Il existe en outre des standards de compétences spécifiques à chacune des catégories de fonctionnaires (appelées « *professions* »). Le cadre le plus remarquable est celui de la *Policy profession* qui regroupe les fonctionnaires impliqués dans la conception des politiques publiques (18 420 personnes en 2017). Les [Policy Profession Standards](#) (voir la [BONNE PRATIQUE N° 2](#) page suivante), créés en 2010, requièrent en effet que les fonctionnaires maîtrisent certaines compétences techniques liées à l'évaluation.

D'autres pays ont cherché à définir et à encadrer les compétences liées à l'évaluation des politiques publiques. Aux États-Unis, la nouvelle loi, [Foundations for Evidence-Based Policymaking Act \(FEBPA\)](#) de 2018, prévoit que l'*Office of Personnel Management* (OPM), administration en charge des ressources humaines à la Maison blanche, en collaboration avec l'OMB, identifie les compétences nécessaires à l'exercice des activités d'évaluation de programmes, définit les professions correspondantes et les parcours professionnels associés au sein des agences gouvernementales. À l'initiative de la [société canadienne d'évaluation \(SCE\)](#), le Canada a défini pour sa part en 2005 un référentiel de compétences professionnelles pour l'activité d'évaluation et mis en place une procédure d'accréditation. Ce référentiel décrit les connaissances, les compétences et les aptitudes que les évaluateurs de programme doivent posséder pour produire des évaluations solides. Au-delà de principes généraux sur l'organisation et le processus évaluatif, il ne donne cependant pas d'orientation sur les méthodes à privilégier ni sur les niveaux de preuve attendus.

BONNE PRATIQUE N° 2 Les Policy Profession Standards au Royaume-Uni

Les standards de compétences de la **Policy profession**, créés en 2010, définissent 18 compétences, déclinées en trois niveaux selon le degré de responsabilité, qui ont notamment pour objet de renforcer l'utilisation des preuves (« *evidence* ») dans la conception des politiques publiques. En particulier, il existe une compétence intitulée « évaluation » qui recouvre notamment les éléments suivants :

- au niveau 1 : compréhension de l'importance de l'évaluation des politiques publiques, utilisation d'évaluations antérieures pour appréhender l'efficacité de la politique publique, connaissance de la différence entre recherche quantitative et qualitative, lectures de graphes, tableaux, statistiques descriptives, etc. ;
- au niveau 2 : utilisation des évaluations pour faire des recommandations de politique publique ; connaissance et pratique des méthodes communes d'évaluation, y compris les RCT ; maîtrise des concepts clés de l'évaluation (significativité statistique, intervalles de confiance, affectation aléatoire) ; capacité à critiquer une évaluation sur une base méthodologique ;
- au niveau 3 : capacité à identifier les pistes de recherche dans son domaine de spécialité ; connaissance et expérience pratique d'un panel de méthodes d'évaluation ; capacité à commanditer des évaluations appropriées et proportionnées et soutien à leur utilisation dans le processus de décision ; promotion des méthodes d'évaluation ; construction d'un réseau de chercheurs afin d'avoir accès à la recherche la plus récente.

La formation des acteurs administratifs en général aux méthodes d'évaluation d'impact des politiques publiques est peu développée dans tous les pays étudiés, à l'exception du Royaume-Uni en lien avec les standards de compétences définis. En France et au Canada où il existe des écoles de formation des hauts fonctionnaires, les formations proposées aux futurs agents publics ne couvrent pratiquement pas les enjeux de l'évaluation.

Plusieurs pays utilisent la formation continue pour sensibiliser leurs cadres dirigeants à l'évaluation. Le **Royaume-Uni est cependant le seul pays étudié à avoir déployé un dispositif complet de sensibilisation et de formation** afin de garantir que les compétences propres à chacune des professions et définies précisément par des standards soient maîtrisées. Si les ministères effectuent des programmes de formation en interne, le rôle de prestataires extérieurs comme l'*Alliance for Useful Evidence* (A4UE) ou la *What Works Team* est important. L'A4UE propose des actions de formation à l'utilisation des preuves aux fonctionnaires de la *Policy Profession*. La *What Works Team* propose des cours sur les méthodes d'évaluation (dont les RCT) aux fonctionnaires économistes (*Government Economic Service*, GES), aux chercheurs en sciences sociales (*Government Social Research service*, GSR) ou impliqués dans la conception des politiques publiques (*Policy profession*), ainsi qu'aux personnes participant au *Future Leaders Scheme* (FLS), formation continue proposée chaque année à quelques centaines de fonctionnaires « talentueux » de niveau chef de projets (Grade 7) ou supérieur.

Dans les autres pays (États-Unis, Suède, Allemagne, France et Canada), la formation aux méthodes d'évaluation d'impact s'adresse essentiellement aux agents publics spécialisés, en position de piloter ou de mener eux-mêmes des évaluations d'impact. En France, l'Insee propose à ses agents et à ceux des services statistiques ministériels des formations régulières (annuelles) aux méthodes d'évaluation d'impact. Quelques acteurs académiques proposent aussi des formations ponctuelles à ces méthodes pour les commanditaires et utilisateurs des évaluations. Aux États-Unis, [J-PAL North America](#), rattaché au MIT, est un acteur important de la formation aux méthodes expérimentales (RCT). Les principaux bureaux d'évaluation au niveau fédéral ont généralement dans leurs activités une mission d'assistance et d'appui au pilotage ou à la conduite d'évaluations auprès des autres agences/bureaux de leur ministère dans le but d'améliorer leur qualité. Par exemple, le [Chief Evaluation Office \(CEO\)](#) au ministère du travail (*Department of Labor*, DOL) accompagne les autres agences du ministère pour garantir la qualité des évaluations qu'elles financent en propre. En Allemagne, certains ministères, comme celui du Travail et des Affaires sociales (BMAS), proposent régulièrement des formations continues sur les méthodes d'évaluation des politiques publiques. Toutefois, c'est la [DeGEval](#), la société allemande d'évaluation, qui participe sans doute le plus à la professionnalisation des évaluateurs¹¹.

3. Des collaborations entre l'administration et la recherche souvent bien exploitées et parfois originales

En dépit de ces constats, on observe que **dans l'ensemble des pays, des collaborations entre l'administration et la recherche sur le champ de l'évaluation sont possibles et sont parfois bien exploitées.**

Les partenariats entre l'administration et la recherche peuvent prendre diverses formes, les plus communes étant le recours aux universitaires et chercheurs pour la réalisation des évaluations d'impact, la mise en place de groupes de travail techniques pour accompagner les équipes de chercheurs et l'organisation de conférences académiques pour débattre des résultats des évaluations d'impact. Par exemple, en Suède, les contacts directs avec des universitaires sont fréquents pour les agences d'évaluation, le plus souvent sous la forme de réseaux et de façon peu formalisée. L'IFAU est l'agence pour laquelle ces contacts sont les plus intenses, en raison de sa proximité avec l'université d'Uppsala qui permet à de nombreux personnels de travailler dans les deux institutions. L'IFAU dispose en outre d'un [conseil scientifique](#) composé de douze membres – majoritairement des universitaires issus de domaines variés (économie, sciences sociales, droit, éducation) – chargé de déterminer les recherches à financer. L'IFAU organise régulièrement des séminaires ouverts aux décideurs publics afin de les sensibiliser aux méthodes scientifiques d'évaluation.

Au-delà de ces partenariats conventionnels, l'administration a également mis en place ou favorisé des modalités plus originales de collaboration avec les chercheurs au travers notamment de **la création de *Nudge units*** chargées d'appliquer les apports des sciences comportementales aux politiques publiques. Ces unités apportent généralement leur soutien

¹¹ La DeGEval propose notamment des cours de formation et de perfectionnement, anime un réseau junior (*Nachwuchsnetzwerk*) et décerne chaque année le Prix du jeune chercheur (*DeGEval-Nachwuchspreis*), relaie les évaluations entre acteurs institutionnels et universitaires, ou encore héberge un forum d'échanges autour de l'évaluation.

aux administrations pour tester de nouvelles solutions et en évaluer rigoureusement les résultats à l'aide de méthodes expérimentales (RCT). On peut en citer plusieurs (voir la **BONNE PRATIQUE N° 3**, ci-dessous) :

- la **Behavioural Insights Team (BIT)** au Royaume-Uni créée en 2010 au sein du Cabinet Office avant de devenir indépendante en 2014 ;
- l'**Office of Evaluation Sciences (OES)** au sein de la General Services Administration (GSA) aux États-Unis, qui a remplacé le Social and Behavioral Sciences Team (SBST), un groupe intergouvernemental créé en 2014 dans l'administration de la Maison blanche.
- l'**Unité de l'impact et de l'innovation (UII)**, précédemment le Centre d'innovation, créée en 2015 et rattachée au Premier ministre au Canada ;

l'équipe en charge du développement des sciences comportementales appliquées aux politiques publiques à la **Direction interministérielle de la transformation publique (DITP)** rattachée au ministère de l'Action et des Comptes publics, en France.

BONNE PRATIQUE N° 3

Le développement de *Nudge units* dans les pays étudiés

Au Royaume-Uni, la création en 2010 de la **Behavioural Insights Team (BIT)** entendait répondre au besoin d'une meilleure prise en compte des recherches en sciences comportementales dans l'élaboration et la conduite des politiques publiques. Entre 2010 et 2014, la BIT a ainsi mené des expérimentations et produit des rapports conjoints avec les principaux ministères britanniques. Son rapport de 2012, *Test, Learn, Adapt*¹², a connu une large diffusion et permis de sensibiliser l'administration à la méthode des RCT. En 2014, la BIT est devenue indépendante, même si elle est encore largement financée par le gouvernement britannique. Elle conduit désormais son propre agenda d'évaluations et cherche à essaimer sa démarche au sein de l'administration centrale et régionale. Le ministère de l'Éducation a ainsi pour projet de créer sa propre Behavioural Insights Unit. Il existe également un bureau régional de la BIT à Manchester, en partenariat avec les autorités régionales¹³.

Aux États-Unis, l'**Office of Evaluation Sciences (OES)** au sein du ministère *General Services Administration (GSA)*, composé d'une équipe d'une quinzaine de chercheurs appliqués issus de disciplines diverses (statistiques, économie, psychologie, santé publique, etc.), apporte son soutien aux différentes agences fédérales pour tester des nouvelles solutions reposant sur l'application des sciences comportementales aux dispositifs fédéraux, et en évaluer rigoureusement les résultats à l'aide de méthodes expérimentales (RCT). L'OES réalise environ une quinzaine d'expérimentations aléatoires contrôlées chaque année, principalement à l'aide de données administratives. Depuis 2015, il en a réalisé plus de soixante pour une douzaine d'agences fédérales.

Au Canada, l'**Unité de l'impact et de l'innovation (UII)**, rattachée au Premier ministre, a été créée en 2015. Cette unité vise à mieux prendre en compte les apports de la recherche, notamment dans le domaine de l'économie comportementale, pour la conception des politiques publiques. En particulier, l'Initiative Impact Canada déployée par l'UII s'appuie sur les RCT pour évaluer les effets des expérimentations sur le modèle de la Behavioural Insights Unit (BIU) qui a été la première unité de ce type développé au Canada par le gouvernement de l'Ontario.

¹² Haynes, Service, Goldacre & Torgerson, BIT, 2012.

¹³ Behavioural Insights Team, 2017.

En France, la Direction interministérielle de la transformation publique (DITP), rattachée au ministère de l'Action et des Comptes publics, a remplacé en 2017 le Secrétariat général pour la modernisation de l'action publique (SGMAP). Cette direction ne porte pas de programme d'évaluation de politiques publiques en tant que tel mais met à disposition des ministères volontaires des marchés d'étude et d'expertise spécifiques (consultants, chercheurs, etc.) et promeut, via des appels à projets, de nouvelles méthodes au sein des administrations : expérimentations, apport des sciences comportementales dans les politiques publiques. En 2018, la DITP a ainsi confié à plusieurs équipes de recherche (dont le CEPREMAP et la Behavioural Insight Team britannique) l'accompagnement et l'évaluation de sept projets d'application des sciences comportementales aux politiques publiques proposés par des administrations.

Une autre pratique exemplaire offrant des points de contact entre chercheurs et administrations est la création en 2015 au Royaume-Uni du **Cross-Government Trial Advice Panel** (CG-TAP), composé d'experts de l'administration et du monde académique. Le CG-TAP forme et aide techniquement les ministères souhaitant mettre en œuvre des évaluations d'impact randomisées. Une vingtaine de ministères et d'agences publiques y ont eu recours depuis sa création (voir la **BONNE PRATIQUE N° 4**).

BONNE PRATIQUE N° 4

Le Cross-Government Trial Advice Panel au Royaume-Uni

Depuis 2015, la What Works Team coordonne le Cross-Government Trial Advice Panel (CG-TAP), composé d'experts de l'administration (18) et du monde académique (28). Le CG-TAP a pour objectif d'inciter et d'aider les ministères à mettre en œuvre des évaluations d'impact de leurs programmes, en se fondant notamment sur les enseignements du rapport *Test, Learn, Adapt* de la Behavioural Insights Team. Il doit permettre de rendre plus accessible aux fonctionnaires la conception et la mise en œuvre d'une évaluation d'impact. S'il encourage l'utilisation des méthodes d'évaluation aléatoire (RCT), le panel propose aussi des solutions quasi-expérimentales (mise en œuvre différée, régression sur discontinuité, double différence, score de propension). Il ne pilote pas lui-même des évaluations d'impact afin de conserver son rôle de conseil et de promouvoir la formation directe des agents des ministères responsables des politiques publiques. Depuis sa création, le CG-TAP a contribué à 52 projets d'évaluation au sein de 18 administrations. Le ministère du Logement, des Communautés et du Gouvernement local (MHCLG) a par exemple bénéficié de l'expertise de ce panel pour la réalisation en mars 2018 d'une RCT permettant d'évaluer l'impact sur l'intégration sociale d'un programme d'enseignement de l'anglais.

Enfin, la France a développé la pratique des comités scientifiques et des commissions d'évaluation chargés de piloter l'évaluation de grands dispositifs nationaux. La Direction de l'animation de la recherche et des études statistiques (Dares) du ministère du Travail ou France Stratégie, service du Premier ministre, ont ainsi mis en place de tels comités dont la composition, variable selon les instances, inclut systématiquement des universitaires et des chercheurs parmi leurs membres. Par exemple, le **comité d'évaluation du Plan d'investissement dans les compétences (PIC)**, dont le secrétariat est assuré par la Dares, est composé exclusivement d'universitaires et de chercheurs issus de diverses disciplines (économie, sciences cognitives, sciences de l'éducation, sciences politiques, sociologie).

PARTIE III – LES DISPOSITIFS FORMELS FAVORISANT LA CONDUITE ET L'UTILISATION DES ÉVALUATIONS D'IMPACT

1. Un cadre structuré et coordonné au Canada et aux États-Unis, plus souple dans les autres pays

Pour les pays de notre échantillon, la pratique de l'évaluation s'inscrit dans un cadre plus ou moins structuré et coordonné. D'un côté, le Canada et, dans une moindre mesure, les États-Unis ont développé un cadre très formalisé comprenant des instances de coordination et des outils de programmation. D'un autre côté, le Royaume-Uni, la Suède, l'Allemagne et la France sont des pays où les structures chargées d'évaluer les politiques publiques opèrent de manière largement indépendante les unes des autres.

Le Canada est sans doute l'exemple du pays ayant poussé le plus loin cette structuration institutionnelle de l'évaluation. Bien qu'il ne s'agisse pas d'évaluation d'impact, cette organisation mérite d'être soulignée en raison de son caractère systématique. En effet, **l'évaluation fait l'objet d'un cadre formalisé et coordonné par le Secrétariat du Conseil du Trésor (SCT)**. Le cadre réglementaire mis en place à partir d'avril 2001 et actualisé en 2009 et 2011¹⁴ impose ainsi à chaque ministère et organisme public d'inclure l'évaluation dans ses pratiques de gestion. Il s'accompagne d'une directive et de normes de qualité spécifiques¹⁵. Cette politique a pour but de favoriser la production d'informations précises, objectives et étayées pour aider les gestionnaires à prendre plus efficacement des décisions éclairées sur leurs politiques et pour rendre des comptes aux citoyens canadiens.

Ce cadre a donné lieu à la création, dans la plupart des ministères, d'**unités d'évaluation** et à la mise en place de **comités de suivi des évaluations** (présidés par un directeur du ministère). Chaque ministère est chargé de formaliser un plan d'évaluation ministériel organisant la production sur les années à venir (aujourd'hui ces plans sont quinquennaux) par son unité d'évaluation (généralement un bureau « audit et évaluation »). Les évaluations de programmes sont réalisées en interne par environ 450 ETP formant le réseau des « évaluateurs de programme » au sein de l'administration fédérale canadienne. **Au sein du SCT, le Centre d'excellence en évaluation (CEE), équipe spécialisée en évaluation, joue un rôle clé pour coordonner les administrations et les accompagner dans cette fonction évaluative.** Le CEE établit chaque année un rapport sur l'activité des unités d'évaluation et la qualité des évaluations produites. Il veille également à la publication des rapports d'évaluation et s'assure de leur transmission au Parlement.

On retrouve également **un dispositif institutionnel de l'évaluation assez avancé aux États-Unis où l'Office of Management and Budget (OMB), administration en charge de la préparation du budget, joue un rôle central** pour impulser des évaluations d'impact au sein du gouvernement et pour encourager les ministères à davantage utiliser les preuves dans les décisions politiques. Les efforts de l'OMB incluent une série de

¹⁴ SCT, [Politique sur l'évaluation](https://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=31306), 2009. Pour un historique de cette politique voir <https://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=31306>.

¹⁵ SCT, [Directive sur la fonction d'évaluation](#), 2009.

directives¹⁶ à destination des ministères les encourageant à faire davantage d'évaluations de programmes rigoureuses, et à conditionner davantage l'octroi de subventions à l'atteinte des résultats.

En particulier, en 2015, l'OMB a mis en place une instance de coordination sur l'évaluation, appelée **Interagency Council on Evaluation Policy (ICEP)** afin d'aider les agences fédérales à adopter des pratiques d'évaluation communes, sur le modèle de l'instance réunissant les principales agences statistiques du gouvernement fédéral (la Federal Interagency Council on Statistical Policy). L'ICEP vise à améliorer la programmation, la conduite et la diffusion des évaluations au sein du gouvernement fédéral. Par ailleurs, **certains ministères sont dotés d'unités d'études et d'évaluation particulièrement développées ayant un rôle avancé de pilotage et de coordination** de ces activités au sein du ministère. Ainsi, le ministère du Travail dispose d'un chef évaluateur (**Chief Evaluation Officer, CEO**, Department of Labor), dont le bureau créé en 2010 est localisé dans l'Office de l'Assistant Secretary for Policy. Cette fonction existe également dans d'autres ministères sous d'autres appellations (voir la **BONNE PRATIQUE N° 5**).

Enfin, **certains ministères et agences fédérales ont défini une politique spécifique pour leurs activités d'évaluation**¹⁷. Cette politique définit plusieurs principes permettant de garantir que les évaluations menées par le gouvernement fédéral respectent les standards scientifiques (principe de rigueur), répondent aux besoins des parties prenantes (principe de pertinence), sont objectives (principe d'indépendance), sont publiques (principe de transparence) et protègent la confidentialité et la vie privée des individus (principe d'éthique).

Certains ont également élaboré des Learning Agendas^{18,19} qui recensent et hiérarchisent les questions politiques et de recherche les plus importantes. Ils permettent de mettre en avant les sujets sur lesquels les connaissances scientifiques sont lacunaires et de coordonner les travaux entre les différents services compétents.

Cette structuration de l'évaluation déjà bien avancée aux États-Unis va se renforcer avec la mise en application des dispositions de la loi de 2018, **Foundations for Evidence-Based Policymaking Act (FEBPA)**, qui prévoit pour chaque agence fédérale la nomination d'un chef évaluateur (*Evaluation Officer*), l'élaboration d'une politique d'évaluation et la mise en place de Learning Agendas.

¹⁶ OMB (2009, 2012, 2013) : M-10-01 ; M-12-14 ; M-13-17.

¹⁷ Voir par exemple [US Department of Labor Evaluation Policy, 2013](#).

¹⁸ Nightingale, Fudge et Schupmann, Evidence-Based Policy making Collaborative, 2018.

¹⁹ Nightingale et Scott, 2018.

BONNE PRATIQUE N° 5

Les chefs évaluateurs pour animer, piloter et coordonner les évaluations dans les ministères aux États-Unis

Le ministère du Travail américain (*Department of Labor*) dispose depuis 2010 d'un chef évaluateur implanté au sein de l'Office de l'*Assistant Secretary for Policy* et qui a plusieurs missions au sein du ministère :

- concevoir, financer et superviser la mise en œuvre d'un portefeuille de projets d'évaluation sur les programmes et les politiques du ministère, qui complètent les activités d'évaluation des bureaux et des agences du ministère (le CEO gère une quarantaine d'évaluations à tout moment dont une quinzaine d'évaluations d'impact) ;
- accroître la qualité des évaluations et la connaissance des méthodes et standards de l'évaluation (par une assistance technique à la réalisation et au pilotage des évaluations au sein du Ministère) ;
- améliorer l'utilisation et la diffusion des résultats des évaluations ;
- améliorer l'accès aux données et leur utilisation.

Cette fonction existe depuis plusieurs années dans d'autres administrations américaines sous différentes appellations²⁰. Par exemple le directeur de l'*Institute of Education Sciences* (IES) joue le rôle de chef évaluateur pour le *Department of Education*, de même que la directrice de l'*Office of Planning, Research and Evaluation* du *Department of Health and Human Services*.

La loi *Foundations for Evidence-Based Policymaking Act* (FEBPA) adoptée en 2018 impose la nomination dans chaque agence fédérale d'un chef évaluateur chargé notamment, pour les agences encore dépourvue de ce type de fonctions, de garantir la cohérence, la qualité et l'indépendance des évaluations de l'agence ; de déterminer les capacités de l'agence pour développer et utiliser les évaluations ; de définir et de mettre en œuvre une politique d'évaluation pour l'agence et enfin, de coordonner, développer et mettre en place une programmation des évaluations.

Du fait de sa tradition juridique, **au Royaume-Uni, l'évaluation reste une pratique souple et peu coordonnée bien que des lignes directrices relatives à la qualité des évaluations soient petit à petit venues structurer les modes de faire.** Chaque ministère gère en effet son propre agenda d'évaluations et les dix What Works Centres fonctionnent de manière indépendante. Néanmoins, **la What Works Team joue un rôle d'intermédiaire entre les centres thématiques et les ministères.** Cette équipe a pour rôle de partager au sein de l'administration centrale les résultats produits et rassemblés par les What Works Centres, et de former les fonctionnaires aux méthodes d'évaluation d'impact (*voir partie II*). Quelques ministères se sont dotés de programmes d'évaluation de leurs politiques publiques, mais ces programmes, non obligatoires, ne sont pas systématiquement publiés. En outre, depuis 2015, la What Works Team encourage les ministères à publier des **Areas of Research Interest (ARI)** à destination des chercheurs,

²⁰ Schupmann et Fudge, 2018.

qui indiquent les domaines où des besoins de preuves scientifiques existent. Tous les ministères se sont depuis prêtés à l'exercice et ont publié une ARI en 2017 ou 2018.

En Suède, comme en Allemagne, la pratique de l'évaluation est généralisée mais peu coordonnée. En Suède, elle est réalisée en grande partie par les **agences spécialisées dans l'évaluation des politiques publiques** (35 selon l'OCDE en 2013²¹), créées pour la plupart depuis 2006, et en complément par les unités d'évaluation dans les agences sectorielles chargées de la mise en œuvre des politiques publiques. Chaque agence d'évaluation choisit de manière autonome ses projets d'évaluations. Il n'existe pas non plus de programmation systématique des évaluations, chaque agence pouvant élaborer un programme d'évaluations sans qu'il y ait d'obligation formelle.

En comparaison, **l'évaluation d'impact est encore moins développée en France** qu'au Royaume-Uni, en Allemagne ou en Suède. **Il n'existe pas non plus d'instance de coordination des évaluations au niveau national**, à l'exception toutefois des services statistiques ministériels opérant une centralisation des besoins en information sur les politiques de leur ministère. Chaque acteur procède à la programmation de ses propres travaux. Au niveau interministériel, si des tentatives ont été conduites dans les années passées pour coordonner les évaluations, elles n'ont pas perduré et ne portaient pas spécifiquement sur les évaluations d'impact²².

2. Des corpus et standards méthodologiques favorisant la montée en qualité, surtout aux États-Unis, au Royaume-Uni et en Allemagne

Les bonnes pratiques identifiées sur le plan des standards méthodologiques sont mises en œuvre majoritairement au Royaume-Uni et aux États-Unis, et aussi mais de manière plus ponctuelle en Allemagne. Ainsi, il n'existe pas en France, en Suède ou au Canada de standards de qualité relatifs aux méthodes d'évaluations d'impact formalisés (guide méthodologique, lignes directrices ou standards de preuves) et faisant référence parmi les commanditaires ou utilisateurs d'évaluations. Au Canada, le guide édité en 2004 par le Centre d'excellence en évaluation (CEE, Secrétariat du Conseil du Trésor) reste au niveau des principes et du processus évaluatif et ne livre aucune indication sur les méthodes en tant que telles ni sur les niveaux de preuve attendus²³.

Plusieurs pays ont mis en place des **procédures de contrôle et de revue de qualité des évaluations**, en faisant appel à des experts indépendants, mais ces dernières sont plus ou moins répandues, avancées et formalisées dans les administrations. Elles consistent très souvent en l'organisation de groupes de travail techniques, chargés d'accompagner les équipes de recherche sélectionnées pendant la conduite des évaluations. Aux États-Unis, un exemple remarquable est la procédure mise en place par l'*Institute of Education Sciences* (IES, *Department of Education*) qui consiste en l'organisation d'une procédure de revue par les pairs. Obligatoire avant la publication, elle prévoit l'examen critique du

²¹ OCDE, 2013, p. 66.

²² Ces tentatives d'institutionnalisation de l'évaluation incluent le Conseil interministériel de l'évaluation et son conseil scientifique (CIME et CSE, 1990-1998), le Conseil national de l'évaluation (CNE, 1998-2002), le Secrétariat d'État au numérique et à l'évaluation des politiques publiques et les Comités interministériels pour la modernisation de l'action publique (CIMAP, 2012-2014) adossés au programme des évaluations de la MAP.

²³ SCT/CEE, 2004

rapport d'évaluation par au moins deux rapporteurs recrutés par l'IES en fonction de leurs compétences académiques et de leur expérience (publication dans les meilleures revues scientifiques, utilisation de méthodologies similaires, réalisation d'études similaires)²⁴.

Si dans chaque pays on retrouve **des guides et des standards méthodologiques**, les publications éditées par le **Trésor britannique** offrent la collection la plus complète de méthodologies d'évaluation à destination des administrations (voir la **BONNE PRATIQUE N° 6**). **Aux États-Unis, la récente loi FEBPA de 2018 prévoit l'élaboration de lignes directrices sur l'évaluation par l'OMB en collaboration avec l'ICEP** pour aider les agences fédérales à mener leurs activités d'évaluation.

BONNE PRATIQUE N° 6

Les guides méthodologiques édités par le Trésor britannique

Depuis plus de quinze ans, le Trésor britannique est à l'initiative d'une collection de guides et de documents méthodologiques de référence en matière d'évaluation. Ces derniers sont diffusés dans toute l'administration centrale britannique : ils sont mentionnés dans les programmes d'évaluation des ministères qui en possèdent ainsi que dans les standards de compétences des fonctionnaires. Ils sont donc supposés servir de référence lors de la mise en œuvre interne ou de la commande externe d'évaluations d'impact par l'administration publique, même si leur utilisation peut être variable selon les administrations. Ils couvrent l'ensemble des dimensions de l'évaluation, de la mise en œuvre à l'impact :

- ***The Green Book. Appraisal and Evaluation in Central Government***, publié en 2003 dans sa première version, actualisé en 2011, puis publié dans une deuxième version en 2018, présente la démarche d'évaluation d'un dispositif dans son ensemble à travers un « cycle de vie » des politiques publiques : logique, objectifs, évaluation *ex ante*, pilotage de la mise en œuvre, évaluation *ex post*, feedback. Il comprend des annexes sur les notions d'efficacité et d'équité, les droits de propriété intellectuelle, l'évaluation socioéconomique, la gestion du risque, les impacts redistributifs et le taux d'actualisation. La version 2018 de ce *Green Book* intègre désormais une définition de l'évaluation englobant de nombreux types d'évaluation (évaluation de mise en œuvre, évaluation d'impact, évaluation socioéconomique).
- ***The Magenta Book. Guidance for evaluation***, publié en 2003 et actualisé en 2011, est un guide pratique sur les différents types d'évaluation *ex post* : évaluation de mise en œuvre, évaluation d'impact, analyse coût-bénéfice. Une première partie, non technique, est destinée aux décideurs politiques. Elle présente les enjeux principaux, les différents types d'évaluation, les éléments clés qui doivent être intégrés à un dispositif pour permettre son évaluation, les difficultés pratiques. Une seconde partie, destinée aux évaluateurs, détaille les différentes étapes d'une évaluation (planification, collecte des données, identification des impacts, présentation et diffusion des résultats).
- ***Quality in policy impact evaluation***, publié en 2012, est un complément du *Magenta Book* qui présente de manière pédagogique les méthodes d'évaluation causales, reposant sur un contrefactuel. Il détaille cinq catégories d'évaluation classées selon leur robustesse, avec leurs forces et leurs faiblesses, et des exemples concrets d'évaluations de politiques publiques.

²⁴ Pour plus de détails sur la procédure de revue par les pairs, voir [le site de l'IES](#).

La diffusion de standards de preuve (« *standards of evidence* ») a également été, dans les pays les plus avancés, comme les États-Unis et le Royaume-Uni, au cœur d'une réflexion sur l'amélioration de la qualité des évaluations. Ces standards de preuve permettent de déterminer la qualité méthodologique des évaluations en fonction de la validité des méthodes mobilisées. Les plus utilisés sont les « **standards of evidence** » **édités par Nesta** depuis 2013 au Royaume-Uni²⁵ et **l'échelle de Maryland** (*Maryland Scientific Methods Scale*)²⁶, qui notent les évaluations d'impact en fonction de la crédibilité du groupe témoin et de la mobilisation d'approches expérimentales (évaluations avec assignation aléatoire) ou quasi-expérimentales. Ces standards peuvent être utilisés pour s'assurer de la robustesse des résultats produits mais, dans la pratique, ils sont le plus souvent mobilisés pour la réalisation de revues systématiques (synthèse des enseignements tirés des évaluations répondant à un standard de qualité élevé) ou pour construire des bases de données de résultats issus des évaluations, appelés Clearinghouses (*voir partie IV*). En pratique, les organisations, gouvernementales ou indépendantes, ont tendance à développer leurs propres standards de preuves, ce qui conduit à une multiplication et à une hétérogénéité des outils. Au Royaume-Uni, l'Alliance for Useful Evidence a ainsi recensé dix-huit standards de preuves différents en mai 2018²⁷, et a appelé à une unification de ces échelles. Aux États-Unis, les standards développés par **Mathematica Policy Research** pour le compte du ministère de l'Éducation font figure de référence et ont servi à l'élaboration de plusieurs Clearinghouses notamment pour les ministères du Travail, de l'Éducation et de la Santé. Cette démarche peut cependant parfois rencontrer des difficultés en raison de la résistance forte de certains acteurs aux méthodes purement quantitatives (en particulier les RCT). Le recours aux standards de preuve ne se retrouve pas dans les autres pays. Le Canada se prête ainsi davantage, pour apprécier la qualité de ses évaluations, à des exercices récurrents de **méta-évaluation**, c'est-à-dire d'audits de la qualité des travaux d'évaluation selon des critères de qualité prenant en compte l'ensemble des dimensions (robustesse des méthodes, transparence du processus, publication des résultats, degré d'utilisation des résultats, etc.).

BONNE PRATIQUE N° 7

Les standards de preuves aux États-Unis

Aux États-Unis, les agences fédérales s'appuient notamment sur des standards de preuve pour garantir la qualité de leurs évaluations. Ils permettent de déterminer la qualité méthodologique des évaluations qu'elles financent et plus généralement des évaluations disponibles. Ils sont également utilisés pour construire des bases de données, appelées les Clearinghouses, qui, pour les plus élaborées d'entre elles, établissent un classement des évaluations selon des critères de qualité méthodologique, de résultats et de coûts. Les standards développés par Mathematica Policy Research pour le compte du ministère de l'éducation font figure de référence. Ils

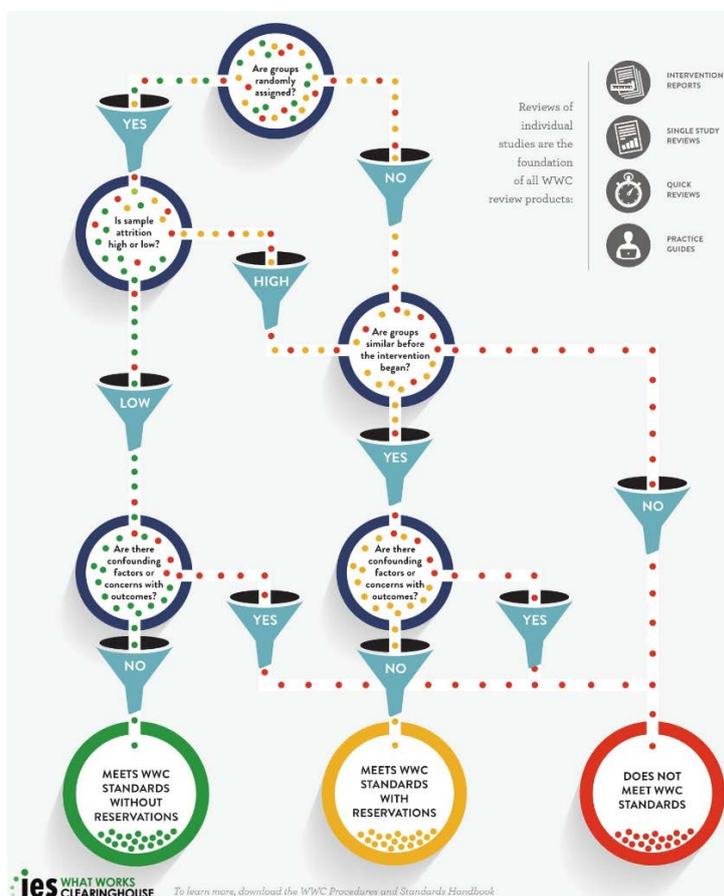
²⁵ Ces standards permettent de hiérarchiser les différentes méthodes d'évaluation selon leur niveau de rigueur et de validité, interne et externe, de la simple description d'un modèle logique argumenté (niveau 1) à la démonstration rigoureuse d'un impact positif d'une évaluation de réplication (niveau 5).

²⁶ L'échelle de Maryland propose une cotation de 1 à 5 des évaluations d'impact en fonction de la crédibilité du groupe témoin ; selon cette échelle, les approches expérimentales (évaluations avec assignation aléatoire) ou quasi expérimentales exploitant des données fiables sont considérées comme les plus robustes.

²⁷ Puttick, A4UE, 2018.

ont servi à l'élaboration de plusieurs Clearinghouses, dont le What Works Clearinghouse (WWC) du ministère de l'Éducation (voir figure 1) ou la Clearinghouse for Labor Evaluation and Research (CLEAR) du ministère du Travail. Ces standards varient généralement avec les agences.

Figure 1 – Standards de preuve du What Works Clearinghouse aux États-Unis



Note : seules les études expérimentales et quasi expérimentales sont considérées. Elles doivent cependant respecter certaines conditions (sur l'attrition, la similarité des populations des groupes test et de comparaison et les facteurs confondants) pour faire partie de la base.
Source : WWC, How the WWC rates a study ?

Enfin, une autre pratique repérée dans le benchmark est la mise en place, au Canada, d'un titre d'évaluateur accrédité²⁸. Cette initiative singulière – que l'on ne retrouve dans aucun des autres pays analysés – a été développée en 2009 par la Société canadienne d'évaluation. Elle relève cependant davantage d'une démarche de certification professionnelle que d'un processus scientifique : il s'agit pour les évaluateurs accrédités selon ce dispositif de se prévaloir du respect de principes déontologiques et de normes professionnelles ainsi que d'un ensemble de 36 compétences théoriques et pratiques. En 2018, le Canada comptait 395 évaluateurs accrédités (voir la **BONNE PRATIQUE N° 8**).

²⁸ Rebaptisé titre d'évaluateur qualifié (EQ) depuis 2019.

BONNE PRATIQUE N° 8 Le titre d'évaluateur accrédité au Canada

Pour encourager la professionnalisation de la pratique de l'évaluation, la Société canadienne d'évaluation (SCE) s'est engagée depuis 2009 dans la mise en place d'un titre d'**évaluateur accrédité** (ÉA). Ce programme des titres professionnels est construit sur trois piliers : un code de déontologie, des normes professionnelles et un ensemble de 36 compétences théoriques et pratiques (connaissance des concepts de base de l'évaluation, maîtrise des méthodes de collecte et de traitement des données, compétences en gestion de projet, qualités interpersonnelles, etc.)²⁹. Les compétences attendues ne mettent pas l'accent sur une discipline en particulier mais sur les dimensions professionnelles générales ; le titre n'est délibérément pas prescriptif du point de vue des méthodes et des techniques utilisées (quantitatives, qualitatives) afin d'encourager la pluridisciplinarité des évaluations.

La procédure d'accréditation se déroule en deux étapes : la soumission par le postulant d'un dossier par lequel il atteste de l'ensemble des aptitudes et compétences requises pour le titre puis la sélection par un jury de professionnels disposant chacun de 25 années d'expérience dans le domaine de l'évaluation. Ce titre d'évaluateur accrédité a fait l'objet en 2015 d'un **bilan intermédiaire** spécifique conduit par une équipe indépendante de la *Claremont Graduate University* qui a analysé la connaissance du titre, les leviers et les obstacles à sa mise en œuvre et ses effets sur les pratiques professionnelles. Sur la base de ce référentiel, le Canada compte, en 2018, 395 évaluateurs accrédités dont la liste est disponible sur demande à la SCE ; les chercheurs en économie y sont minoritaires.

Un cadre juridique non contraignant dans tous les pays, à l'exception du Canada

L'évaluation est inscrite dans la Constitution en Suède et en France. En Suède, elle concerne le pouvoir exécutif depuis 2007³⁰ et le Parlement depuis 2011³¹. En France, le Parlement s'est vu ajouter la mission d'évaluer les politiques publiques par la révision constitutionnelle de 2008 (article 24)³².

Au Canada, la directive sur la fonction d'évaluation de 2001 (dont les dispositions ont été actualisées en 2009 et 2011) du Secrétariat du Conseil du Trésor (SCT) a rendu obligatoire l'évaluation pour tous les ministères et agences du gouvernement et a donné lieu à la création, dans la plupart des ministères, d'unités d'évaluation et à l'adoption de plans d'évaluation quinquennaux (voir plus haut).

Dans la plupart des pays, par ailleurs, les textes législatifs portant de nouveaux programmes peuvent explicitement inclure un principe d'évaluation dont les dispositions sont plus ou moins contraignantes. Le programme phare du gouvernement Blair pour l'emploi, *Pathways to Work*, a par exemple fait l'objet d'une telle disposition d'évaluation. De même, le Crédit d'impôt pour la compétitivité et l'emploi (CICE)³³ ou la loi

²⁹ Voir : <https://evaluationcanada.ca/fr/competences-pour-les-evaluateurs-canadiens>

³⁰ *Ordinance on impact analysis of Regulation*

³¹ *Instrument of Government*, chap. 4, art. 8

³² *Constitution*, article 24

³³ *Loi n°2012-1510 du 29 décembre 2012 de finances rectificatives pour 2012*, article 66

Pacte³⁴ en France comprenaient un article instituant la mise en place d'un comité d'évaluation *ex post*. Dans certains cas, les lois organisant ces évaluations peuvent s'avérer très prescriptives sur le plan des méthodes. Toutefois, l'inscription de l'évaluation dans la loi instituant les dispositifs ne fait l'objet d'aucune obligation systématique dans les pays étudiés, mais elle y est souvent considérée comme une bonne pratique.

Aux États-Unis, au Royaume-Uni et en Suède, la demande sociale d'évaluations, émanant notamment du Parlement, des partenaires sociaux, des citoyens et des médias, est forte et conduit les ministères ou les agences gouvernementales à évaluer ou à faire évaluer dans une large mesure les programmes dont les enjeux économiques, sociaux, budgétaires ou environnementaux sont importants. En France, la demande sociale d'évaluations demeure au contraire faible. Face au poids de la production statistique dans leurs activités et dans un contexte budgétaire tendu, les services statistiques ministériels (SSM), ayant aussi dans leurs missions d'évaluer les dispositifs portés par leur ministère, opèrent parfois des arbitrages défavorables à l'évaluation.

Des financements provenant principalement des crédits des ministères

Dans tous les pays étudiés, le principal financeur d'évaluations d'impact est le gouvernement. Les travaux d'évaluation sont d'abord financés sur les crédits propres des administrations qui décident elles-mêmes de la part consacrée aux évaluations. Le financement des évaluations n'est donc pas garanti d'une année sur l'autre, et dépend à la fois de la dotation globale allouée annuellement aux administrations concernées (résultant d'arbitrages budgétaires) et du choix laissé à leur discrétion de répartir cette enveloppe globale entre leurs différentes missions.

Dans quelques rares cas de programmes ou de dispositifs publics, il peut exister des fonds réservés aux évaluations, lorsque par exemple une part de leur coût y est spécifiquement dédiée. Aux États-Unis, la loi *Elementary and Secondary Educational Act* (ESEA) de 1965, remplacée par la loi *Every Student Succeeds Act* (ESSA) de 2015³⁵, autorise ainsi le ministère de l'Éducation à financer les évaluations de ce programme par des fonds représentant jusqu'à 0,5 % des crédits alloués au programme ([IES Evaluation Plan](#)). Pour 2018, ces fonds représentaient environ 19,5 millions de dollars. En France, le Programme des investissements d'avenir (PIA) lancé en 2010 prévoyait aussi qu'une part des 47 milliards d'euros alloués au programme soit réservée à l'évaluation des actions par les opérateurs (BPI, Caisse des dépôts et consignations, Ademe, ANR, etc.), sous la forme d'un pourcentage ou d'un montant maximum par rapport au montant total de l'action financée détaillée dans la convention signée entre l'État et l'opérateur. De même, toujours en France, il est prévu que les études et les recherches relatives au Plan d'investissement dans les compétences (PIC) soient financées par un fonds géré par la Dares, sous l'autorité du Haut Commissaire aux compétences et à l'inclusion dans l'emploi.

Dans plusieurs pays, des réflexions ont lieu sur la mise en place d'une disposition systématique réservant une partie du budget des programmes majeurs à leur évaluation. Aux États-Unis, par exemple, la *Commission on Evidence-Based Policymaking*

³⁴ Loi n°2017-1837 du 30 décembre 2017 de finances pour 2018, article 28

³⁵ <https://www.ed.gov/essa?src=rn>

(CEP) avait recommandé dans un rapport³⁶ publié en septembre 2017, pour les nouveaux programmes définis par la loi ou pour les programmes reconduits, que le Congrès et la présidence des États-Unis permettent l'utilisation de 1 % au maximum des ressources dévolues au programme pour la collecte des données, la recherche et l'évaluation. Cette recommandation n'a pas été retenue par la loi FEBPA de 2018.

Beaucoup plus rarement, certaines administrations bénéficient de fonds réservés pour conduire globalement leurs activités d'évaluation. Aux États-Unis, depuis 2012, le [Chief Evaluation Office](#) (*Department of Labor*) se voit octroyer des financements qui représentent jusqu'à 0,75 % des crédits affectés à certaines agences opérationnelles du ministère « *evaluation set-asides* ». Le CEO disposait ainsi en 2018 d'un budget de 21,4 millions de dollars (dont 8 millions de crédits directs et la somme restante correspondant aux crédits transférés des autres agences du ministère)³⁷. Il gère un portefeuille d'une quarantaine d'études dont une dizaine d'évaluations d'impact.

Aux États-Unis et au Royaume-Uni, on observe par ailleurs que le financement des évaluations est également assuré par des fondations privées, soit sur la base de financements pérennes attribués à des structures spécialisées, sur le modèle du [Big Lottery Fund](#) au Royaume-Uni qui soutient les What Works Centres, soit sur la base de commandes d'évaluation, comme la [Laura and John Arnold Foundation \(LJAF\)](#), la [Mac Arthur Foundation](#) ou encore [The Pew Charitable Trusts](#) qui financent de nombreuses évaluations expérimentales dans leurs domaines d'activité.

En Allemagne, la *Leibniz Gemeinschaft*, qui regroupe près de 90 centres de recherches scientifiques, économiques et sociales, est un financeur important d'évaluations, doté d'un budget de 1,73 milliard d'euros en 2015 (toutes activités confondues).

Au Canada enfin, les financements dédiés aux évaluations de programme (dans leur ensemble) sont sanctuarisés. Selon les années, ces montants peuvent varier mais ils se situent en moyenne, depuis 2009 et pour le niveau fédéral, autour de 60 millions de dollars canadiens, soit 40 millions d'euros³⁸. Déterminés par le ministère dans le cadre de son dialogue de gestion avec le secrétariat du Conseil du Trésor, ces montants correspondent pour l'essentiel aux salaires versés aux équipes d'évaluation. L'enveloppe est globale et ne flèche pas les crédits vers un type de méthodes d'évaluation plutôt qu'un autre.

³⁶ Commission on Evidence-Based Policymaking, 2017.

³⁷ [Results for America, 2018](#).

³⁸ SCT/CEE, 2009.

PARTIE IV – LE DEGRÉ DE DIFFUSION ET D'INFLUENCE DES ÉVALUATIONS

1. Un principe général de publication dans tous les pays, mais de portée inégale

La **publication des travaux d'évaluation d'impact est un principe général** au sein des administrations des six pays étudiés, mais qui est **plus ou moins formalisé ou contraignant selon les pays**.

D'un côté, en Suède et au Canada, la publication des évaluations par les administrations est systématique ou quasi systématique car garantie dans les textes juridiques. D'un autre côté, aux États-Unis, au Royaume-Uni, en Allemagne et en France, elle est plus inégale car laissée à la discrétion des administrations. Dans ces derniers pays, la publication semble davantage assurée lorsque les évaluations sont confiées à des chercheurs extérieurs à l'institution, en raison notamment des exigences que ces derniers peuvent avoir à cet égard. Toutefois, même lorsque la publication est autorisée par le commanditaire, les évaluations ne sont pas toujours facilement accessibles à partir des sites internet des administrations.

Plus précisément, **la Suède a choisi d'ériger le principe de publication au rang d'obligation constitutionnelle**, au même titre que l'ensemble des documents produits par les administrations et organismes publics du pays³⁹. Ce principe constitutionnel de publicité et de transparence s'applique notamment, selon les experts interrogés, aux travaux d'évaluation. **Au Canada, le principe de publication est inscrit dans le cadre juridique de la politique nationale sur l'évaluation de 2009** disposant que les rapports d'évaluation de programme réalisés par les ministères et organismes publics doivent faire l'objet d'une publication sur les sites internet des administrations concernées⁴⁰. L'obligation semble bien respectée sur la période récente, même si la publication peut intervenir avec délai.

Dans les autres pays, plusieurs évolutions favorables sont en cours. Aux États-Unis, le principe de publication des évaluations devrait se généraliser dans l'administration à la suite de **l'application de la loi FEBPA de 2018**. Cette dernière prévoit la mise en œuvre d'une politique ou charte d'évaluation dans chaque agence fédérale – il en existe aujourd'hui dans les principales administrations productrices d'évaluations – qui repose sur plusieurs principes dont celui de la transparence. De même, au Royaume-Uni, la définition **d'un protocole de publication par le Government Social Research Service (GSR)**, unité qui chapeaute les chercheurs en sciences sociales travaillant dans l'administration, a conduit plusieurs ministères à renforcer récemment la publication des évaluations. Publié en 2010 et actualisé en 2015, ce protocole met en avant la publication systématique des études – dont les évaluations font partie – dans un délai maximum de douze semaines après validation du document final par la hiérarchie. Par ailleurs, depuis leur création en 2013, **les What Works Centres** ont développé une politique de publication systématique et rapide grâce à leurs outils de classification, de synthèse et de visualisation des résultats (*voir plus bas*). Dans une moindre mesure, en France, la mise en place de **conseils scientifiques ou de**

³⁹ *Freedom of the Press Act*, chap. 2, art. 1.

⁴⁰ Voir l'article 4.3.18 de la directive : <https://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=31300>.

commissions d'évaluation par certaines administrations permet également de garantir la transparence et la publication des évaluations. En Allemagne, la mise à disposition des évaluations se fait également de façon discrétionnaire, à l'initiative des ministères ou des donneurs d'ordre. La publication de l'évaluation des réformes Hartz par exemple est parcellaire. En signe d'ouverture, et à l'inverse, l'[Agence de coopération internationale allemande \(GIZ\)](#) a publié en 2016 l'ensemble des études qu'elle mettait en œuvre en son nom et avec l'accord du ministère fédéral allemand de la coopération économique et du développement (BMZ).

2. Des outils de capitalisation et de diffusion des résultats particulièrement innovants au Royaume-Uni et aux États-Unis

La publication n'implique pas la diffusion automatique des résultats et leur appropriation par les utilisateurs. C'est pourquoi dans la plupart des pays ont été mis en place par les administrations et les laboratoires de recherche des outils pour **favoriser le partage et la reprise des résultats des évaluations** par les médias, les décideurs et le grand public. Il s'agit pour l'essentiel de notes de synthèse (de type *Policy Brief*) ou des « *executive summary* » exposant de manière concise et pédagogique les questions étudiées, la méthode utilisée et les principaux résultats dans les rapports d'évaluation.

Certains pays comme le Royaume-Uni et les États-Unis sont cependant allés plus loin en proposant **des dispositifs plus innovants** pour faciliter la synthèse et la diffusion des résultats. L'exemple le plus abouti est donné par la mise en place de **plateformes ou « toolkits » adossés aux What Works Centres au Royaume-Uni et de Clearinghouses développés par certaines administrations, universités et fondations aux États-Unis**. Ces outils ont pour but de centraliser les résultats des évaluations et de les rendre accessibles au plus grand nombre sous la forme de boîtes à outils. Généralement ciblés sur un domaine particulier (l'éducation, le travail, la santé, la justice, etc.), ils se présentent comme une liste d'interventions ou de dispositifs regroupés par thématique, qu'il est possible de classer et de comparer selon notamment leur performance et la robustesse des méthodes utilisées. Ils adoptent généralement des présentations visuelles pour faciliter leur compréhension par un public non expert des méthodes statistiques. Les *toolkits* s'adressent en particulier aux professionnels du domaine (les enseignants dans le domaine de l'éducation, les policiers ou magistrats dans le domaine de la prévention de la criminalité et de la justice, etc.). Les études recensées reposent généralement sur des méthodes expérimentales et quasi expérimentales, et sont publiées dans des revues scientifiques ou sous la forme de documents de travail, de rapports d'évaluation, etc. (voir la **BONNE PRATIQUE N° 9** page suivante).

La multiplication et l'hétérogénéité des outils développés dans chaque pays – chacun adoptant son propre système de classement des dispositifs – peut cependant rendre leur utilisation complexe. Aux États-Unis, une initiative intéressante est celle entreprise par deux organisations, Pew Charitable Trusts et MacArthur Foundation. Elles ont développé une base de données, **Results First Clearinghouse Database**, qui rassemble l'ensemble des études collectées dans neuf Clearinghouses et qui offre un système de classement commun. La base comprend au total près de 3 000 dispositifs dans le domaine des politiques sociales.

BONNE PRATIQUE N° 9

Les What Works Centres et les *Clearinghouses*, deux vecteurs de diffusion des résultats des évaluations d'impact

Au Royaume-Uni, les What Works Centres sont des organismes destinés à servir de plateformes uniques d'accès aux résultats des évaluations d'impact de politiques publiques dans leur domaine de spécialité. Cinq d'entre eux – EEF, WWCCR, EIF, What Works Centre for Local Economic Growth et What Works Centre for Wellbeing – ont développé des *toolkits* pour présenter de manière visuelle les résultats des évaluations et faciliter la comparaison des dispositifs. Les plus connus sont le Teaching and Learning Toolkit de l'Education Endowment Foundation (EEF) et le Crime Reduction Toolkit du What Works Centre for Crime Reduction (WWCCR). Ces *toolkits* se présentent sous la forme d'une liste d'interventions qu'il est possible de filtrer en fonction de leurs impacts, de leurs coûts et de la robustesse des preuves scientifiques associées. Le Crime Reduction Toolkit a aussi la spécificité de préciser les modalités d'efficacité de l'intervention (« *how it works* » et « *where it works* ») pour permettre d'adapter le dispositif à chaque contexte local (voir Figure 2).

Figure 2 – Crime Reduction Toolkit du What Works Centre for Crime Reduction

Intervention ▲▼	Impact on crime ▲▼ Effect	How it works ▲▼ Mechanism	Where it works ▲▼ Moderator	How to do it ▲▼ Implementation	What it costs ▲▼ Economic cost	
After school clubs	X✓ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	?	£ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	
Aggression Replacement Training (ART) for reducing antisocial behaviour	X✓ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	?	£ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	
Alcohol pricing	✓✓ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	?	£ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	
Alley gating	✓✓ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	?	£ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	
Boot camps	X✓ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	 <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	?	£ <div style="width: 100%; height: 10px; background: linear-gradient(to right, #009682 10%, #ccc 10% 90%, #ccc 90%);"></div>	

Note : Toolkit disponible sur le site du What Works Centre for Crime Reduction. La première colonne liste des interventions qui ont fait l'objet d'évaluations. La deuxième représente les connaissances sur l'efficacité du programme : effet positif, négatif ou nul (pictogrammes) et robustesse (barre horizontale). Les colonnes suivantes présentent d'autres informations (sur le mécanisme, les effets de contexte, etc.).

Aux États-Unis, plusieurs agences ont développé des *Clearinghouses* qui sont des outils élaborés de collecte et de diffusion des évaluations, généralement ciblés sur un domaine particulier (l'éducation, le travail, la santé, la justice, etc.). Ils permettent de rassembler les études disponibles dans ce domaine, de les synthétiser dans un format lisible pour les non spécialistes et de les distinguer selon les méthodes qu'elles utilisent et les résultats auxquels elles conduisent. Les études sont regroupées par grande thématique dans le domaine et par intervention ou dispositif à l'intérieur du domaine. Elles reposent généralement sur des

méthodes expérimentales et quasi-expérimentales. Il s'agit d'études publiées dans des revues scientifiques ou de littérature grise (documents de travail, rapports d'évaluation, etc.). Les Clearinghouses permettent de déterminer quelles interventions sont les plus efficaces et pour quelles populations et quels contextes. Il existe aujourd'hui plusieurs Clearinghouses. Certaines sont propres aux ministères ou aux agences fédérales, par exemple la What Works Clearinghouse (WWC) du Department of Education (cf. l'exemple du programme Head Start pour mieux préparer à l'école les jeunes enfants issus de milieux défavorisés, figure 3), le Research-Tested Intervention Programs (RTIP) du Department of Health and Human Services, ou la Teen Pregnancy Prevention (TPP) Evidence Review du même ministère. D'autres Clearinghouses ont été développées par des universités, des fondations ou des administrations locales, comme par exemple les Blueprints for Healthy Youth Development (Institute of Behavior Science, University of Colorado Boulder, Annie E. Casey Foundation), les Social Programs That Work (Laura and John Arnold Foundation) ou la California Evidence-Based Clearinghouse for Child Welfare (California Department of Social Services).

Figure 3 – Présentation du programme Head Start à partir de la base de données What Works Clearinhouse du ministère de l'Éducation

Head Start

Early Childhood Education

Head Start was found to have potentially positive effects on general reading achievement and no discernible effects on mathematics achievement and social-emotional development for 3- and 4-year-old children.

Head Start is a national, federally-funded program that provides services to promote school readiness for children from birth to age 5 from predominantly low-income families. These services are provided to both children and their families and include education, health and nutrition, family engagement, and other social services. Head Start program administrators are given the flexibility to design service delivery to be responsive to cultural, linguistic, and other contextual needs of local communities, leading to considerable variability in the services offered. Head Start service models also vary according to family needs, such that children and families may be served through center-based or family child care, home visits, or a combination of programs that operate full or half days for 8–12 months per year.

Findings

1 STUDIES THAT MET STANDARDS OUT OF

40 ELIGIBLE STUDIES REVIEWED

Outcome Domain	Effectiveness Rating	Grades	Improvement Index
Mathematics achievement	0	PK	--
Reading achievement	+	PK	-50 0 +50 13
Social-emotional development	0	PK	--

Last Updated: July 2015

Note : le niveau de preuve d'efficacité associé au programme Head Start est évalué séparément sur trois variables d'intérêt : la réussite en mathématiques, la réussite en lecture et le développement socio-affectif. Un indice indique la progression qu'aurait connue un enfant « moyen » du groupe de comparaison s'il avait bénéficié du programme : un indice de 13 signifie que sur un groupe de 100 élèves, il aurait gagné 13 places au classement en passant de la 50^e à la 37^e place.

Source : What Works Clearinghouse. Topics : Early Childhood.

Il n'existe pas à notre connaissance d'outils de capitalisation et de diffusion des résultats similaires dans les autres pays. **Au Canada, deux démarches récentes ont été initiées en 2017 dans le domaine des politiques sociales, *Social Systems Evidence (SSE)* et**

Health Systems Evidence (HSE). Elles sont financées principalement par l'Université McMaster (Ontario). Elles se présentent sous la forme d'un moteur de recherche permettant de retrouver des études, évaluations et synthèses dans une base de référence comptant plus de 15 000 références. Ces dernières proviennent quant à elles de sources internationales, principalement américaines et britanniques⁴¹. En France, la réflexion sur ce sujet est encore à un stade embryonnaire. Avec le soutien de la Direction générale de la Santé (DGS), **Santé Publique France a lancé en 2018 un répertoire des interventions efficaces ou prometteuses** dans le domaine de la prévention et de la promotion de la santé. Ce répertoire compte à ce jour près de 80 références (en grande partie des évaluations randomisées⁴²) couvrant une quinzaine de thématiques (addiction, santé mentale, vieillissement, parentalité, etc.) et en synthétise les résultats obtenus.

3. Des incitations à l'utilisation des résultats en vigueur aux États-Unis et au Royaume-Uni, mais encore peu employées

Comment s'assurer que la connaissance produite soit la plus utile pour la décision publique ? C'est la question à laquelle plusieurs pays cherchent à répondre en mettant l'accent, ces dernières années, sur l'enjeu de l'utilisation des résultats des évaluations par les décideurs et les citoyens. En ce qui concerne les évaluations d'impact, la réflexion est la plus avancée au Royaume-Uni et aux États-Unis.

Cela s'est traduit d'abord par **le financement d'études et de recherches** afin d'identifier les facteurs permettant de renforcer l'utilisation des évaluations dans la prise de décision. Au Royaume-Uni, par exemple, des chercheurs de l'EPPI-Centre ont mené, en 2016, sous la direction de l'*Alliance for Useful Evidence*, une méta-analyse⁴³ passant en revue l'ensemble des interventions visant à améliorer l'utilisation des preuves scientifiques par les décideurs (*evidence-informed decision-making*) afin de distinguer les méthodes de diffusion les plus efficaces. Cette étude conclut par exemple que les interventions qui facilitent l'accès aux résultats sous un format simplifié comme les répertoires de preuve et celles qui permettent de former les décideurs à l'utilisation des données scientifiques sont celles ayant l'impact le plus important et le plus robuste sur le taux d'utilisation de preuves.

De manière plus singulière, le Royaume-Uni a mis en place dans les ministères une série de quatre « audits sur les preuves » (« *evidence audits* ») par la *What Works Team* en 2013 afin d'identifier les domaines de politique publique où l'utilisation des preuves pourrait être renforcée. Le premier de ces audits a été effectué sur le ministère du Développement international (DFID) et a permis à la *What Works Team* de faire des suggestions pour améliorer la production, la diffusion et la prise en compte des évaluations d'impact et des preuves scientifiques⁴⁴.

Plus largement, les pays les plus avancés sur cette question ont cherché à mettre en place des formes d'incitations et de contraintes à l'utilisation des évaluations dans la prise de décision publique ou dans les pratiques des professionnels. Les administrations canadienne

⁴¹ Campbell Collaboration, Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), International Initiative for Impact Evaluation (3ie) et les What Works Centres au Royaume-Uni.

⁴² Voir le détail de la méthode de sélection et la base de publications [sur le site Santé publique France](#).

⁴³ Langer, Tripney et Gough, EPPI-Centre, 2016.

⁴⁴ Department for International Development, 2014.

et britannique ont ainsi édicté des normes et des standards professionnels ([Policy Profession Standards](#) au Royaume-Uni, [Cadre national sur l'évaluation](#) au Canada). En particulier, les fonctionnaires britanniques qui appartiennent à la Policy profession – c'est-à-dire qui sont impliqués dans la conception des politiques publiques – sont individuellement tenus par leur standard de compétences d'utiliser les données scientifiques dans leur travail, et notamment d'utiliser les résultats d'évaluation pour appréhender l'efficacité d'un dispositif et pour faire des recommandations de politiques publiques (*voir partie 2*). Le levier financier constitue une autre voie d'action⁴⁵, comme en témoigne le développement récent dans plusieurs pays des dispositifs de type **Social Impact Bonds (SIB)**. Ce système des SIB apparu en 2007 au Royaume-Uni et repris depuis par plusieurs pays permet à des investisseurs privés de financer des interventions délivrées par des organisations non gouvernementales ou à but non lucratif qui, si elles sont fructueuses, seront ensuite remboursées par le gouvernement (*voir la BONNE PRATIQUE N° 10*)⁴⁶. Les interventions financées par SIB font l'objet d'une évaluation qui permet d'établir si le dispositif a ou non atteint ses objectifs, et donc si l'investisseur privé sera remboursé par le décideur public. Un exemple emblématique est donné par le [SIB mis en place par le ministère de la Justice \(MOJ\) et le Big Lottery Fund](#) à Peterborough en 2010 afin de lutter contre la récidive auprès de 3 000 prisonniers ayant été condamnés à une incarcération de moins de 12 mois. Connu sous le nom de **Pay for Success (PFS)** aux États-Unis, ce mécanisme est apparu pour la première fois en 2010. La plupart des projets de type PFS utilisent des méthodes rigoureuses d'évaluation, comme les méthodes expérimentales ou quasi expérimentales.

BONNE PRATIQUE N° 10

Les Social Impact Bonds au Royaume-Uni et aux États-Unis

Le mécanisme des *Social Impact Bonds* (SIB) est originaire du Royaume-Uni. Il permet de lever des fonds du secteur privé et philanthropique pour financer des interventions sociales préventives délivrées par des organisations non gouvernementales ou à but non lucratif, qui en cas de succès seront remboursées par le gouvernement avec un intérêt. Il fait également intervenir un intermédiaire qui gère les contrats entre les parties prenantes et surveille la mise en œuvre du programme ainsi qu'un évaluateur indépendant qui détermine si le programme atteint ses objectifs. Ce mécanisme permet un transfert de risque (celui de devoir payer les coûts d'un programme qui n'atteindrait pas ses objectifs) du gouvernement aux investisseurs. Si le programme est efficace, le gouvernement paie l'intermédiaire qui à son tour paie les investisseurs (avec un intérêt). Dans le cas contraire, le gouvernement ne paie rien et les investisseurs supportent entièrement le coût du programme. Les montants payés sont en théorie fonction des économies de coût générées par le programme. Lorsque la rentabilité espérée du programme est trop faible au regard de son coût, un tiers intervient parfois pour se porter garant et supporter une partie du coût du programme en cas d'échec de celui-ci.

⁴⁵ Certaines administrations ont aussi permis de faire entrer l'utilisation des résultats des évaluations dans les processus d'arbitrages budgétaires pour justifier leur demande de crédits. Au Royaume-Uni, ces discussions font intervenir les résultats d'évaluations d'impact démontrant l'efficacité des programmes.

⁴⁶ Gustafsson-Wright, Gardiner et Putch, 2015.

Le premier SIB a été mis en place par le ministère de la Justice (MOJ) et le Big Lottery Fund à Peterborough en 2010 afin de lutter contre la récidive auprès de 3 000 prisonniers ayant été condamnés à une incarcération de moins de douze mois. Aujourd'hui le gouvernement britannique finance sept fonds de SIB pour une valeur cumulée de plus de 150 millions de livres et dispose d'un Centre for Social Impact Bonds afin d'encourager leur développement.

Aux États-Unis, *Pay for Success* (PFS) est un mécanisme de financement de programmes similaire. Il incite les parties prenantes à choisir des programmes appuyés par des preuves solides d'efficacité pour limiter les risques d'échec. Ces preuves permettent également de déterminer les termes du contrat, c'est-à-dire les objectifs à atteindre. Par ailleurs, il requiert que les programmes soient évalués de manière indépendante pour déterminer s'ils atteignent leurs objectifs. La plupart des projets de type PFS utilisent des méthodes rigoureuses, comme les méthodes expérimentales ou quasi expérimentales. Elles garantissent que les résultats mesurés sont bien attribuables au programme⁴⁷⁴⁸.

En 2016, 11 projets de type PFS ont été lancés aux États-Unis dans les domaines de l'éducation de la petite enfance, du sans-abrisme chronique et de la récidive⁴⁹. Ils émanent de gouvernements locaux exclusivement (États, villes, comtés). Mais le gouvernement fédéral, notamment le ministère du travail (DOL), octroie des fonds pour soutenir les projets Pay for Success. Ainsi, le DOL s'est engagé aux côtés des États de New York et du Massachusetts à financer les projets PFS portés par eux pour un montant total de 24 millions de dollars en cas de succès (12 millions pour chaque projet⁵⁰).

De même, aux États-Unis, sous l'administration Obama, de nouveaux outils de financement des programmes ont été mis en place reposant sur la preuve de leur efficacité. Il s'agit des « **Tiered-Evidence Grants** », aussi appelés « **Innovation Fund Grants** ». Dans ces nouveaux modèles, deux ou trois types de subventions sont généralement distingués, dont les montants attribués sont d'autant plus importants que les dispositifs proposés sont appuyés par des preuves solides de leur efficacité. La force de la preuve est fonction notamment de la méthode utilisée (méthodes expérimentales versus méthodes quasi expérimentales) et du nombre d'études à l'appui. L'évaluation joue alors un rôle essentiel dans la décision de financer un dispositif et dans le montant qui lui est octroyé (voir la **BONNE PRATIQUE N° 11**).

Ces nouvelles structures de subventionnement ont plusieurs avantages par rapport aux structures classiques. D'abord, elles incitent les candidats à choisir des dispositifs efficaces pour obtenir des montants de subvention plus élevés, ces derniers étant fonction de la force de la preuve disponible. Celle-ci est évaluée à l'aide des standards académiques disponibles qui établissent une hiérarchie des preuves, les méthodes expérimentales apparaissant

⁴⁷ GAO, 2015

⁴⁸ Non Profit Finance Fund, 2016

⁴⁹ Huit nouveaux projets PFS ont été lancés en 2017 par les localités. Pour une liste mise à jour des projets PFS, se reporter au site <https://www.payforsuccess.org/projects/comparison>.

⁵⁰ Le montant total maximal à rembourser est de 21,5 millions de dollars pour le projet de l'État de New York et de 28 millions pour le projet de l'État du Massachusetts. Le DOL s'est engagé à apporter 12 millions pour chacun des projets. La somme restante doit être supportée par les États.

comme les plus performantes. Ensuite, elles exigent généralement que les projets soient évalués rigoureusement (à l'aide de méthodes expérimentales ou quasi expérimentales), ce qui permet la production de nouvelles évaluations qui alimenteront à leur tour les décisions futures de financement de projets. Enfin, elles encouragent l'innovation en finançant des projets prometteurs mais peu testés.

BONNE PRATIQUE N° 11

Les Tiered-Evidence Grants aux États-Unis

L'administration Obama a mis en place de nouveaux outils de financement de programmes reposant sur la preuve de leur efficacité, appelés « *Tiered-Evidence Grants* » ou « *Innovation Fund Grants* ». Le gouvernement fédéral octroie des subventions importantes aux États, aux collectivités locales et à des organisations à but non lucratif (des universités, des organismes de bienfaisance, etc.) pour mettre en œuvre des programmes sociaux dans les domaines de l'éducation, de la santé, de l'emploi ou encore du logement. L'administration Obama a encouragé les ministères en charge de ces programmes à orienter davantage leurs financements vers les dispositifs ayant apporté la preuve de leur efficacité. L'évaluation joue alors un rôle essentiel dans la décision de financer un dispositif et dans le montant qui lui est octroyé.

Dans ces nouveaux modèles, trois types de subventions sont généralement distingués. Le premier type octroie les montants de financement les plus importants et vise les dispositifs ayant apporté des preuves solides de leur efficacité (par exemple sur la base d'évaluations de type RCT multi-sites). Il s'agit de les répliquer sur d'autres territoires ou de les étendre à une population plus large : c'est l'étape de mise à l'échelle des projets ou « *scale up tier* ». Le second type accorde des montants plus faibles que le premier niveau pour des dispositifs prometteurs mais ayant apporté des preuves moins fortes de leur efficacité (par exemple sur la base d'une évaluation de type RCT uni-site ou utilisant des méthodes quasi expérimentales bien conçues) ; il s'agit de l'étape de validation des projets ou « *validation tier* ». Enfin, le troisième type octroie les montants les plus faibles pour des dispositifs innovants, à haut potentiel mais dont les résultats n'ont pas encore été mesurés par des évaluations rigoureuses ; il s'agit de l'étape de la preuve du concept.

Depuis 2010, neuf programmes adoptant cette structure de financement ont été mis en place par cinq agences fédérales. En 2016, on en comptabilise six dans quatre agences fédérales pour un montant total de près de 800 millions de dollars⁵¹. Cela représente cependant une part faible des subventions accordées aux États et aux collectivités locales⁵².

⁵¹ Feldman et Haskins, *Evidence-Based Policymaking Collaborative*, 2016

⁵² Le gouvernement fédéral dépense environ 600 milliards de dollars chaque année en subventions pour les États et les collectivités locales (voir GAO, 2015).

Conclusion

Au terme de ce panorama des pratiques et des enjeux de l'évaluation d'impact dans les pays les plus avancés, il serait tentant d'opposer à l'effervescence des pratiques académiques observés dans l'ensemble des pays les difficultés récurrentes liées à leur institutionnalisation au sein des systèmes politico-administratifs et, ce faisant, leur faible capacité à peser sur les décisions publiques.

L'analyse approfondie des différents pays montre cependant un paysage plus contrasté. Elle atteste, d'abord, qu'il existe bel et bien des déterminants importants à prendre en compte pour comprendre la réussite de l'évaluation d'impact dans les différents contextes nationaux. Qu'il s'agisse de l'ouverture du milieu administratif aux acteurs de la recherche aux États-Unis ou de la place des agences exécutives d'évaluation des politiques publiques en Suède, les exemples étudiés révèlent l'existence de leviers d'actions propres à chaque pays. Ce faisant, l'étude invite à considérer le développement de l'évaluation d'impact comme un processus dynamique qui concourt, sur une période longue, à l'extension des savoirs et à l'amélioration des dispositifs publics.

La prise de conscience de l'importance de l'évaluation d'impact et des politiques fondées sur la preuve apparaît également comme une préoccupation récente dans la majorité des pays. Les pratiques les plus matures, notamment aux États-Unis et au Royaume-Uni, sont dans une certaine mesure moins liées aux avancées méthodologiques qu'à une volonté politique d'améliorer l'efficacité et la transparence de l'action publique et de relier les enjeux de l'évaluation à un ensemble plus large d'enjeux : association des parties prenantes, recours à l'expérimentation et à l'essai, apport des sciences comportementales, diffusion des résultats sous des formats innovants, etc.

Plus largement, la comparaison des différents pays montre aussi qu'en matière de développement d'évaluation d'impact, il n'existe pas de recette miracle ni de modèle à suivre. La situation diffère d'un pays à l'autre, et chaque pays développe sa propre trajectoire en fonction de son contexte politico-administratif, de l'état de son milieu académique et des rapports de pouvoir au sein des processus de décision – trois composantes qui restent fondamentales dans la constitution de systèmes évaluatifs. L'exemple singulier du Canada où une culture ancienne de l'évaluation des programmes publics n'a pas encore donné lieu au développement de l'évaluation d'impact montre aussi que le *continuum* qui irait du contrôle de la performance de l'action publique à la mesure scientifique de ses résultats ne va pas de soi.

Ces constats généraux invitent dès lors à rester modeste sur les recommandations qui pourraient être tirées à l'issue de cette étude. Ils appellent aussi à se prémunir contre la tentation de la « greffe institutionnelle » ou l'importation, sans adaptation préalable, d'une bonne pratique repérée ici ou là. Pour autant, au-delà de ces bonnes pratiques, plusieurs orientations de nature générale peuvent être retenues de cette étude et constituent sans nul doute des voies à envisager pour soutenir l'élan de l'évaluation d'impact en France.

La **première orientation** porte sur une meilleure articulation entre la demande et la production d'évaluation d'impact. Les succès des pays anglo-saxons résident en partie dans le fait d'avoir su agir sur la demande d'évaluation, en associant les universitaires à l'élaboration de l'agenda évaluatif des gouvernements et des parlements mais aussi en

pointant les « manques de preuve » dans certains secteurs d'action publique et en faisant de la commande d'évaluation une politique nationale coordonnée. Cela suppose un changement de posture de la part des décideurs publics qui doivent accepter de ne pas avoir toutes les réponses aux questions et de partager leurs besoins de connaissance. Mais cela garantit que l'évaluation puisse intervenir aux moments clés d'une politique publique, lorsque les décideurs sont confrontés à un manque d'information sur ce qui fonctionne ou non dans l'intervention publique. D'une conception de l'évaluation d'impact comme une démarche scientifique d'apport de connaissances nouvelles à destination de décideurs plus ou moins réceptifs, il convient de passer à une approche partant des besoins des décideurs en veillant à mieux les programmer et à davantage les coordonner.

La **deuxième orientation** concerne les conditions de déploiement de l'évaluation d'impact, et la définition de principes communs. À rebours des débats qui se concentrent généralement sur la création de dispositifs institutionnels (nouvelles structures, réglementation, etc.), l'étude montre que dans les pays les plus avancés, le développement de l'évaluation s'appuie d'abord sur des principes et des valeurs partagés : l'indépendance de l'évaluation, son inscription dans la durée, le niveau d'exigence méthodologique, la disponibilité des données, le souci d'expérimenter avant de généraliser, l'attention portée à la transférabilité des résultats, la place de la connaissance scientifique dans la décision publique... L'ensemble de ces principes forme un environnement favorable à la pratique de l'évaluation et au développement des capacités d'évaluation. Les expériences étrangères démontrent que la mise en place d'instances spécifiques ou de budgets dédiés – qui permettent assurément d'intensifier les pratiques – sont plus souvent la conséquence que la cause de cet environnement propice à l'évaluation.

Enfin, la **troisième orientation** invite à se placer du côté des acteurs de l'évaluation et à réfléchir à la mise en place d'une alliance réunissant une large communauté d'acteurs intéressés. Quel que soit le domaine d'intervention concerné ou les méthodes utilisées, dans tous les pays les acteurs font face aux mêmes défis pour concevoir et mettre en œuvre des politiques fondées sur la preuve : ressources limitées, manque de visibilité, faible connaissance des travaux... C'est pourquoi, à l'instar de l'Alliance for Useful Evidence lancée au Royaume-Uni et étendue aujourd'hui à plusieurs pays, il est nécessaire d'encourager la mise en commun des efforts afin de réunir autour de l'enjeu de l'évaluation d'impact un ensemble de compétences, techniques mais aussi stratégiques, visant à intéresser la communauté la plus large de chercheurs, professionnels de terrain et parties prenantes des politiques publiques. L'enjeu consiste à favoriser la diffusion d'une culture de l'évaluation d'impact par la mise en place d'un écosystème et la mise en commun des efforts, dans une logique d'ouverture, de pluralisme et d'interdisciplinarité. Cette dernière orientation de création d'une alliance des acteurs de l'évaluation est sans doute la plus ardue à concrétiser, dans un modèle français où les cloisonnements académiques et sectoriels sont forts et où les incitations à coopérer sont faibles. Elle est pourtant celle qui pourrait remporter *in fine* l'adhésion des décideurs et du grand public pour l'évaluation d'impact et garantir que des évaluations plus nombreuses soient non seulement produites mais également diffusées et utilisées.

Références

- **Références transversales**

Bachtler J., Polverari L., Taylor S., Ashcroft B. et Swales K. (2000), « Methodologies used in the Evaluation of the Effectiveness of European Structural Funds: A Comparative Assessment. Final Report to the Scottish Executive », European Policies Research Centre and Fraser of Allander Institute.

Bono P.-H., Debu S., Desplatz R., Hayet M., Lacouette-Fougère C. et Trannoy A. (2018), « [Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique](#) », *Document de travail*, n° 2018-06, France Stratégie, décembre.

Bozio A. (2015), « L'évaluation des politiques publiques: enjeux, méthodes et institutions », in *Revue française d'économie*, vol. 29, n° 4, p. 59-85.

Bozio A. et Geoffard P.-Y. (2017), *L'accès des chercheurs aux données administratives – État des lieux et propositions d'actions*, rapport au secrétaire d'État chargé de l'Industrie, du Numérique et de l'Innovation, groupe de travail du CNIS.

CAE (2013), « Évaluation des politiques publiques », *Les notes du Conseil d'analyse économique*, n° 1, février.

Commission on Evidence-Based Policymaking (2017), *The promise of Evidence-Based Policymaking*, rapport, septembre.

Deloitte (2013), *Regards croisés sur la démarche d'évaluation des politiques publiques. Benchmark international*.

Desplatz R. et Ferracci M. (2016), « Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et des praticiens », France Stratégie.

Ferracci M. et Wasmer É. (2012), *État moderne, État efficace : évaluer les dépenses publiques pour sauvegarder le modèle français*, Paris, Odile Jacob.

Fougère D. (2010), « Les méthodes économétriques d'évaluation », in *Revue française des affaires sociales*, 2010/1-2, p. 105-128.

Furubo J.-E., Rist R. C. et Sandahl R. (2002), *International Atlas of Evaluation*, New Brunswick & Londres, Transaction Publishers.

Lázaro B. (2015), *Comparative study on the institutionalization of evaluation*, EUROsociAL, State of the Art Study, n° 15, juillet, p. 41-46.

OECD (2015), *OECD Regulatory Policy Outlook 2015*, Paris, OECD Publishing.

OECD (2019), *Institutionalisation, quality and use of policy evaluation: Governance Lessons from Countries' Experience* (Preliminary consolidated results from the OECD Policy Evaluation Survey)

Polverari L. et Bachtler P. (2004), « Assessing the Evidence: the Evaluation of Regional Policy in Europe », *European Policy Research Paper*, n° 56, European Policies Research Centre.

Sharma M. R., Freudl D., Münch C., Schindler E. et Wegrich K. (2013), « Expert report on the implementation of ex-post evaluation: Good practice and experience in other countries », Prognos, Berlin.

Sustainable Governance Indicators 2017

Varone F. et Jacob S. (2004), « Institutionnalisation de l'évaluation et nouvelle gestion publique : un état des lieux comparatif », *Revue internationale de politique comparée*, 11(2), p. 272-292.

- **Références par pays**

Pour la bibliographie complète, se reporter aux documents de travail consacrés aux pays.

Canada

SCT/CEE (2004), *Guide for the Review of Evaluation Reports Prepared by the Centre of Excellence for Evaluation*, janvier.

SCT/CEE (2004), *Examen de la qualité des évaluations dans les ministères et les organismes*, Ottawa.

SCT/CEE (2009), [Politique sur l'évaluation](#).

SCT/CEE (2009), [Directive sur la fonction d'évaluation](#).

SCT (2014), [Évaluation de la Politique sur l'évaluation de 2009](#).

États-Unis

Feldman A. et Haskins R. (2016), [Tiered-Evidence Grantmaking](#), Toolkit, septembre.

GAO (2015), *Pay For Success, Collaboration among Federal Agencies Would Be Helpful as Governments Explore New Financing Mechanisms*, [GAO-15-646](#), septembre.

Nightingale D.S., Fudge K. et Schupmann W. (2018), [Evidence Toolkit: Learning Agendas](#), Research Report, mars.

Nightingale D.S. et Scott M.M. (2018), [Building Evidence Culture and Capacity in Federal Agencies](#), Brief, Urban Institute, avril.

Non Profit Finance Fund (2016), [Pay for Success: The first Generation, A comparative Analysis of the First 10 Pay for Success Projects in the United States](#), rapport.

Office of Management and Budget (OMB) :

Budget of the US Government, [Major Savings and Reforms](#), Fiscal Year 2019.

Increased Emphasis on Program Evaluations, M-10-01, Washington, D.C., 27 octobre 2009.

Use of Evidence and Evaluation in the 2014 Budget, M-12-14, Washington, D.C., 18 mai 2012.

Next Steps in the Evidence and Innovation Agenda, M-13-17, Washington, D.C., 26 juillet 2013.

Results for America (2018), [Invest in what works federal standard of excellence](#).

Schupmann W. et Fudge K. (2018), [Chief Evaluation Officers](#), Toolkit, juin.

US Department of Labor, (2013), [Evaluation policy](#).

France

Barrot J.-N. et Eliaou J.-F. (2018), *Propositions du groupe de travail sur les moyens de contrôle et d'évaluation*, Assemblée nationale, [second rapport](#), juin.

CEC (2018), *L'évaluation des dispositifs d'évaluation des politiques publiques*, [rapport d'information n° 771](#), présenté par Pierre Morel-à-L'Huissier et Valérie Petit, Comité d'évaluation et de contrôle des politiques publiques, mars.

Insee (2017), [Charte des services statistiques ministériels](#)

[Constitution](#), article 24

[Loi n° 2012-1510 du 29 décembre 2012 de finances rectificatives pour 2012](#), article 66

[Loi n° 2017-1837 du 30 décembre 2017 de finances pour 2018](#), article 28

Royaume-Uni

Allard C. et Rickey B. (2017), *What Works Centres britanniques: quels enseignements pour des politiques fondées sur la preuve en France ?*, Agence nouvelle des solidarités actives, mars.

Behavioural Insights Team (2017), *Update Report 2016-17*, octobre.

Evidence into Action and What Works Team (2014), *What Works Review of the Use of Evidence in the Department for International Development (DFID)*, juin.

Gustafsson-Wright E., Gardiner S. et Putchá V. (2015), *The Potential and Limitations of Impact Bonds. Lessons from the first five years of experience worldwide*, Brookings, juillet.

Haynes L., Service O., Goldacre B. et Torgerson D. (2012), *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*, Behavioural Insights Team, juin.

Her Majesty's Treasury (2003), *The Green Book: Central Government Guidance on Appraisal and Evaluation*, deuxième édition publiée en mars 2018 (première édition 2003).

Her Majesty's Treasury (2011), *The Magenta Book: Guidance for evaluation*, avril.

Her Majesty's Treasury (2012), *Quality in policy impact evaluation: understanding the effects of policy from other influences (supplementary Magenta Book guidance)*, décembre.

Langer L., Tripney J. et Gough D. (2016), *The Science of Using Science: Researching the Use of Research Evidence in Decision-Making*, EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London, Londres, avril.

Puttick R. (2018), *Mapping the Standards of Evidence used in UK social policy*, Alliance for Useful Evidence, mai.

Suède

Organisation de coopération et le développement économiques (2013), *Value for Money in Government: Sweden 2013*, OECD Publishing, Paris.

[Freedom of the Press Act](#), chap. 2, art. 1

[Instrument of Government](#), chap. 4, art. 8

[Ordinance on impact analysis of Regulation](#)

Annexe 1

Grille d'analyse par pays

Introduction : le contexte institutionnel de l'évaluation des politiques publiques

- **Dates et définitions clés concernant l'évaluation**
- **Cartographie des principaux acteurs de l'évaluation des politiques publiques** (entendu au sens large et pas seulement dans le sens plus étroit de l'évaluation d'impact avec contrefactuel).

Axe I – La production d'évaluations d'impact

- **Volume et évolution de la production académique**
 - Combien d'évaluations d'impact ont été publiées depuis une quinzaine d'années dans ce pays ?
 - Y a-t-il eu une augmentation notable du nombre d'évaluations d'impact ces dernières années ?
- **Principaux producteurs et focus sur quelques structures remarquables**
 - Description succincte des structures (origine, objectifs, ressources...).
 - Domaines concernés, méthodes utilisées, ouverture aux expérimentations.
 - Origine des évaluations : auto-saisines ou commandes extérieures (par qui ?).
- **Thématiques couvertes et évaluations emblématiques**
 - Domaines de la politique publique les plus traités (éducation, emploi, logement, santé, environnement, exclusion, fiscalité, etc.) et explications possibles : offre (de données) plus importante, demande sociale ou politique dans certains secteurs ?

Encadré : exemples de projets emblématiques, auteurs des évaluations, budget, résultats.

Axe II – Place des évaluations d'impact dans les processus de décision et les pratiques

Critère 1 : Degré d'ouverture du milieu administratif aux chercheurs en économie

- **Politique de recrutement des chercheurs dans l'administration et de formation des fonctionnaires à l'évaluation d'impact**

- Q1 : Existe-t-il des procédures RH permettant de recruter des chercheurs en économie sur des postes de cadres dans l'administration et permettant ensuite de valoriser ces expériences dans les carrières académiques ?
- Q2 : Existe-t-il des actions de sensibilisation et des formations (initiale et continue) à l'évaluation d'impact des politiques publiques pour les cadres de l'administration et les décideurs publics ?
- **Mise en place de partenariats entre l'administration et les chercheurs**
 - Q3 : Existe-il des formes de partenariats entre l'administration et la recherche en économie ?
 - *Si oui, lesquelles (conventions, conseils scientifiques, chaires, financement de structures, etc.) ?*
 - *Pour quoi faire ? (réfléchir aux modalités d'évaluation ex post d'une nouvelle mesure avant son instauration ; réaliser des évaluations ; accompagner des fonctionnaires pour qu'ils puissent eux-mêmes les réaliser ; synthétiser la littérature académique ; former aux méthodes)*

Critère 2 : Qualité des dispositifs formels favorisant la conduite d'évaluations d'impact des politiques publiques

• **Instances et structures**

- Q4 : Existe-t-il une instance de coordination des évaluations d'impact *ex post* ?
- Q5 : Existe-t-il une programmation annuelle ou pluriannuelle des évaluations d'impact ?
- Q6 : Existe-t-il une réflexion sur la qualité des évaluations d'impact réalisées par l'administration ?
 - *Quelles en sont les modalités : procédures classiques dans le champ académique (revue par les pairs, comités de lecture...), élaboration de guides méthodologiques, mise en place d'un système de suivi et de contrôle de la qualité de ces évaluations, fait par un acteur dédié, indépendant ?*

• **Processus formalisés et cadre juridique**

- Q7 : Existe-t-il dans les textes juridiques (législatifs, réglementaires) des dispositions relatives à l'évaluation d'impact⁵³ ?
 - *Les exigences en termes d'évaluation sont-elles proportionnées à l'importance de la politique publique (par exemple, en termes budgétaires) ? Ces dispositions*

⁵³ Par exemple, il peut s'agir de *post-implementation review* (PIR) dans le cas des études d'impact préalables des projets de loi, ou bien de *sunset* ou *review clause* pour les projets de loi. Une *sunset clause* est une disposition législative qui prévoit l'expiration d'une mesure à une date fixée et qui permet au Parlement de décider à cette date si la mesure doit être reconduite (en l'état ou après quelques modifications) ou supprimée compte tenu de ses résultats. Une *review clause* prévoit le réexamen d'une mesure dans un délai préétabli sans extinction automatique de la mesure.

concernent-elles l'évaluation d'impact des politiques publiques (mesure des effets) ? Sont-elles prescriptives sur les méthodes à employer ?

- Ces dispositions sont-elles nombreuses aujourd'hui dans le pays considéré (volume) ? Sont-elles systématiques (par exemple au-delà d'un certain seuil de dépenses) ? Concernent-elles tous les domaines de la politique publique ?
- Prévoient-elles une remise des travaux d'évaluation au Parlement ou au gouvernement ? Leur calendrier de réalisation est-il synchronisé avec l'agenda politique/législatif ?

- **Financement**

- Q8 : Existe-t-il des financements dédiés à l'évaluation d'impact des politiques publiques ?

Critère 3 : Degré de diffusion et d'influence des évaluations

- **Transparence**

- Q9 : Pour les acteurs administratifs seulement : les évaluations d'impact *ex post* sont-elles systématiquement publiées ? Sous quel délai ?
- Q10 : La démarche d'évaluation d'impact est-elle ouverte aux parties prenantes (comité pluraliste, consultation préalable...) ?

- **Diffusion et réception**

- Q11 : Des collections spécifiques sont-elles disponibles pour le grand public, les médias, les décideurs, les professionnels (*Policy Brief*, synthèses de la recherche, etc.) ?
- Q12 : Existe-t-il des plateformes d'accumulation et de vulgarisation des résultats des évaluations d'impact de politiques publiques ?
- Q13 : Existe-t-il des lieux ou des temps d'échanges et de mise en débat des résultats des évaluations d'impact ?

- **Utilisation et influence**

- Q14 : Existe-t-il une réflexion sur l'utilisation des résultats des évaluations d'impact dans la décision publique ou les pratiques ?
- Q15 : Existe-t-il des formes d'incitations ou de contraintes à l'utilisation des évaluations d'impact (par les décideurs, les fonctionnaires, les professionnels) ?

Encadré : exemples d'évaluations emblématiques et suites données à leurs résultats

RETROUVEZ
LES DERNIÈRES ACTUALITÉS
DE FRANCE STRATÉGIE SUR :



www.strategie.gouv.fr



[@Strategie_Gouv](https://twitter.com/Strategie_Gouv)



[france-strategie](https://www.linkedin.com/company/france-strategie)



[FranceStrategie](https://www.facebook.com/FranceStrategie)



[@FranceStrategie_](https://www.instagram.com/FranceStrategie_)



[StrategieGouv](https://www.youtube.com/StrategieGouv)



FRANCE STRATÉGIE

Institution autonome placée auprès du Premier ministre, France Stratégie contribue à l'action publique par ses analyses et ses propositions. Elle anime le débat public et éclaire les choix collectifs sur les enjeux sociaux, économiques et environnementaux. Elle produit également des évaluations de politiques publiques à la demande du gouvernement. Les résultats de ses travaux s'adressent aux pouvoirs publics, à la société civile et aux citoyens.