

L'évaluation des politiques publiques aux États-Unis

Rozenn Desplatz



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

Table des matières

Résumé.....	3
Introduction : le contexte institutionnel	5
Axe I – La production d'évaluations d'impact.....	11
1. Un volume très élevé de travaux sur les États-Unis.....	11
2. Les universités sont les principales institutions productrices	12
3. Les évaluations d'impact portent en particulier sur les domaines de la santé et de l'éducation	15
Axe II – Place des évaluations d'impact dans les processus de décision et les pratiques	18
1. Une ouverture importante du milieu administratif aux universitaires, se traduisant par de nombreuses collaborations.....	18
1.1. Des recrutements d'universitaires dans l'administration possibles mais parfois difficiles.....	18
1.2. Des formations aux méthodes d'évaluation généralement prévues mais surtout pour les spécialistes.....	19
1.3. Des partenariats nombreux et divers entre l'administration et les universitaires	20
2. De nombreux dispositifs formels favorisant la conduite et l'utilisation des évaluations d'impact.....	22
2.1. Une structuration institutionnelle bien avancée et qui va se renforcer.....	22
2.2. Un cadre juridique favorable à l'évaluation d'impact mais pas systématique	27
2.3. Des sources de financement diverses	27
3. Des efforts sur la diffusion et la vulgarisation des évaluations, une réflexion et une pratique avancées sur leur utilisation dans la décision publique.....	29
3.1. Un principe de transparence dans plusieurs agences fédérales	29
3.2. Des outils élaborés de diffusion des résultats des évaluations : les Clearinghouses	30
3.3. De nouveaux outils de financement des programmes reposant sur la preuve de leur efficacité : Tiered-Evidence Grants et Pay for Success (PFS)	32
Annexe 1 – Grille d'analyse par pays	41
Annexe 2 – Dates clés de l'évaluation	45
Annexe 3 – Principaux acteurs de l'évaluation.....	47
Annexe 4 – Les critères de qualité méthodologique du What Works Clearinghouse	51
Annexe 5 – Les Clearinghouses.....	53

Annexe 6 – Tiered evidence grants.....	57
Annexe 7 – Pay for success	67
Références et sources	71

Résumé

Ce document de travail examine la production d'évaluations d'impact et leur place dans le processus de décision aux États-Unis. Il s'inscrit dans une série de six monographies ayant pour objectif de comparer la France aux pays considérés comme les plus avancés dans le domaine de l'évaluation des politiques publiques (outre les États-Unis, le Royaume-Uni, la Suède, l'Allemagne et le Canada).

Ces monographies reposent sur les résultats de [l'analyse bibliométrique publiée par France Stratégie en décembre 2018](#), sur des analyses documentaires (littérature scientifique et documentation institutionnelle) et sur des entretiens avec les acteurs nationaux de l'évaluation. L'influence des évaluations est examinée selon trois critères : l'ouverture du milieu administratif aux chercheurs en économie ; l'existence de dispositifs formels favorisant la mise en œuvre d'évaluations d'impact ; enfin, le degré de diffusion et d'influence de ces évaluations sur le public et sur les décideurs.

Aux États-Unis, l'évaluation des politiques publiques est une activité ancienne que l'on peut dater des années 1960. Des impulsions politiques fortes émanant de l'exécutif ont favorisé le développement des évaluations d'impact dans quasiment tous les domaines de la politique publique. Récemment, 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.

Les demandes d'évaluations d'impact de politiques publiques émanent principalement des ministères et des agences fédérales, en particulier les secteurs du marché du travail, de l'éducation et de la santé. L'externalisation est très largement pratiquée dans l'administration auprès d'acteurs multiples qui se réunissent en consortium pour répondre aux commandes. Les fondations privées financent également de nombreuses évaluations d'impact. Les principaux producteurs sont les grandes universités nationales en économie. Plusieurs grands organismes de recherche indépendants sont très souvent sollicités par le gouvernement pour évaluer ses programmes (comme MDRC ou *Mathematica Policy Research*).

Les États-Unis ont développé un cadre formalisé pour l'évaluation comprenant des instances de coordination et des outils de programmation. Certains ministères sont ainsi dotés d'unités d'études et d'évaluation ayant un rôle avancé de pilotage et de coordination pour leurs différentes agences. La loi FEBPA de 2018 prévoit pour chaque agence fédérale la mise en place de *Learning Agendas* (outils de programmation) ainsi que la nomination d'un chef évaluateur (*Evaluation Officer*). Les États-Unis disposent également de « standards de preuve » pour juger de la qualité méthodologique des évaluations produites en fonction des méthodes utilisées.

La publication des travaux d'évaluation d'impact est un principe général mais de portée inégale au sein des administrations. Il devrait se généraliser dans l'administration à la suite de l'application de la loi FEBPA de 2018. Certaines administrations, universités et fondations ont

développé des outils de diffusion, appelés Clearinghouses, dont le but est de centraliser les résultats des évaluations et de les rendre accessibles au plus grand nombre.

Enfin, les États-Unis ont récemment mis en place de nouveaux outils de financement de programmes ou dispositifs, qui s'appuient sur des méthodes rigoureuses d'évaluation, comme les méthodes expérimentales ou quasi expérimentales. Il s'agit des mécanismes de type *Pay for Success* (PFS) et *Tiered-Evidence Grants*.

Mots clés : évaluation d'impact, politiques publiques, *ex post*, *ex ante*, administration, États-Unis

L'évaluation des politiques publiques aux États-Unis

Rozenn Desplatz¹

Introduction : le contexte institutionnel²

Une longue histoire de l'évaluation aux États-Unis

L'évaluation des politiques publiques est une activité ancienne aux États-Unis puisqu'elle a véritablement commencé dans les années 1960 à l'occasion de l'évaluation des grands programmes sociaux de lutte contre la pauvreté mis en place par l'administration Johnson. Elle trouve à l'époque sa justification dans la forte augmentation des dépenses publiques qui découle de ces programmes. Elle s'est fortement développée tant au niveau fédéral qu'au niveau des États.

L'approche de l'évaluation – sa définition, les méthodes associées – a également évolué au cours du temps (voir Annexe 2). D'abord conçue comme une activité consistant à mesurer les résultats propres des programmes (l'évaluation causale), l'évaluation s'est ensuite développée sous des formes diverses pour juger plus largement de la mise en œuvre, de la pertinence et de l'efficacité des programmes. Les méthodes associées ont en conséquence évolué : reposant essentiellement sur les méthodes quantitatives dans les années 1960, en particulier sur les expérimentations aléatoires contrôlées (RCT), elles accordent aujourd'hui une place plus grande aux méthodes qualitatives et participatives³.

Selon l'International Atlas of Evaluation⁴, deux grandes vagues peuvent être distinguées dans le développement des évaluations aux États-Unis sur la période 1960-2000. La première

¹ Experte référente en évaluation des politiques publiques à France Stratégie au moment de la rédaction de cette étude. Remerciements pour les relectures et les compléments apportés aux versions antérieures : Sandra Debu, Clément Lacouette-Fougère et Alain Trannoy.

² La rédaction de ce document de travail s'appuie sur la grille d'analyse établie collectivement pour le benchmark des six pays et disponible en Annexe 1. Elle se fonde sur une analyse de la littérature disponible, la consultation des sites internet des institutions concernées et plusieurs entretiens dont la liste figure dans l'annexe Références et sources.

³ Perret (2008).

⁴ Furubo, Rist et Sandahl (2002).

période de 1960 à 1980 est caractérisée par une forte demande d'évaluations et par la mise en place de structures d'évaluation tant du côté du pouvoir exécutif que du côté législatif. La seconde de 1980 à 2000 est marquée par une baisse importante des moyens consacrés à l'évaluation. Plus récemment, l'évaluation a reçu de nouvelles marques d'attention lors des présidences de Bush et Obama. Sous l'administration Bush, les efforts en direction de l'évaluation se sont surtout traduits par la mise en œuvre, en 2004, d'un outil de revue de l'efficacité des programmes fédéraux, appelé Program Assessment Rating Tool (PART). Sous l'administration Obama, plusieurs directives de l'Office of Management and Budget (OMB) à l'attention des ministères ont demandé aux responsables des agences fédérales de développer les activités d'évaluation et de mieux utiliser les résultats pour définir et réformer les politiques publiques. Cette nouvelle impulsion s'est concrétisée par une série d'initiatives, comme la promotion d'évaluations rapides et peu coûteuses et la mise en place de nouveaux mécanismes de financement de dispositifs fondés sur la « preuve », comme les « tiered-evidence grants » ou le « Pay for success » (voir plus loin).

Sous l'administration Trump, l'importance des politiques fondées sur la preuve est réaffirmée dans le budget fédéral de 2018 et 2019 : « *An effective and efficient Federal government requires evidence – evidence about where needs are greatest, what works and what does not work, where and how programs could be improved, and evidence about how programs of yesterday may no longer be suited for today or prepare us for tomorrow* »⁵. Très récemment, une nouvelle loi, **Foundations for Evidence-Based Policymaking Act (FEBPA) of 2018**⁶, a été votée pour améliorer l'accès aux données sensibles, renforcer les protections entourant la confidentialité des données et la capacité du gouvernement fédéral à produire et à utiliser les éléments de preuve. Cette loi reprend pour partie les recommandations formulées par une commission mise en place sous l'administration Obama, la Commission on Evidence-Based Policymaking (CEP)⁷, dans un rapport⁸ publié en septembre 2017.

Outre le niveau fédéral, les États sont également actifs dans ce domaine à des degrés cependant divers. Le développement de l'évaluation au niveau des États a pris appui dans les années 1980 et 1990 sur la dérogation accordée par le gouvernement fédéral d'utiliser les fonds fédéraux pour expérimenter des interventions innovantes dans le domaine des politiques sociales. Cette possibilité de dérogation, accordée en vertu de l'article 115 de la loi sur la sécurité sociale⁹, leur est offerte à la condition que ces nouveaux dispositifs fassent l'objet d'une évaluation rigoureuse (à l'aide par exemple des méthodes d'assignation aléatoire, RCT).

⁵ OMB (Office of Management and Budget), Budget of the U.S. Government, Analytical Perspectives (2019).

⁶ <https://bipartisanpolicy.org/blog/congress-provides-new-foundation-for-evidence-based-policymaking/>.

⁷ La CEP était composée de quinze membres (académiques, experts, responsables de programmes). Dans le rapport remis en septembre 2017 au Président et au Congrès, la commission a formulé vingt-deux recommandations pour améliorer l'accès des chercheurs aux données publiques : renforcer les protections relatives à la confidentialité des données ; mettre en place une infrastructure permettant d'accéder aux données et de réaliser des appariements de données sécurisés ; renforcer la capacité d'analyse et d'évaluation du gouvernement. En particulier, la commission a recommandé la mise en place au sein de chaque ministère d'un chef évaluateur (Chief Evaluation Officer, CEO) et d'une programmation pluriannuelle de travaux axée sur des questions prioritaires de recherche et de politique publique (Learning Agendas).

⁸ Commission on Evidence-Based Policymaking (2017).

⁹ Social Security Act, Section 115.

Les cinq États les plus actifs aujourd'hui dans le domaine de l'évaluation sont les États de Washington, de l'Utah, du Minnesota, du Connecticut et de l'Oregon¹⁰.

Principaux acteurs

L'évaluation est fortement légitimée et institutionnalisée aux États-Unis à la fois du côté de l'exécutif et du côté législatif. Les principales institutions ont été mises en place sur la période 1960-1980. Pendant cette période, la branche exécutive a commencé à financer des évaluations importantes dans le domaine de la politique sociale (éducation, santé, criminalité et logement). Pour ne pas dépendre exclusivement de l'expertise de la branche exécutive, le Congrès s'est de plus en plus appuyé sur le General Accounting Office (GAO) qui a vu l'évaluation apparaître dans ses missions à la fin des années 1960 (voir Annexe 3).

Au sein de la branche exécutive, la fonction d'évaluation est organisée et structurée très différemment dans chaque ministère¹¹. Les moyens humains et financiers sont en outre très inégaux entre les ministères. Certains sont dotés d'unités d'études et d'évaluation particulièrement développées, qui ont un rôle avancé de pilotage et de coordination de ces activités pour les différentes agences du ministère et qui travaillent avec de nombreux acteurs extérieurs (autres ministères, prestataires, chercheurs). Selon un classement des agences fédérales établi en 2018 par Results for America¹², une organisation américaine à but non lucratif, les principaux bureaux d'évaluation sont les suivants :

- l'**Office of Planning, Research, and Evaluation, OPRE** (Administration for Children and Families, U.S. Department of Health and Human Services, HHS), en charge des activités d'évaluation, de recherche et d'études de l'Administration for Children and Families. Il comprend 61 personnes réparties dans quatre divisions et est doté d'un budget de 165 millions de dollars en 2018 ;
- l'**Institute of Education Sciences (IES)** (U.S. Department of Education, ED) qui est chargé de produire des statistiques et de conduire des recherches et des évaluations dans le domaine de compétence du ministère. Il s'appuie sur quatre centres dirigés par des commissaires, dont le **National Center for Education Evaluation and Regional Assistance (NCEE)**¹³ pour les activités d'évaluation. L'IES comprend 170 personnes et est doté d'un budget de 613,5 millions de dollars en 2018. Le NCEE compte environ 20 personnes et est doté d'un budget de 38 millions de dollars en 2018 ;
- l'**Office of Policy Development and Research (PD&R)** (U.S. Department of Housing and Urban Development, HUD), l'unité en charge de produire des statistiques, des études et de la recherche ainsi que des évaluations pour le ministère HUD. Il comprend plusieurs bureaux, dont l'Office of Research, Evaluation and Monitoring au sein duquel se trouve une division spécialisée sur l'évaluation de programmes (Program Evaluation Division). PD&R comprend 139 personnes et est doté d'un budget de 89 millions de dollars en 2018 ;
- le **Chief Evaluation Office (CEO)** (U.S. Department of Labor, DOL) qui pilote ses propres évaluations et coordonne plus généralement l'ensemble des évaluations du ministère du

¹⁰ Pew-Mac Arthur First Initiative (2017).

¹¹ OMB, *Overview of Federal Evidence-Building Efforts* (2016).

¹² Results for America, *Invest in what works federal standard of excellence* (2018).

¹³ Les trois autres centres de l'IES sont le National Center for Education Research (NCER), le National Center for Education Statistics (NCES) et le National Center for Special Education Research (NCSER).

travail (conduites par les agences du ministère elles-mêmes comme l'Employment and Training Administration (ETA)). Le CEO comprend 15 personnes et a un budget de 21,4 millions de dollars en 2018.

Certaines agences américaines d'aide internationale au développement sont également dotées de bureaux d'évaluation puissants (Millenium Challenge Corporation, MCC ; United State Agency for International Development, USAID).

La demande d'évaluation de dispositifs fédéraux émane soit du Congrès, soit des ministères eux-mêmes. Les évaluations d'impact sont pour la plupart confiées à des tiers, organismes de recherche ou universités, qui peuvent également se regrouper en consortiums pour les programmes importants (voir ci-dessous les exemples d'évaluations emblématiques).

L'**Office of Management and Budget (OMB)**, administration en charge de la préparation du budget aux États-Unis, soutient les activités d'évaluation au sein du gouvernement. Il 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. Dans plusieurs de ses *memorandum*¹⁴, l'OMB a également appelé les responsables des agences fédérales à mettre en place de nouvelles initiatives pour renforcer la place des évaluations dans les agences (définition d'un programme d'évaluations, utilisation de nouveaux modes de subvention de programmes pour les États reposant davantage sur les résultats des évaluations, mise en place d'outils de capitalisation et de diffusion des évaluations, etc.). L'OMB organise également des ateliers et des formations à destination des agences fédérales, produit des lignes directrices sur l'évaluation comme celles sur le suivi et l'évaluation des programmes d'aide au développement international¹⁵ et héberge un groupe intergouvernemental sur l'évaluation (l'**Interagency Council on Evaluation Policy, ICEP**).

Au sein du gouvernement, l'évaluation est considérée comme un moyen pour améliorer les programmes fédéraux. Elle fait partie d'un large ensemble d'outils concourant à renforcer les éléments de preuve, aux côtés de la mesure de la performance, de l'analyse des données et de la recherche. La préférence est cependant donnée aux études expérimentales et quasi expérimentales de « haute qualité » quand il s'agit de mesurer un lien de cause à effet entre un dispositif et ses résultats^{16,17}.

Côté législatif, le Congrès s'appuie sur trois instances d'expertise : le **Government Accountability Office (GAO)** créé en 1921, le **Congressional Research Service (CRS)**¹⁸

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

¹⁵ OMB (2018) : M-18-04.

¹⁶ OMB (2013) : M-13-17. Selon ce memorandum, « *Such evidence includes evaluation results, performance measures, and other relevant data analytics and research studies, with a preference for high quality experimental and quasi-experimental studies* » (page 2).

¹⁷ OMB, Budget of the U.S. Government, Analytical Perspectives (2019). Les éléments de preuve sont ainsi définis : « *Government agencies should use a range of evidence types and analytical and management tools to learn what works and what does not, for whom and under what circumstances, and how to improve results. Evidence refers to facts or information indicating whether a belief or proposition is true or valid. Evidence can be quantitative or qualitative and may come from a variety of sources, including performance measurement, program evaluations, statistical series, retrospective reviews, data analytics, and other science and research. A portfolio of evidence may include: impact evaluations [...], process or implementation evaluations [...], performance monitoring and measurement [...], statistics and other forms of research and analysis [...]* ».

¹⁸ Le CRS est composé de plus de 400 employés, parmi lesquels on compte des analystes de politiques publiques, des avocats, des professionnels de l'information dans des domaines très variés (le droit,

créé en 1946 et le **Congressional Budget Office (CBO)**¹⁹ créé en 1974. Les agences du Congrès ne réalisent pas d'évaluation scientifique causale. Elles peuvent s'appuyer sur celles réalisées ou commandées par l'exécutif. Le GAO est la plus importante des trois agences rattachées au Congrès. Il est doté d'un budget de plus de 575 millions de dollars et compte près de 3 000 personnes. Il comprend quatorze équipes thématiques et onze bureaux régionaux. L'une des équipes comprend une unité de sept ou huit personnes sur l'évaluation²⁰ chargée d'accompagner les autres équipes sur les questions méthodologiques. Les évaluations de programmes réalisées par le GAO consistent en des audits de performance réalisés à l'aide de méthodes qualitatives (entretiens, questionnaires, observations, etc.) ou quantitatives (exploitation de données). Le GAO réalise des études à la demande des membres du Congrès mais peut aussi s'autosaisir. Il collecte et analyse également des données. Le GAO a environ 900 études en cours. Tous les rapports qu'il produit sont destinés à être publics.

Le milieu professionnel est également structuré et influent. L'**American Evaluation Association (AEA)**, fondée en 1986, compte plusieurs milliers de membres (7 300 membres provenant des 50 États des États-Unis et de 80 pays étrangers). Elle contribue à l'élaboration de référentiels et de standards (principes directeurs s'appliquant aux évaluateurs), dispense des formations, publie des revues (*American Journal of Evaluation* auparavant appelé *Evaluation Practice, New Directions for Evaluation*) et tient des congrès scientifiques importants (voir par exemple Evaluation 2018).

Plusieurs fondations financent et promeuvent les évaluations scientifiques *ex post*, comme la **MacArthur Foundation** fondée en 1970 ou **Laura and John Arnold Foundation (LJAF)** fondée en 2008 et faisant partie d'Arnold Ventures. Ces dernières financent et promeuvent presque exclusivement les méthodes expérimentales (RCT) comme approches pouvant apporter des preuves rigoureuses²¹. La fondation LJAF soutient par ailleurs une coalition de chercheurs, **Evidence-Based Policymaking Collaborative**, issus de différents organismes et instituts (Urban Institute, Brookings Institution, American Enterprise Institute, Pew Charitable Trusts et MacArthur Foundation), dont le but est de formuler des propositions pour développer l'élaboration de politiques publiques fondées sur la preuve, notamment au niveau fédéral. La fondation LJAF développe également une base de données sur les programmes sociaux qui

l'économie, les affaires étrangères, la sécurité nationale, l'administration publique, l'éducation, la santé, l'immigration, l'énergie, l'environnement, les sciences et la technologie). Au service du Congrès, il analyse les politiques publiques en cours et les réformes envisageables. Ses travaux ne sont pas publics.

¹⁹ Le CBO est composé de 235 employés à temps plein, pour l'essentiel des économistes et des spécialistes des politiques publiques américaines dans les domaines de la santé mais aussi de la sécurité nationale, du travail, de la fiscalité, de l'énergie et de la macroéconomie. Le CBO élabore, pour le compte du Congrès, principalement des projections économiques et budgétaires sur 10 ans et 30 ans et produit des estimations de coût des projets de loi et des options législatives envisageables. Pour garantir son objectivité, le CBO ne fait pas de recommandations de politique économique. Pour plus de détails, voir www.cbo.gov/about/10-things-to-know.

²⁰ Il s'agit du « Center for Evaluation Methods and Issues » in « Applied Research and Methods ». Il est à noter que le GAO s'était doté dans les années 1980 d'une direction spécialisée dans l'évaluation de programmes (Institute for Program Evaluation rebaptisé en 1983 Program Evaluation and Methodology Division, PMED), composée d'une centaine de chercheurs en sciences sociales. Elle conduisait des évaluations de programmes et proposait un accompagnement technique. Cette direction a été supprimée en 1996, suite à des restrictions budgétaires et au départ à la retraite de son directeur (Furubo, Rist et Sandahl, 2002).

²¹ Voir en particulier les [projets financés en janvier 2019 par l'Evidence-Based Policy Team](#) de la fondation LJAF.

ont apporté la preuve de leur efficacité, appelée *Social Program that Works* (voir plus loin). La fondation MacArthur soutient pour sa part deux initiatives, *Pew-MacArthur 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) et *Actionable Intelligence for Social Policy initiative (AISP)* à l'université de Pennsylvanie pour apparier les données administratives des agences fédérales entre elles à un niveau individuel.

Il existe également aux États-Unis d'importants organismes de recherche indépendants, producteurs d'évaluations d'impact²², comme le **Washington State Institute for Public Policy (WSIPP)** au niveau local créé en 1983, **MDRC** fondé en 1974 et **Mathematica Policy Research Group** créé en 1968 (voir Axe 1 pour plus de détails) et plusieurs think tanks influents ayant diverses activités dont celle d'évaluer les politiques publiques (**Rand Corporation, Brookings Institution, Urban Institute**).

Le milieu académique est également très dynamique. Les principales universités productrices d'évaluations d'impact sont aussi les meilleures universités américaines (Stanford, MIT, Harvard, Princeton, Chicago, Yale, Columbia, Pennsylvania, Duke, Cornell, etc.) (voir Axe 1).

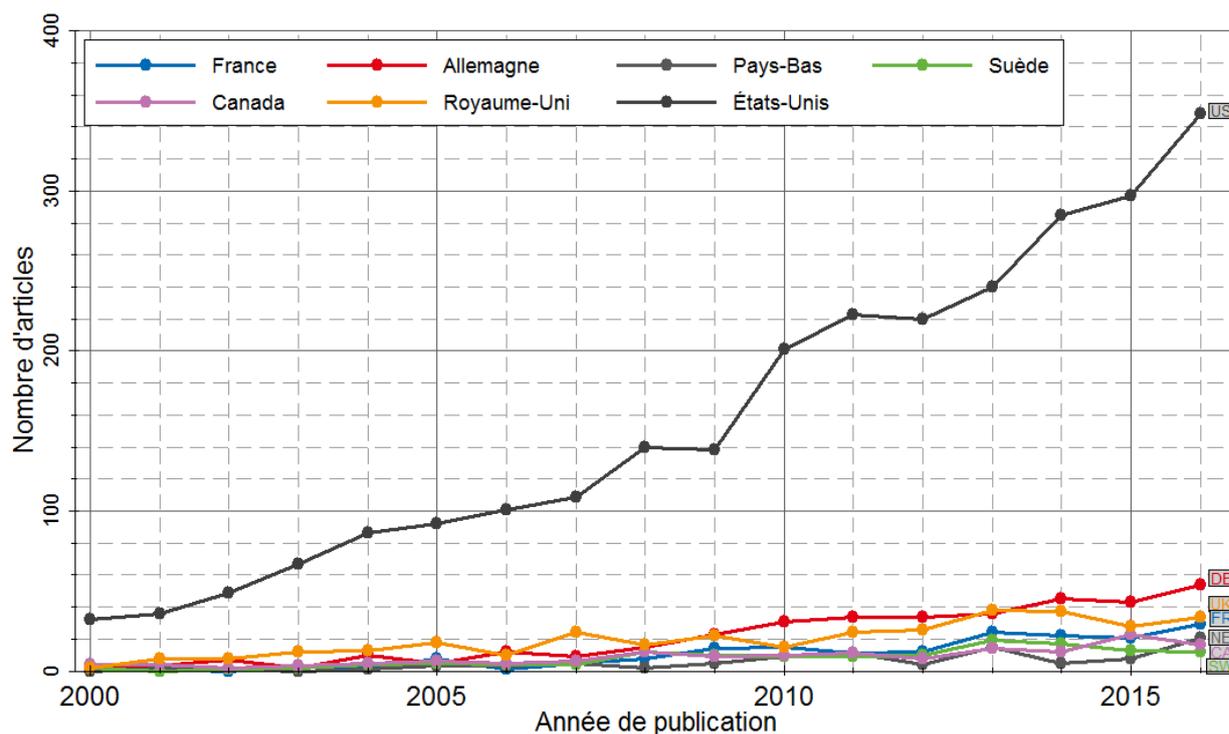
²² Lenihan (2013).

AXE I – LA PRODUCTION D'ÉVALUATIONS D'IMPACT

1. Un volume très élevé de travaux sur les États-Unis

Dans notre étude bibliométrique²³, qui se restreint aux publications dans des revues à comité de lecture, près de 2 900 évaluations d'impact ont été identifiées sur les États-Unis, ce qui en fait le producteur le plus important – loin devant la Chine, le Royaume-Uni, l'Allemagne, l'Italie et l'Inde. Si on rapporte le nombre de publications à la population, les États-Unis apparaissent en quatrième position, derrière les pays d'Europe du Nord (Norvège, Danemark, Suède) et devant notamment la Suisse, les Pays-Bas, le Royaume-Uni, l'Allemagne et le Canada. Les premières évaluations d'impact sont anciennes aux États-Unis, puisqu'elles datent des années 1960. Depuis, le flux d'évaluations annuelles portant sur les États-Unis s'accroît chaque année sensiblement : en 2016, près de 350 nouvelles évaluations ont été réalisées, contre 200 en 2010 et près de 100 en 2005 (voir Figure 1).

Figure 1 – Évolution du nombre d'articles d'évaluation d'impact publiés pour les pays de l'échantillon entre 2000 et 2016



Source : EconLit, traitement des auteurs, Bono et al. (2018).

²³ La base de données EconLit recense les publications académiques en sciences économiques depuis 1886. Elle est alimentée et mise à jour par les éditeurs de revue qui signalent les nouvelles publications. L'analyse est restreinte aux seuls articles scientifiques (publiés dans des revues à comité de lecture). Elle exclut en particulier les documents de travail. Voir Bono, Debu, Desplatz, Hayet, Lacouette-Fougère et Trannoy (2018).

2. Les universités sont les principales institutions productrices

Les premières institutions productrices d'évaluations d'impact aux États-Unis sont les universités. En effet, parmi les vingt premières institutions productrices d'évaluations d'impact ne figurent que des universités selon la métrique pondérée (où on répartit un article entre ses auteurs et les institutions auxquelles ils appartiennent). En tête du classement se trouvent les universités de Harvard, de Chicago, du Michigan et de Pennsylvanie ainsi que le Massachusetts Institute of Technology (MIT) (voir Figure 2 et Figure 3). Au sein du département d'économie du MIT, on peut citer le **Abdul Latif Jameel Poverty Action Lab North America (J-Pal North America)** qui a été fondé en 2003 dans l'objectif de réduire la pauvreté en s'appuyant sur les résultats de la recherche scientifique. J-Pal est un réseau de 171 professeurs associés venant d'universités du monde entier et spécialisés dans la conduite d'évaluations d'impact randomisées dans des domaines très variés (agriculture, criminalité, éducation, environnement et énergie, finance, entreprises, genre, santé, travail, économie politique et gouvernance). Il dispose aujourd'hui de sept bureaux régionaux, dont l'un en Amérique du Nord (J-Pal North America) situé au sein du MIT et un autre en Europe (J-Pal Europe) situé au sein de l'École d'économie de Paris (PSE). J-Pal a trois activités principales : la production d'évaluations d'impact à l'aide des méthodes expérimentales ; la formation des décideurs publics aux méthodes d'évaluation d'impact aléatoires (RCT) ainsi que la dissémination des résultats des évaluations aux décideurs (conférences, ateliers, publications).

Selon l'autre métrique (métrique non pondérée), d'autres organisations se distinguent, comme la RAND Corporation et le National Bureau of Economic Research (NBER). Créée en 1948, la **RAND Corporation** est un organisme indépendant à but non lucratif de conseil et de recherche au service de la décision politique et économique. Basée en Californie, elle emploie 1 175 chercheurs présents dans 50 pays, dont 53 % détiennent un doctorat et 38 % un diplôme de master dans des disciplines très diverses. Ces chercheurs sont répartis dans quatre départements de recherche, organisés en groupes de discipline (sciences comportementales et politiques ; défense et sciences politiques ; économie, sociologie et statistiques ; ingénierie et sciences appliquées). Les domaines de recherche sont très variés et couvrent : les enfants, la famille, les communautés ; l'énergie et l'environnement ; les infrastructures et les transports ; la sécurité nationale et le terrorisme ; les sciences des données ; la santé et le vieillissement ; les affaires internationales ; la science et la technologie ; l'éducation ; la sécurité publique ; le droit et les affaires ; le travail. La RAND Corporation publie le *RAND Journal of Economics*, une revue à comité de lecture en économie. Pour sa part, le **NBER**, fondé en 1920, est le plus grand organisme de recherche en économie des États-Unis, regroupant plus de 1 400 professeurs en économie des universités nord-américaines dont plusieurs proviennent du département d'économie de l'université Harvard. Ses travaux sont organisés en vingt programmes de recherche et quinze groupes de travail. Ils s'adressent aux décideurs politiques, au monde des affaires et au monde académique. Ses bureaux sont situés à Cambridge dans le Massachusetts. Il dispose d'une antenne dans la ville de New York.

Il existe également aux États-Unis d'autres organismes de recherche indépendants qui sont des producteurs importants d'évaluations d'impact²⁴ et qui sont très souvent sollicités par les ministères et les agences fédérales pour évaluer leurs programmes :

- au niveau local, le **Washington State Institute for Public Policy (WSIPP)**²⁵ est une organisation non partisane, créée en 1983 et financée par l'État de Washington pour expertiser ses politiques dans les domaines de la justice criminelle, de la santé, de l'éducation et du logement. L'institut réalise des analyses coûts-bénéfices à partir de méta-évaluations reposant sur des méthodologies rigoureuses (méthodes expérimentales principalement). Ses modèles d'analyses coûts-bénéfices pourraient être étendus à plusieurs autres États dans le cadre du projet Results First (RF)²⁶, développé en partenariat pas The Pew Charitable Trusts et MacArthur Foundation ;
- **MDRC** (connu sous le nom de Manpower Demonstration Research Corporation avant 2003) est un organisme de recherche indépendant, créé en 1974 par des agences gouvernementales et la fondation Ford. MDRC conduit des évaluations dans le domaine des politiques sociales à l'aide de différentes méthodes, à la fois qualitatives et quantitatives. MDRC est impliqué dans près de 80 projets dans divers domaines de la politique (comme les familles, l'éducation, la jeunesse, la justice pénale, l'emploi, la santé)²⁷. Il contribue également à développer et à faire connaître les méthodologies d'évaluation, notamment les méthodes expérimentales et quasi expérimentales, ainsi que les méthodes combinées (qualitatives et quantitatives)²⁸. MDRC comprend 250 experts et dispose d'un budget annuel de 90 millions de dollars (en 2017) ;
- **Mathematica Policy Research Group** est un organisme de recherche indépendant créé en 1968 et détenu à 100 % par ses 1 200 employés. Il mène des évaluations dans le domaine des politiques sociales pour le gouvernement et le secteur privé²⁹. Il a notamment conduit des évaluations d'impact à l'aide de méthodes expérimentales pour le ministère du travail (Department of Labor). Il a défini et met à jour régulièrement 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). Il a créé son propre centre, le **Center for Improving Research Evidence (CIRE)**, qui dispense de la formation et une assistance technique à ses clients pour la conception, la mise en œuvre et l'évaluation de dispositifs.

²⁴ Lenihan (2013).

²⁵ www.wsipp.wa.gov.

²⁶ Results First Initiative est un projet qui vise à aider les États à adopter des approches innovantes d'élaboration des politiques publiques fondées sur la preuve.

²⁷ Pour plus de détails sur les projets, voir www.mdrc.org/projects.

²⁸ MDRC dispose d'une [collection spécifique dédiée aux questions méthodologiques](#). MDRC, en collaboration avec six partenaires financiers, a créé en 2004 le fonds Judith Gueron pour l'innovation méthodologique dans le domaine de la recherche sur les politiques sociales. Voir www.mdrc.org/investing-future#overlay-context=about/about-mdrc-overview-0.

²⁹ www.mathematica-mpr.com/our-publications-and-projects/projects#q1=&q2=&keyword=&p=1&sortBy=date&sortdir=desc&facets=&datefacets=

Figure 1 – Top 20 des institutions pour les États-Unis (métrique non pondérée)

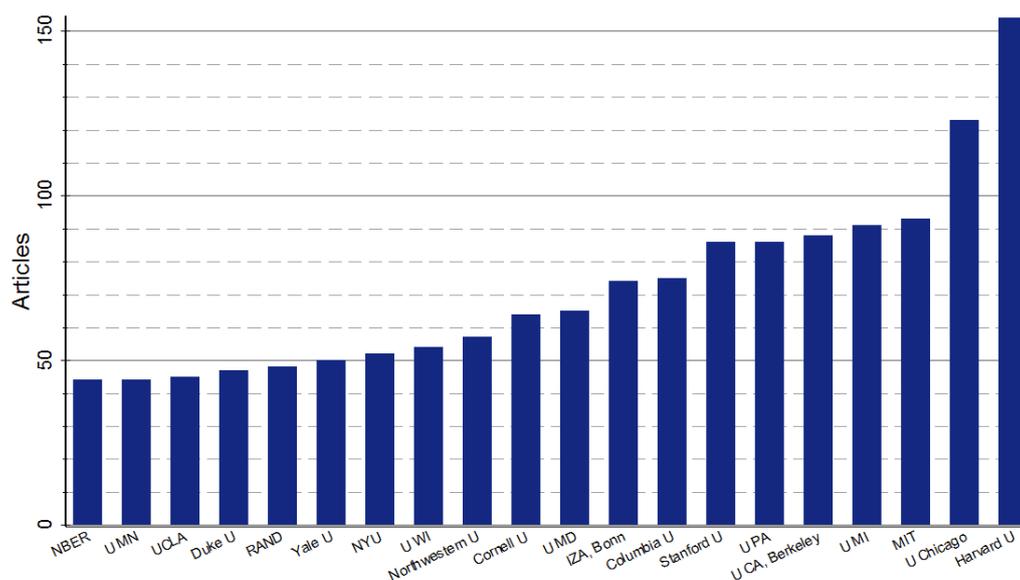
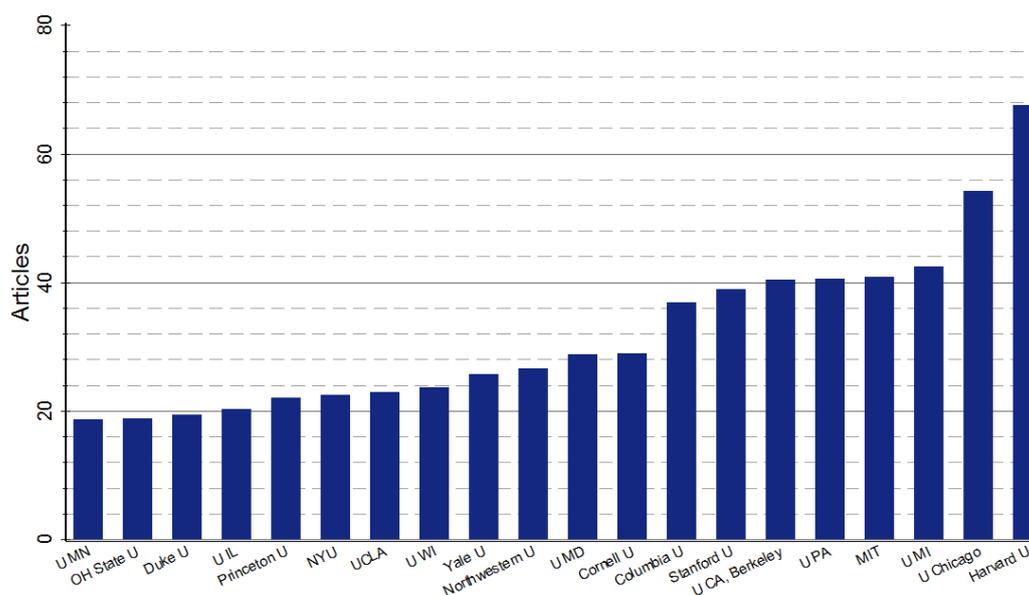


Figure 2 – Top 20 des institutions pour les États-Unis (métrique pondérée)



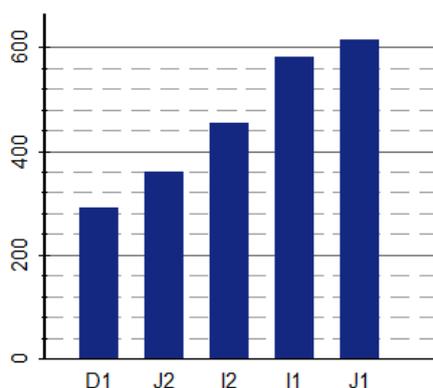
Note : la métrique non pondérée compte les occurrences des institutions, indépendamment du nombre d'auteurs et d'institutions à l'origine de l'article. À l'inverse, la métrique pondérée pondère les occurrences des institutions par le nombre d'auteurs et d'institutions à l'origine de l'article. Elle assure que chaque article « vaut » au total un point, qui est ensuite partagé entre auteurs et entre institutions. Par ailleurs, le classement repose sur les institutions (universités, grandes écoles) et non sur les laboratoires de recherche. Enfin, le comptage est réalisé en cumulé sur l'ensemble de la période, depuis la date d'entrée des revues dans la base EconLit, variable selon les revues, et jusqu'en 2016.

Source : EconLit, traitement des auteurs, Bono et al. (2018).

3. Les évaluations d'impact portent en particulier sur les domaines de la santé et de l'éducation

Les domaines les plus étudiés sont, dans l'ordre, l'économie de la population (J1), la santé (I1), l'éducation (I2), le marché du travail (J2) et le comportement des ménages et l'économie de la famille (D1) (voir Figure 4). Par ailleurs, parmi les trois premières revues publiant des évaluations d'impact sur les États-Unis se trouvent deux revues généralistes en économie (*Review of Economics and Statistics*, *American Economic Review*) et une revue spécialisée dans les questions de santé (*Journal of Health Economics*).

Figure 3 – Top 5 des codes JEL pour les États-Unis



Note de lecture : nombre d'apparitions d'un code JEL de catégorie 2 jusqu'en 2016 (en cumulé sur l'ensemble de la période depuis la date d'entrée des revues dans la base).

Liste des codes JEL : **D1**-Comportement des ménages et économie de la famille ; **I1**-Santé; **I2**-Éducation, enseignement supérieur et recherche; **J1**-Économie démographique; **J2**-Demande et offre de travail.

Source : *EconLit*, traitement des auteurs, [Bono et al. \(2018\)](#).

Les évaluations d'impact sont réalisées à la demande du Congrès, des ministères et des agences fédérales ou des fondations. Les exemples figurant dans l'encadré ci-dessous ont été indiqués par les personnes interrogées lors des entretiens. Ils montrent que les programmes fédéraux sont bien souvent évalués par les grands organismes de recherche indépendants en partenariat avec des cabinets privés. Réalisées sur plusieurs années, les évaluations comprennent différents types de travaux portant sur la mise en œuvre, sur les effets du programme à différents horizons et parfois sur les coûts financiers attachés aux programmes au regard de leurs effets (analyse coûts-bénéfices).

Encadré 1 – Évaluations d'impact emblématiques

Workforce Investment Act (WIA) Adult and Dislocated Worker Programs Gold Standard Evaluation

Votée en 1998, la Workforce Investment Act (WIA) définit les deux principaux programmes d'accompagnement et de formation pour les demandeurs d'emploi, celui pour les adultes et celui ciblé sur certaines catégories de travailleurs comme les salariés licenciés suite à la fermeture de leur entreprise ou à une réduction substantielle de leur effectif (Adult and Dislocated Worker programs). Elle a été remplacée en 2014 par la Workforce Innovation and Opportunity Act

(WIOA), qui a cependant maintenu en l'état ces deux programmes (mêmes prestations, mêmes populations servies).

En 2008, l'Employment Training Administration (ETA), l'une des agences fédérales du ministère du travail (DOL), a décidé de lancer une évaluation nationale aléatoire de ces deux programmes (The WIA Adult and Dislocated Worker Programs Gold Standard Evaluation). L'objectif de cette évaluation est triple : décrire la mise en œuvre de ces deux programmes ; fournir une évaluation nationale de leurs effets notamment sur le retour à l'emploi et les salaires des demandeurs d'emploi et enfin établir une analyse coûts-bénéfices des deux programmes. L'évaluation a été confiée, pour une durée totale de sept ans, à un consortium d'acteurs, piloté par Mathematica Policy Research et rassemblant également Social Policy Research Associates (SPR), MDRC et Corporation for a Skilled Workforce (CSW). Trois rapports ont été publiés récemment : un premier rapport publié en 2015 sur la mise en œuvre des programmes (D'Amico *et al.*) ; un second en 2016 sur les effets à 15 mois de ces programmes (McConnell *et al.*) et un troisième en 2017 sur leurs effets à 30 mois (Fortson *et al.*).

L'évaluation repose sur un tirage aléatoire de 28 zones géographiques parmi les 487 couvrant les différents États et, au sein de chaque zone, des demandeurs d'emploi dans un des trois groupes définis (deux groupes de traitement et un groupe de contrôle selon les prestations reçues). Ces groupes ont été constitués entre novembre 2011 et avril 2013. L'évaluation repose sur des données d'enquête collectées sur les différents groupes à 15 mois et à 30 mois après le tirage aléatoire et sur des données administratives (National Directory of New Hires, NDNH) portant sur l'emploi, les salaires et l'indemnisation chômage mobilisées 36 mois après le tirage aléatoire. Les résultats à 30 mois montrent que l'accompagnement intensif accroît les salaires ultérieurs (de 20 %) mais que la formation n'aurait pas d'effet positif sur les salaires.

Mother and Infant Home Visiting Program Evaluation (MIHOPE)

MIHOPE est l'évaluation nationale requise par la loi du programme de visites à domicile appelé Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV), créé en 2010 par la Patient Protection and Affordable Care Act. Elle a été lancée en 2011 par l'Administration for Children and Families (ACF) en collaboration avec une autre agence du ministère de la santé, Health Resources and Services Administration (HRSA, Department of Health and Human Services, HHS).

MIHOPE s'intéresse à quatre dispositifs remplissant les critères d'efficacité définis par le ministère de la santé : Early Head Start – Home-based option, Healthy Families America, Nurse-Family Partnership et Parents as Teachers. Elle comporte quatre études : une évaluation d'impact par assignation aléatoire pour estimer les effets sur une large gamme de variables d'intérêt précisées dans la loi créant le programme MIECHV ; une évaluation de mise en œuvre décrivant de manière détaillée les services fournis aux familles et leurs différences éventuelles ; une analyse des coûts financiers de ces programmes au regard de leurs effets et enfin une analyse des besoins des territoires et des États tels que figurant dans leur candidature initiale de 2010 et 2011. L'évaluation a été réalisée par MDRC en partenariat avec James Bell Associates, Johns Hopkins University, Mathematica Policy Research, University of Georgia et Columbia University. Trois rapports ont d'ores et déjà été publiés dont un premier rapport remis au Congrès en 2015 (Michalopoulos *et al.*), un second sur la mise en œuvre en 2018 (Duggan *et al.*) et un troisième sur les effets des programmes en 2019 (Michalopoulos *et al.*). Un quatrième rapport devrait être publié en 2019 sur les coûts financiers associés aux programmes.

L'évaluation MIHOPE couvre 88 programmes mis en œuvre localement dans 12 États (relevant de l'un des quatre dispositifs précités) et porte sur 4 200 femmes de quinze ans et plus, enceintes ou ayant des enfants de moins de six mois, affectées par tirage au sort dans l'un de ces programmes ou dans le groupe de contrôle entre 2012 et 2015. Elle mobilise plusieurs sources de données (enquêtes auprès des familles et des visiteurs à domicile ; données administratives ;

interviews auprès de différents acteurs ; observations directes des parents, des enfants et des visiteurs à domicile). Les résultats montrent que les programmes de visites à domicile ont été mis en œuvre comme prévu et trouvent des effets positifs sur plusieurs variables d'intérêt pour les enfants âgés de 15 mois (amélioration de l'environnement familial, réduction des agressions psychologiques à l'égard de l'enfant, moins de visites médicales d'urgence, moins de problèmes comportementaux pour l'enfant).

District of Columbia Opportunity Scholarship Program (OSP)

Créé en 2004 par le Congrès, le DC Opportunity Scholarship Program est un programme octroyant des bons (« *vouchers* ») aux familles à faibles revenus pour que leur enfant puisse fréquenter des écoles privées. La loi Scholarships for Opportunity and Results Act (SOAR) de 2011 a prévu de nouveaux crédits pour ce programme, qui n'acceptait plus de nouveaux candidats depuis 2009. Les bons sont d'une valeur de 8 000 dollars par année pour un élève de l'école primaire ou collège (de Kindergarten jusqu'au 8^e grade) et de 12 000 dollars par année pour un élève de lycée (à partir du 9^e grade). L'argent est versé directement aux écoles privées concernées.

La loi SOAR a demandé à l'Institute of Education Sciences (IES) une évaluation indépendante de ce programme et la remise d'un rapport chaque année au Congrès. L'évaluation a été confiée, pour une durée de six ans et un coût total de près de 9,5 millions de dollars, à Westat et à Pemberton Research. Cinq rapports ont déjà été publiés : les trois premiers (Feldman *et al.*, 2014 ; Dymarski *et al.*, 2016 ; Betts *et al.*, 2016) décrivent les caractéristiques des bénéficiaires et des écoles privées choisies, les raisons pour lesquelles les parents décident de participer au programme et en quoi les écoles privées visées diffèrent des écoles publiques où étaient inscrits les élèves ; les deux rapports suivants (Dymarski *et al.*, 2017 et 2018) évaluent les effets du programme respectivement un an et deux ans après l'entrée dans le programme. Un rapport final évaluera les effets à trois ans.

L'évaluation repose sur un tirage aléatoire d'environ 1 800 élèves candidats au programme OSP à un groupe de traitement (pouvant percevoir les bons) et à un groupe de contrôle (ne recevant pas les bons). Trois cohortes ont été constituées (printemps 2012, printemps 2013 et printemps 2014). Les données utilisées reposent notamment sur des données d'enquêtes collectées pour chacune des cohortes jusqu'à trois ans après leur entrée dans la cohorte et des données administratives sur les évaluations en mathématiques et en lecture des élèves. Les résultats à deux ans montrent que le programme OSP a un effet négatif sur les performances en mathématiques des élèves mais un effet positif sur la sécurité dans l'école telle qu'elle est perçue par les parents et les élèves. Par ailleurs, le programme n'a pas d'effet significatif sur les performances en lecture des élèves, sur la satisfaction des parents et des élèves vis-à-vis de l'école ou encore sur l'implication des parents dans l'éducation de leur enfant à l'école ou à la maison.

Références

Mathematica Policy Research : D'Amico *et al.* (2015) ; Fortson *et al.* (2017) et McDonnell *et al.* (2016).

Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation : Michalopoulos *et al.* (2015) ; Duggan *et al.* (2018) ; Michalopoulos *et al.* (2019).

Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance : Betts *et al.* (2016) ; Dymarski *et al.* (2016, 2017, 2018) ; Feldman *et al.* (2014).

AXE II – PLACE DES ÉVALUATIONS D'IMPACT DANS LES PROCESSUS DE DÉCISION ET LES PRATIQUES

1. Une ouverture importante du milieu administratif aux universitaires, se traduisant par de nombreuses collaborations

1.1. Des recrutements d'universitaires dans l'administration possibles mais parfois difficiles

Le recrutement d'universitaires au sein du gouvernement fédéral est possible, mais peut se révéler difficile pour des raisons tenant à la fois aux salaires (moindre attractivité) et à la valorisation d'un passage dans une agence gouvernementale dans la suite du parcours universitaire. En effet, la carrière académique repose pour les universitaires sur leur capacité à publier leurs travaux dans les meilleures revues scientifiques, et impose de ce fait une pression importante pour produire des travaux de recherche de grande qualité. Un passage dans une agence gouvernementale peut ralentir cette dynamique de production et de publication d'articles académiques. Pour ces raisons, les personnes ayant un doctorat ou un post-doctorat doivent souvent faire le choix entre une carrière académique à l'université et une carrière administrative au sein du gouvernement.

Toutefois, des passerelles existent entre les milieux académique et administratif qui peuvent être plus ou moins importantes selon les agences gouvernementales. Au niveau fédéral, ces dernières ont notamment 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 qu'ils pourront 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, d'une part, et les entités non fédérales (les États et les collectivités locales, les universités, les centres de recherche, etc.) d'autre part, 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 par la General Services Administration (GSA) (voir ci-dessous) et spécialisé dans l'application des résultats des sciences comportementales aux programmes/dispositifs fédéraux, est composé d'une équipe interdisciplinaire de chercheurs issus notamment de l'université (Harvard, Reed College, université de Washington, etc.) et recrutés pour des périodes courtes (de 1 à 4 ans) à l'aide

du programme IPA³⁰. Ce dernier est cependant plus ou moins utilisé par les agences gouvernementales au niveau fédéral.

1.2. Des formations aux méthodes d'évaluation généralement prévues mais surtout pour les spécialistes

Au sein du gouvernement fédéral, les unités ou bureaux en charge des activités d'évaluation disposent pour les plus importants d'entre eux (voir précédemment) d'un personnel qualifié et formé aux méthodes d'évaluation d'impact (de niveau master, doctorat et post-doctorat). Ces unités prévoient également régulièrement des formations pour mettre à jour les compétences de leur personnel et ainsi lui permettre de suivre les travaux d'évaluation réalisés par des chercheurs. Dans la plupart des cas, les évaluations d'impact de programmes fédéraux sont en effet externalisées et confiées à des chercheurs appartenant à des grands organismes de recherche indépendants (comme MDRC, Mathematica Policy Research) et/ou à des universités. Les formations aux méthodes d'évaluation d'impact sont bien souvent dispensées par des universitaires. J-PAL North America, rattaché au MIT, est un acteur important dans le paysage pour la formation aux méthodes expérimentales (RCT).

En dehors des principaux bureaux d'évaluation ayant un rôle de coordination et de pilotage des évaluations pour leur ministère, la connaissance des méthodes d'évaluation d'impact par les agents fédéraux est plus inégale. Chaque année, l'OMB demande aux agences d'indiquer comment elles utilisent les preuves et quels obstacles elles rencontrent (« evidence submission »). La barrière la plus fréquente concerne les compétences du personnel et sa capacité à produire et à utiliser les preuves. Ce constat a été réitéré récemment dans le rapport³¹ de la Commission on Evidence-Based Policymaking (CEP) qui montre que 54 % des répondants (des agences fédérales) rencontrent des difficultés pour embaucher et retenir du personnel qualifié correspondant aux activités de production de preuves³². Au plus haut niveau de l'administration, les responsables (Senior Executive Service, SES) ne sont pas formés aux méthodes associées, et les qualifications qui leur sont demandées, appelées « Executive Core Qualifications »³³, n'incluent aucun élément sur les preuves, l'évaluation de programmes, les données et l'innovation³⁴. Il en est de même dans les bureaux en charge de la conception et du suivi des dispositifs fédéraux (programs offices) où la connaissance des méthodes, notamment d'évaluation d'impact, peut être faible.

Les principaux bureaux d'évaluation au niveau fédéral ont généralement dans leurs activités une mission d'assistance pour former les autres agences ou bureaux aux méthodes d'évaluation ou pour les aider à évaluer leurs programmes dans le but d'améliorer la qualité des évaluations réalisées. C'est le cas du Chief Evaluation Office (CEO) au sein du ministère

³⁰ https://oes.gsa.gov/assets/files/GSA_OES_SolicitationFY19v2.pdf.

³¹ Commission on Evidence-Based Policymaking (2017).

³² Il s'agit d'une enquête réalisée auprès de 79 agences fédérales déclarant occuper des missions dans les domaines de la statistique (collecte et traitement des données), de la recherche, de l'évaluation et de l'analyse des politiques. Voir page 91 du rapport.

³³ Ces qualifications de base sont au nombre de cinq : Leading Change (ECQ 1) ; Leading People (ECQ 2) ; Results Driven (ECQ 3) ; Business Acumen (ECQ 4) ; Building Coalitions (ECQ 5). À chacune de ces qualifications correspondent plusieurs compétences associées. Voir www.opm.gov/policy-data-oversight/senior-executive-service/executive-core-qualifications/#url=Overview.

³⁴ Feldman (2017).

du travail (Department of Labor, DOL) qui accompagne les autres agences du ministère ayant une activité d'évaluation pour garantir la qualité des évaluations qu'elles financent en propre. C'est aussi le cas du National Center for Education Evaluation and Regional Assistance (NCEE) de l'Institute of Education Sciences (IES) au sein du ministère de l'éducation (Department of Education) dont les agents dispensent des formations à ceux chargés de la conception et du suivi des dispositifs fédéraux. Ces dernières sont encore peu fréquentes.

Certaines agences ont également développé des ressources en ligne ouvertes à tous et concernant l'évaluation en général³⁵.

Enfin, on peut souligner le rôle central de l'OMB qui prévoit des ateliers et des séminaires d'échanges sur l'évaluation à destination des agences fédérales. La nouvelle loi, Foundations for Evidence-Based Policymaking Act (FEBPA) de 2018 (voir précédemment), renforce ce rôle puisqu'elle a demandé à l'OMB de produire en collaboration avec l'ICEP, dans l'année suivant la promulgation de la loi, des lignes directrices sur l'évaluation pour aider les agences fédérales à mener leurs activités d'évaluation conformément aux standards généralement admis. Elle a également demandé à l'Office of Personnel Management (OPM), en collaboration avec l'OMB, d'identifier les compétences nécessaires à l'exercice des activités d'évaluation de programmes, de définir les professions correspondantes et les parcours professionnels associés au sein des agences. Aujourd'hui, s'il existe des métiers de statisticiens, d'économistes, de spécialistes de l'information, d'analystes de politiques publiques au sein de l'administration américaine, il n'y a cependant pas de métier associé à l'évaluation de programmes³⁶.

1.3. Des partenariats nombreux et divers entre l'administration et les universitaires

Pour les principaux bureaux d'évaluation, les partenariats avec le milieu académique sont nombreux et variés. Ils s'appuient sur les universitaires et les chercheurs d'organisations de recherche indépendantes pour la réalisation des évaluations d'impact des dispositifs fédéraux par le biais d'appels d'offres (voir les exemples cités précédemment). Ils peuvent aussi octroyer des subventions aux académiques, pour des durées plus courtes, pour les inciter à travailler sur les sujets d'intérêt pour le ministère. Par exemple, le CEO (DOL) a attribué récemment près de deux millions de dollars à neuf équipes universitaires de chercheurs pour une durée de deux ans (janvier 2017-janvier 2019) pour des activités d'évaluation et de recherche (Labor Research and Evaluation Grants)³⁷. Ils mettent également parfois en place des groupes de travail techniques composés d'universitaires pour accompagner les équipes de recherche en charge d'évaluer les programmes fédéraux (voir plus loin). Enfin, ils

³⁵ [USAID Learning Lab Evaluation Toolkit](#) (U.S. Agency for International Development) ; EPA Logic Model Training (U.S. Environmental Protection Agency) : www3.epa.gov/evaluate/lm-training/ ou www.epa.gov/evaluate ; [CNCS Evaluation Resources](#) (U.S. Corporation for National and Community Service) ; [CDC Evaluation Resources](#) (Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, HHS).

³⁶ Voir le rapport de la Commission on Evidence-Based Policymaking à la page 91 : « *Today, while the U.S. Office of Personnel Management (OPM) recognizes statisticians, economists, information specialists, and policy analysts as official and distinct occupations within government, no occupational series exists specifically for the field of program evaluation* ». Voir aussi le [Handbook of occupational groups and families](#) (décembre 2018).

³⁷ [Labor Research and Evaluation Grants](#) (CEO, DOL).

organisent des conférences auxquelles participent les universitaires pour diffuser et débattre des résultats des évaluations.

Au-delà de ces formes communes de partenariat, une initiative plus novatrice est celle de l'**Office of Evaluation Sciences (OES)** au sein de la General Services Administration (GSA). Composé d'une équipe d'une quinzaine de chercheurs appliqués issus de différentes disciplines (statistiques, économie, psychologie, santé publique, etc.), l'OES 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). Il réalise environ une quinzaine d'expérimentations aléatoires contrôlées chaque année, principalement à l'aide de données administratives. Depuis 2015, l'OES a mené plus d'une soixantaine d'évaluations aléatoires contrôlées pour une douzaine d'agences fédérales. Parmi les projets réalisés, l'OES a récemment évalué, pour le compte du ministère de la santé (Centers for Medicare and Medicaid Services, HHS), les effets de l'envoi d'un courrier informant les médecins les plus prescripteurs de quétiapine – un antipsychotique – de leurs pratiques divergentes de celles de leurs homologues. L'évaluation, basée sur le tirage au sort d'environ 5 000 médecins prescrivant beaucoup de quétiapine, a montré que cette intervention réduisait la prescription de quétiapine dans les neuf mois suivants³⁸.

À sa création, l'OES a repris les projets en cours portés par un groupe intergouvernemental, **Social and Behavioral Sciences Team (SBST)**, lorsque ce dernier a cessé ses activités. Le SBST visait à inciter les agences gouvernementales à concevoir des expérimentations utilisant les sciences comportementales et à en évaluer les effets pour améliorer leurs programmes³⁹. Créé en 2014 et situé au sein de l'administration de la Maison-Blanche, le SBST était un sous-comité du National Science and Technology Council (NSTC)⁴⁰. Il comprenait une douzaine d'organisations gouvernementales dont les ministères de la défense, de l'agriculture, de la santé, de l'éducation, etc.

Enfin, il existe au niveau local des partenariats entre les universités et les villes, les États, etc. Par exemple, J-PAL North America a lancé en 2015 le **J-PAL State and local Initiative**⁴¹ dans le but d'aider les gouvernements locaux à utiliser les expérimentations aléatoires contrôlées pour évaluer leurs dispositifs et à en intégrer les apports dans leurs décisions. Chaque année, JPAL invite les gouvernements locaux à présenter des questions de politique publique qu'ils aimeraient voir évaluées par des méthodes aléatoires contrôlées. Les candidats sélectionnés reçoivent un soutien technique du personnel de JPAL, des financements pour développer leur projet et une mise en relation avec le réseau de chercheurs universitaires de JPAL. Ceux

³⁸ Voir la [description du projet](#) et l'[abstract](#).

³⁹ Le SBST doit aider les ministères et les agences fédérales à remplir les objectifs qui leur ont été assignés dans l'Executive Order « [Using Behavioral Science Insights to Better Serve the American People](#) » de septembre 2015 : (i) identifier les politiques, les programmes et les actions pour lesquelles les sciences comportementales peuvent conduire à des améliorations substantielles ; (ii) développer des stratégies pour mettre en œuvre ces idées et en évaluer rigoureusement les effets ; (iii) embaucher des experts en sciences comportementales au sein du gouvernement fédéral ; (iv) rechercher des opportunités pour renforcer les relations entre les agences fédérales et la communauté académique afin de mieux utiliser les résultats des sciences comportementales.

⁴⁰ Le National Science and Technology Council est situé au sein de l'Office of Science and Technology Policy ; <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/CoT-SBST-Charter-0315-SIGNED.pdf>.

⁴¹ Pour plus d'informations, voir <https://www.povertyactionlab.org/stateandlocal/aboutus>.

d'entre eux qui ont établi un partenariat avec un chercheur de JPAL pour concevoir une évaluation aléatoire contrôlée peuvent ensuite bénéficier de financements supplémentaires pour la conduite de l'évaluation. Douze partenariats avec des gouvernements locaux sont aujourd'hui en cours. Cette initiative est financée par la fondation Laura and John Arnold. De même, il existe des **Labs dans plusieurs universités** pour aider les gouvernements locaux à évaluer leurs dispositifs, dont les travaux sont principalement financés par des fondations⁴².

2. De nombreux dispositifs formels favorisant la conduite et l'utilisation des évaluations d'impact

2.1. Une structuration institutionnelle bien avancée et qui va se renforcer

Des responsables prévus dans chaque agence

L'**OMB** joue un rôle central pour impulser des évaluations 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 directives incitant les ministères à faire davantage d'évaluations de programmes rigoureuses et à conditionner l'octroi de subventions à l'atteinte des résultats. L'expérimentation rapide et d'autres stratégies fondées sur la preuve (mise en place de nouveaux instruments de financement de programmes) constituent également des axes de promotion de l'OMB (voir ci-dessous).

En particulier, en 2015, l'OMB a mis en place une instance de coordination sur l'évaluation, appelée **Interagency Council on Evaluation Policy (ICEP)**, pour aider les agences fédérales à adopter des pratiques d'évaluation communes, dans un esprit similaire à la fonction exercée par l'instance réunissant les principales agences statistiques, Federal Interagency Council on Statistical Policy (ISCP). L'objectif de cette instance est d'améliorer la programmation, la conduite et la diffusion des évaluations au sein du gouvernement fédéral. Ses missions comprennent la définition de principes communs, comme ceux publiés par les agences statistiques ; l'élaboration d'un cadre de référence méthodologique commun ; la conception de formations pour les agents en charge de l'évaluation (par exemple sur la gestion des contrats de recherche et d'évaluation). Elle comprend des représentants de plusieurs ministères (éducation, justice, logement, santé, travail) et d'agences indépendantes (Millenium Challenge Corporation, National Science Foundation, Corporation for National and Community Service).

L'ICEP a initié un dialogue entre ses membres sur les principes communs garantissant la crédibilité des évaluations réalisées par le gouvernement (comme l'intégrité, l'objectivité, l'impartialité, l'utilité, la rigueur et la pertinence des évaluations). Il a donné lieu à un séminaire et à des actes sur ce thème en octobre 2016⁴³. En novembre 2016, dans un memorandum adressé à la Commission on Evidence-Based Policymaking, l'ICEP a également identifié cinq problèmes majeurs auxquels sont confrontés les bureaux d'évaluation fédéraux (relatifs

⁴² Voir par exemple [California Policy Lab](#), UCLA et UC Berkeley ; [Urban Labs](#), University of Chicago ; [Lab for Economic Opportunities](#), University of Notre Dame.

⁴³ National Academy of Sciences, Engineering, and Medicine (2017).

notamment aux données et aux moyens humains et financiers) et a formulé des recommandations pour progresser⁴⁴.

Certains ministères disposent eux-mêmes de bureaux d'évaluation importants ayant un rôle de coordination et de pilotage des activités d'évaluation pour l'ensemble du ministère (voir ci-dessus). Ainsi, le ministère du travail dispose d'un chef évaluateur (**Chief Evaluation Officer, CEO**⁴⁵, Department of Labor), dont le bureau est localisé dans l'Office de l'Assistant Secretary for Policy. Institué en 2010, le Chief Evaluation Officer 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 sont des 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 / 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 également dans d'autres ministères sous d'autres appellations⁴⁶. Par exemple, le directeur de l'**Institute of Education Sciences (IES)** joue le rôle de chef évaluateur pour le ministère de l'éducation. L'IES est composé de plusieurs centres, dont le **National Center for Education Evaluation and Regional Assistance (NCEE)** pour les activités d'évaluation. Ce dernier est chargé de conduire et de promouvoir le développement des évaluations d'impact rigoureuses des programmes ministériels. Il en est de même de la directrice de l'**Office of Planning, Research and Evaluation (OPRE)** au sein du ministère de la santé (Department of Health and Human Services), dont le rôle est de conseiller l'Assistant Secretary for Children and Families sur les façons d'accroître l'efficacité des dispositifs fédéraux destinés à améliorer le bien-être des enfants et de leur famille. On peut également citer l'**Office of Policy Development and Research (PD&R)** du ministère du logement (Department of Housing and Urban Development).

La nouvelle loi, **Foundations for Evidence-Based Policymaking Act (FEBPA)** de 2018, prévoit la nomination dans chaque agence fédérale d'un chef évaluateur (Evaluation Officer) qui aura pour fonction de garantir en particulier la cohérence, la qualité et l'indépendance des évaluations de l'agence ; de déterminer les capacités de l'agence à 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 (voir ci-dessus).

⁴⁴ Interagency Council on Evaluation Policy (2016).

⁴⁵ www.dol.gov/asp/evaluation/.

⁴⁶ Schupmann et Fudge (2018), *Evidence-Based Policymaking Collaborative*.

Une programmation pluriannuelle prévue dans chaque agence

Certains ministères et agences fédérales ont mis en place des **Learning Agendas**^{47,48} qui recensent les questions politiques et de recherche les plus importantes à un moment donné, et les hiérarchisent en fonction des contraintes politiques et budgétaires. Ces outils permettent d'identifier les sujets sur lesquels les connaissances scientifiques manquent et de coordonner les travaux entre les différents services compétents. Leur élaboration donne lieu à une large consultation de parties prenantes, incluant les hauts responsables de l'administration, les directions en charge des programmes, les chercheurs externes et le Congrès. Il en existe plusieurs aujourd'hui à différents niveaux de l'administration. On peut en citer deux particulièrement développés :

- **Department of Labor, Chief Evaluation Office, Evaluation Plan**⁴⁹ : chacune des quinze agences du ministère du travail (Department of Labor) élabore, en collaboration avec le chef évaluateur du ministère (Chief Evaluation Officer), un « Learning Agenda » sur cinq ans, actualisé chaque année. Il inclut différents types d'études reposant sur des méthodologies variées, comme de l'évaluation d'impact (méthodes expérimentales et quasi expérimentales), de l'analyse statistique de base, de la recherche, des études sur la mise en œuvre des programmes, etc. À partir des programmes d'études des agences, le chef évaluateur du ministère prépare un programme annuel d'évaluations au niveau du ministère. Elles portent sur les évaluations financées par des fonds réservés (set-asides) qui octroient un pourcentage du budget du ministère au CEO. Ce programme annuel est publié dans le registre fédéral pour recueillir les avis du public et des parties prenantes. Il est également adressé au Congrès et à l'OMB ;
- **HUD Research Roadmap** : il s'agit du programme de recherche et d'évaluation sur cinq ans du ministère du logement et du développement urbain (Department of Housing and Urban Development), préparé par l'entité en charge de l'élaboration des politiques et de la recherche (Office of Policy Development and Research, PD&R). Élaboré pour la première fois en 2013 (pour la période allant de 2014 à 2018), il a été mis à jour en 2017. Son actualisation est prévue tous les cinq ans. Il s'appuie sur les contributions de la communauté scientifique, de praticiens et de décideurs publics aux différents niveaux (fédéral et local).
- **IES Evaluation Plan** : l'Institute for Education Sciences (IES), en collaboration avec d'autres acteurs du ministère de l'éducation via un groupe réunissant des hauts responsables du ministère, Evidence Learning Group, établit un plan prévisionnel des évaluations sur deux ans couvrant toutes les évaluations des programmes d'éducation, obligatoires ou discrétionnaires, financées par la loi Every Student Succeeds Act (ESSA) de 2015. Ce plan fait office de Learning Agenda pour le ministère de l'éducation.

Reprenant la recommandation de la CEP (Commission on Evidence-Based Policymaking), la nouvelle loi **FEBPA de 2018** demande aux agences fédérales de mettre en place des Learning Agendas pour définir les questions prioritaires de politique publique, recenser les besoins en données et décrire les études et évaluations à conduire.

⁴⁷ Nightingale, Fudge et Schupmann (2018), *Evidence-Based Policy making Collaborative*.

⁴⁸ Nightingale et Scott (2018).

⁴⁹ Chief Evaluation Officer, [Evaluation Plan](#), 2018.

Des standards de qualité hétérogènes

Certaines agences fédérales ont développé une politique spécifique pour leurs activités d'évaluation (**ACF Evaluation Policy, 2012**⁵⁰ ; **DOL Evaluation Policy, 2013**⁵¹ ; **PD&R Program Evaluation Policy, 2016**⁵² ; **IES Evaluation Principles and Practices, 2017**⁵³). Celle-ci définit plusieurs principes s'appliquant à ces activités et 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), publiques (principe de transparence) et protègent la confidentialité et la vie privée des individus (principe d'éthique). Le principe de rigueur impose notamment l'application des méthodes les plus rigoureuses, entendues au sens large (évaluations de processus, études descriptives, évaluations d'impact, études quantitatives et qualitatives), en réponse aux questions évaluatives posées. Pour l'évaluation des effets causaux d'un programme, le principe de rigueur requiert la mobilisation des méthodes expérimentales en premier lieu et des méthodes quasi expérimentales de « haute qualité » en second lieu (quand les premières ne sont pas possibles). **L'adoption d'une politique d'évaluation par chaque agence fédérale est une des exigences de la récente loi FEBPA de 2018. Elle prévoit aussi 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 (voir précédemment). Dans le domaine de l'aide internationale au développement, des lignes directrices sur le suivi et l'évaluation ont été produites en 2018 par l'OMB à destination des agences fédérales ayant une telle mission, en réponse aux exigences de la loi, Foreign Aid Transparency and Accountability Act de 2016⁵⁴.

Pour garantir la qualité de leurs évaluations, les agences fédérales s'appuient notamment sur des **standards de preuve**. Ceux-ci 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 (voir Annexe 4). Ils ont servi à l'élaboration de plusieurs Clearinghouses, dont le What Works Clearinghouse (WWC) du ministère de l'éducation, la Clearinghouse for Labor Evaluation and Research (CLEAR) du ministère du travail ou encore le Teen Pregnancy Prevention (TPP) Evidence Review du ministère de la santé (voir ci-dessous). Ces standards diffèrent généralement selon les agences.

Certaines agences ont également mis en place des **procédures de contrôle et de revue de qualité des évaluations**, en faisant appel à des experts indépendants. **La procédure de contrôle de qualité des évaluations la plus élaborée aujourd'hui est celle établie par l'IES** (ministère de l'éducation)⁵⁵. Elle consiste en la mise en place de **groupes de travail**

⁵⁰ [U.S. Department of HHS-ACF Evaluation Policy \(2012\)](#).

⁵¹ [U.S. Department of Labor Evaluation Policy \(2013\)](#).

⁵² U.S. Department of Housing and Urban Development, [HUD Program Evaluation Policy \(2016\)](#).

⁵³ U.S. Department of Education, Institute of Education Sciences, [Evaluation Principles and Practices \(2017\)](#).

⁵⁴ OMB (2018) : M-18-04.

⁵⁵ Pour plus de détails sur la procédure de revue par les pairs de l'IES, voir : https://ies.ed.gov/director/sro/peer_review/report_review.asp.

techniques, chargés d'accompagner le personnel de l'IES et les équipes de recherche sélectionnées pendant la conduite des évaluations et en **l'organisation d'une procédure de revue par les pairs** rigoureuse par l'IES (par le Standard and Review Office) conformément aux exigences de la loi Education Sciences Reform Act (ESRA)⁵⁶ de 2002, fondatrice de l'institut. Obligatoire avant la publication, cette procédure repose sur l'examen critique du rapport d'évaluation par au moins deux *referees* 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). Cette procédure a été approuvée en 2006 par le **National Board for Education Sciences**, composé de 15 membres nommés par la présidence des États-Unis sur les conseils et avec l'accord du Sénat (chercheurs éminents, administrateurs d'établissements scolaires et dirigeants d'entreprise).

Enfin, certaines agences fédérales ont élaboré des guides pour leurs activités d'évaluation, dont certains sont communs à plusieurs agences. Par exemple, l'Institute of Education Sciences (IES) (Department of Education) et la National Science Foundation (NSF) ont élaboré le **Guidelines for Education Research and Development** en 2013. Cet ouvrage précise comment différents types d'études contribuent à la définition de preuves. C'est sur la base de ce guide que l'OPRE a développé son propre guide, le **Common Framework for Research and Evaluation**, pour les activités de recherche et d'évaluation des dispositifs de l'Administration for Children and Families (Department of Health and Human Services).

En dehors de la sphère exécutive, la société américaine d'évaluation (American Evaluation Association, AEA) a également établi des principes s'appliquant aux évaluateurs dans la conduite de leurs travaux (**Guiding Principles for Evaluators, 2018**⁵⁷) et a formulé des propositions pour améliorer l'évaluation au sein du gouvernement fédéral (**An Evaluation Roadmap for A More Effective Government, 2013**⁵⁸). Elle a notamment recommandé ainsi que chaque agence fédérale définisse sa politique d'évaluation et ses normes de qualité, consulte les parties prenantes et établisse un programme d'évaluations pour alimenter la prise de décision. On peut également citer les standards établis par le **Joint Committee on Standards for Educational Evaluation**⁵⁹ (**Educational Program Evaluation Standards**⁶⁰ et **Personnel Evaluation Standards**).

Du côté du Congrès, le **GAO** a également élaboré un guide sur l'évaluation (**Designing Evaluations, 2012**⁶¹), et une note clarifiant la distinction entre l'évaluation et la mesure de la performance (**Performance Measurement and Evaluation, 2011**⁶²). Ces outils à destination des évaluateurs du GAO visent à présenter les différentes méthodes à leur disposition pour répondre aux questions posées (mise en œuvre, efficacité, etc.).

⁵⁶ <https://legcounsel.house.gov/Comps/Education%20Sciences%20Reform%20Act%20Of%202002.pdf>.

⁵⁷ American Evaluation Association (2018).

⁵⁸ American Evaluation Association (2013).

⁵⁹ Le Joint Committee on Standards for Educational Evaluation rassemble des associations professionnelles américaines et canadiennes (dont l'American Evaluation Association, AEA et la Canadian Evaluation Society, CES).

⁶⁰ Yarbrough, Shulha, Hopson et Caruthers (2011).

⁶¹ GAO (2012).

⁶² GAO (2011).

Les **fondations** peuvent également disposer de leur propre guide et lignes directrices pour garantir la qualité des évaluations qu'elles financent. Par exemple, l'Evidence-Based Policy Team de la fondation LJAF a élaboré en 2016 une checklist⁶³ dont l'objectif est de s'assurer que l'expérimentation aléatoire contrôlée est correctement conduite et permettra de produire des résultats robustes.

2.2. Un cadre juridique favorable à l'évaluation d'impact mais pas systématique

De nombreuses lois instaurant de nouveaux programmes contiennent des exigences plus ou moins fortes en matière d'évaluation. Elles peuvent également comprendre des dispositions sur le financement des évaluations, en réservant une part des crédits du programme à l'évaluation. Il n'y a pas de cadre formel obligatoire mais la pratique est répandue depuis les années 1960.

Un exemple intéressant est celui du programme **District of Colombia Opportunity Scholarship Program (OSP)** instauré par la loi Scholarships for Opportunity and Results Act (SOAR) de 2011⁶⁴ (voir Encadré 1). La loi confie à l'IES l'évaluation de ce dispositif, en lui demandant de remettre annuellement un rapport au Congrès. Elle contient par ailleurs des dispositions très prescriptives sur la méthodologie. En effet, elle précise que l'évaluation doit reposer sur la méthode la plus robuste pour mesurer les effets du programme⁶⁵ et mentionne les variables de résultat devant être évaluées⁶⁶. Enfin, elle indique que le financement dédié à l'évaluation ne doit pas dépasser 5 % des crédits accordés au programme.

2.3. Des sources de financement diverses

Le gouvernement fédéral est le plus gros financeur d'évaluations d'impact, en particulier les ministères de l'éducation (Department of Education), de l'emploi (Department of Labor) et de la santé (Department of Health and Human Services).

Les sources de financement des évaluations sont diverses. Les agences bénéficient de crédits directs dont elles décident elles-mêmes de la répartition entre leurs différentes activités et de la part consacrée aux évaluations en particulier (« direct appropriations »). Mais certaines agences bénéficient également de **crédits réservés** pour leurs activités d'évaluation spécifiquement. Il peut s'agir d'un pourcentage des crédits d'un ministère ou d'une agence qui est mis de côté pour financer les évaluations (« **evaluation set-asides** »). Par exemple, depuis 2012, le Chief Evaluation Officer (CEO) bénéficie de financements réservés à hauteur de 0,75 % au maximum des crédits affectés à certaines agences opérationnelles du Department of Labor (DOL)⁶⁷. Mais il peut également s'agir

⁶³ Voir « [Key items to get right when conducting randomized controlled trials of social programs](#) ».

⁶⁴ Scholarships for Opportunity and Results Act of 2011, sect. 9. Evaluations.

⁶⁵ Voir SOAR (2011), p. 21 : « *The evaluation is conducted using the strongest possible research design for determining the effectiveness of the opportunity scholarship program* ».

⁶⁶ Il s'agit des progrès scolaires des élèves, de la satisfaction des parents et des élèves, de la perception de la sécurité à l'école, et de l'implication des parents dans l'éducation de leur enfant. Voir SOAR (2011), page 22, Issues to be evaluated.

⁶⁷ Consolidated Appropriations Act (2017).

d'une partie des crédits d'un programme ou d'un dispositif ministériel, réservée à son évaluation. Par exemple, la loi Elementary and Secondary Educational (ESEA) Act de 1965, remplacée ensuite par la loi Every Student Succeeds (ESSA) Act de 2015, autorise 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. Ces fonds peuvent être mis en commun lorsque les financements alloués à certains dispositifs du programme sont trop faibles pour financer des évaluations d'impact de qualité. Pour 2018, les fonds réservés aux évaluations du programme représentent environ 19,5 millions de dollars, dont 11,5 millions pour de nouvelles évaluations et 8 millions pour des évaluations en cours. Sur ces 19,5 millions, environ 11,6 millions sont des fonds mutualisés à partir de plusieurs dispositifs et qui serviront à financer une ou plusieurs évaluations mentionnées dans le programme d'évaluations de l'IES (IES Evaluation Plan⁶⁸).

S'agissant des principaux bureaux d'évaluation, le CEO (DOL) dispose en 2018 d'un budget de 21,4 millions de dollars (dont 8 millions de crédits directs, la somme restante correspondant aux crédits transférés des autres agences du ministère)⁶⁹ et gère un portefeuille d'une quarantaine d'études dont une dizaine sont des évaluations d'impact. Le NCEE (IES, Department of Education) dispose pour sa part d'un budget d'environ 38 millions de dollars en 2018⁷⁰ et d'un portefeuille d'une quarantaine d'évaluations de mise en œuvre ou d'impact (hors évaluations financées dans le cadre du programme ESSA).

Selon une étude récente du GAO, ces fonds apparaissent globalement insuffisants au regard des besoins⁷¹. La Commission on Evidence-Based Policymaking (CEP) avait recommandé, 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.

Les fondations financent également des évaluations d'impact (Laura and John Arnold Foundation, MacArthur Foundation, The Pew Charitable Trusts, etc., voir plus haut). LJAF est celle qui en finance le plus : son équipe dédiée, l'Evidence-Based Policy Team, dispose ainsi d'un budget de 17 millions de dollars en 2018 et finance une vingtaine d'expérimentations aléatoires contrôlées par an⁷².

Au niveau local, les États, les villes, etc., financent également des évaluations, en particulier d'impact.

⁶⁸ [Evaluation Plan for Fiscal Years 2018 and 2019 Appropriations.](#)

⁶⁹ [Results for America \(2018\).](#)

⁷⁰ *Ibidem.*

⁷¹ GAO (2017).

⁷² <https://craftmediabucket.s3.amazonaws.com/uploads/Grants-for-Randomized-Controlled-Trials-Administered-by-the-LJAF-Evidence-Based-Policy-Team-2019.01.10.pdf>.

3. Des efforts sur la diffusion et la vulgarisation des évaluations, une réflexion et une pratique avancées sur leur utilisation dans la décision publique

3.1. Un principe de transparence dans plusieurs agences fédérales

Pour les agences fédérales disposant d'une politique d'évaluation, le principe de transparence est un principe important qui requiert des agences fédérales qu'elles publient les rapports d'évaluation sur leur site internet dans un délai approprié (dans les deux mois suivant la fin d'un rapport pour ACF et DOL et dans les quatre mois suivants pour PD&R). Ce principe précise aussi que les rapports publiés doivent présenter l'ensemble des résultats, y compris les résultats défavorables ou nuls. Les agences fédérales s'engagent également à conserver les données primaires ayant servi aux évaluations pour une utilisation secondaire par des chercheurs intéressés. Selon ce principe, les informations sur les évaluations en cours ou prévues doivent également être accessibles au public. Pour l'Institute of Education Sciences, le principe de transparence est inscrit dans sa loi fondatrice de 2002 (Education Sciences Reform Act⁷³, 2002) et s'impose donc à l'institut. La récente loi FEBPA de 2018 prévoit l'adoption d'une politique d'évaluation par chaque agence fédérale. Cette politique reposera sur plusieurs principes dont celui de transparence.

Pour les principaux bureaux d'évaluation, les informations sur les évaluations conduites sont accessibles, ainsi que les rapports. Quelques exemples sont mentionnés :

- le CEO (Department of Labor) publie sur son site internet dans deux rubriques distinctes les rapports en cours (**current studies**) et les rapports terminés (**completed reports**). Pour les premiers, des informations sont données sur l'objectif de l'étude, les questions évaluatives posées, la durée du projet, sa date de fin, le nom des contractants. Les liens vers les rapports sont accessibles dans la rubrique des rapports terminés ;
- le NCCE (IES, Department of Education) recense sur son site internet les nouveaux rapports d'évaluation disponibles dans une rubrique **What's New** (en précisant les liens) et plus généralement l'ensemble des évaluations dans une **base de données**. Pour chaque évaluation de la base, des informations synthétiques (**Evaluation Profiles**⁷⁴) sont présentées à la fois sur le contractant réalisant l'évaluation (les noms, la durée du contrat et son coût) et sur le projet lui-même (le contexte, les questions posées, la méthode et les principaux résultats) ;
- l'OPRE (ACF, Department of Health and Human Services) publie toutes les deux semaines une **newsletter** qui met à la une certains projets et mentionne les rapports publiés (avec des liens vers les documents). Plus généralement, tous les rapports de l'OPRE sont disponibles dans une **base de données, avec à chaque fois** des informations sur l'objectif de l'étude, la méthode utilisée et les **principaux résultats obtenus**. Il n'y a cependant pas de possibilité de sélectionner les seuls rapports d'évaluation.

⁷³ <https://legcounsel.house.gov/Comps/Education%20Sciences%20Reform%20Act%20Of%202002.pdf>

⁷⁴ Voir par exemple https://ies.ed.gov/ncee/projects/evaluation/choice_impactmagnet.asp.

Par ailleurs, il existe différentes bases de données qui répertorient les évaluations d'impact en cours ou réalisées. Certaines ne portent que les expérimentations aléatoires contrôlées, d'autres incluent les méthodes quasi expérimentales. Parmi les plus utilisées, on peut citer :

- [American Economic Association \(AEA\) RCT Registry](#) ;
- [Open Science Framework](#) ;
- [U.S. National Institutes of Health Clinical Trials Registry](#) ;
- [Society for Research on Education Evaluation \(SREE\) Registry](#) ;
- [International Initiative for Impact Evaluation \(3ie\) Registry](#) for International Development Impact Evaluation ;
- [JPAL Database](#).

Enfin, dans un double souci de pertinence et de transparence, les évaluations sont ouvertes aux parties prenantes. En particulier, ces dernières sont consultées largement au moment de l'élaboration des Learning Agendas. Les méthodes de consultation des parties prenantes sont variées. Par exemple, pour l'élaboration du HUD Research Roadmap (Department of Housing and Urban Development) de 2017, les consultations se sont déroulées en 2016 sous forme de réunions, de forums sur internet, de mails ciblés s'adressant à des experts dans les domaines étudiés. Elles ont permis de collecter environ 500 idées ou questions de recherche de parties prenantes, internes ou externes à l'organisation. De même, le programme annuel d'évaluations du Department of Labor est publié dans le registre fédéral pour recueillir les avis du public et des parties prenantes.

3.2. Des outils élaborés de diffusion des résultats des évaluations : les Clearinghouses

Les agences fédérales disposent de différents moyens pour rendre accessibles les résultats des évaluations aux non-spécialistes. Les rapports d'évaluation contiennent systématiquement un « **executive summary** » qui résume les principaux résultats. Plusieurs agences fédérales disposent également de collections spécifiques de vulgarisation des résultats des évaluations. Par exemple, on peut citer :

- le NCCE (IES, Department of Education) publie, outre les rapports d'évaluation, des synthèses courtes et accessibles sous différentes formes ([Evaluation Brief](#), [Study Snapshot](#) et [Study Highlights](#)) ;
- le PD&R (Department of Housing and Urban Development) publie une collection trimestrielle, appelée [Evidence Matters](#) ;
- l'OES (General Services Administration) publie des [abstracts](#) de deux pages sur les évaluations réalisées.

Une autre initiative intéressante est celle de la Laura and John Arnold Foundation (LJAF), [Straight Talk on Evidence](#), qui présente d'une manière accessible et critique les résultats d'évaluations d'impact disponibles (dont celles commandées par le gouvernement), en particulier ceux issus d'expérimentations aléatoires contrôlées.

Outre ces formats spécifiques, 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 employées et les résultats obtenus. Les études sont regroupées par grande thématique dans le domaine et par intervention ou dispositif à l'intérieur du domaine (voir l'exemple en Annexe 5). Les études recensées 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 :

- **What Works Clearinghouse, WWC** (U.S. Department of Education) ;
- **Research-Tested Intervention Programs, RTIP** (U.S. Department of Health and Human Services) ;
- **Teen Pregnancy Prevention Evidence Review** (U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation) ;
- **Home Visiting Evidence of Effectiveness, HomVEE** (U.S. Department of Health and Human Services, Administration for Children and Families) ;
- **CrimeSolutions.gov** (U.S. Department of Justice) ;
- **Clearinghouse for Labor Evaluation and Research, CLEAR** (U.S. Department of Labor).

D'autres Clearinghouses ont été développées par des universités, des fondations ou des administrations locales, par exemple :

- **Blueprints for Healthy Youth Development** (Institute of Behavior Science, University of Colorado Boulder, and the Annie E. Casey Foundation) ;
- **Social Programs That Work** (Laura and John Arnold Foundation, LJAF) ;
- **California Evidence-Based Clearinghouse for Child Welfare** (California Department of Social Services, CDSS).

Même si les Clearinghouses adoptent une démarche similaire, chacune a développé ses propres standards pour évaluer dans quelle mesure un dispositif est appuyé par des preuves solides d'efficacité. Le classement établi reflète à la fois la rigueur de la méthodologie employée et les résultats obtenus dans les études. Par exemple, le What Works Clearinghouse (Department of Education) considère qu'un dispositif est appuyé par des preuves fortes d'efficacité (niveau ++ ou positif) si on dispose pour lui à la fois⁷⁵ :

⁷⁵ Voir [WWC Standard Handbook \(version 4.0\)](#).

- d'au moins deux études respectant les critères de qualité méthodologique du WWC⁷⁶ et indiquant des effets positifs et statistiquement significatifs ;
- d'aucune étude rigoureuse montrant des résultats négatifs, statistiquement significatifs ou substantiellement importants.

Au total, le WWC distingue six niveaux de preuves (voir Annexe 3), qui suppose chacun le respect d'un certain nombre de conditions définies précisément dans les manuels des procédures du WWC⁷⁷.

Le Teen Pregnancy Prevention Evidence Review (Department of Health and Human Services) est un autre exemple d'outil relativement proche du précédent dans sa logique. Il ne distingue cependant que quatre niveaux de preuves d'efficacité (voir Annexe 6).

L'utilisation des données des Clearinghouses par les décideurs resterait cependant complexe et limitée. Il existe plusieurs raisons à cela : elles sont nombreuses ; elles se renouvellent rapidement (certaines se créent et d'autres disparaissent ou fusionnent) ; elles fonctionnent indépendamment les unes des autres, chacune adoptant son propre système de classement des dispositifs ; elles sont dotées de moyens limités pour accroître leur visibilité⁷⁸.

Pour améliorer la visibilité et la lisibilité des Clearinghouses auprès des décideurs, deux organisations, Pew Charitable Trusts et MacArthur Foundation, ont développé une nouvelle base de données, **Results First Clearinghouse Database**, qui rassemble l'ensemble des études collectées dans neuf Clearinghouses et offre un système de classement commun, distinguant cinq niveaux de preuves d'efficacité. La base de données RF Clearinghouse renvoie ensuite vers les sites des Clearinghouses pour des informations détaillées relatives aux dispositifs et à leurs résultats. La base comprend au total 2 867 dispositifs dans le domaine des politiques sociales (éducation, emploi et formation professionnelle, santé publique, justice criminelle (voir Annexe 3).

3.3. De nouveaux outils de financement des programmes reposant sur la preuve de leur efficacité : Tiered-Evidence Grants et Pay for Success

La récente loi FEBPA de 2018 marque une étape importante dans la conception de politiques publiques fondées sur la preuve. Elle demande aux agences fédérales notamment de nommer des chefs évaluateurs, de se doter d'une politique d'évaluation et de mettre en place une programmation des évaluations (Learning agenda). L'OMB est chargé d'élaborer, dans l'année suivant la promulgation de la loi, des lignes directrices pour aider les agences fédérales à conduire leurs activités d'évaluation. Cette loi a repris pour partie les recommandations de la Commission on Evidence-Based Policymaking mise en place sous l'administration Obama.

⁷⁶ L'une des deux études doit respecter les critères de qualité méthodologique du WWC sans réserve. Il s'agit de méthodes expérimentales (RCT) avec un taux d'attrition faible.

⁷⁷ Voir [WWC Procedures Handbook \(version 4.0\)](#).

⁷⁸ www.pewtrusts.org/-/media/assets/2015/06/results_first_clearinghouse_database_user_guide.pdf.

Sous l'administration Obama, de nouveaux outils de financement de dispositifs ont été mis en place reposant sur les résultats des évaluations disponibles. Il s'agit des Tiered-Evidence Grants (TEG) et des Pay for success (PFS).

Tiered-Evidence 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, du logement, etc. 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. Il s'agit des « **tiered-evidence grants** » aussi appelés « **innovation fund grants** ».

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 (l'étape de mise à l'échelle des projets, ou « scale up tier ») ;
- le second type accorde des montants plus faibles 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) (l'étape de validation des projets, « 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 (l'étape de la preuve du concept).

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 comme les plus performantes (voir les deux exemples ci-dessous). 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.

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 dollars⁷⁹. Cela représente cependant une part faible des subventions accordées aux États et aux collectivités locales⁸⁰.

Par exemple, le programme du ministère de la santé, Teen Pregnancy Prevention (TPP), établi en 2010, vise à diminuer les grossesses juvéniles. Il a une structure de subventionnement à deux étages seulement : un premier niveau (« Tier 1 ») pour des dispositifs appuyés par des preuves solides de leur efficacité et un second niveau (« Tier 2 ») pour des dispositifs prometteurs mais peu évalués. Le ministère doit allouer 75 % des crédits du programme à des subventions de niveau 1 et 25 % à des subventions de niveau 2 (il s'agit d'une obligation légale)⁸¹. Les dispositifs éligibles aux subventions de niveau 1 sont identifiés par le ministère de la santé en amont de la procédure (voir Encadré 2 et Annexe 6).

**Encadré 2 – Comment sont déterminés les dispositifs efficaces ? Et à quel moment ?
L'exemple du programme Teen Pregnancy Prevention Program (ministère de la santé)**

Avant le lancement de la procédure d'appel à projets, le ministère de la santé (HHS), en collaboration avec Mathematica Policy Research, identifie une liste de dispositifs considérés comme efficaces et relevant donc des subventions de niveau 1. Pour faire partie de la liste, les dispositifs doivent remplir trois critères :

- un critère d'éligibilité (sur les dispositifs, les bénéficiaires, les variables d'intérêt, les méthodes utilisées : expérimentales ou quasi expérimentales, la date de l'étude) ;
- un critère de qualité méthodologique garantissant la robustesse des résultats, établi à partir des standards disponibles (dérivés du What Works Clearinghouse du ministère de l'éducation) ;
- et un critère de résultat (au moins une étude doit conclure à des résultats positifs).

Les candidats aux subventions de niveau 1 doivent choisir un des dispositifs de la liste en fonction des besoins de la population ciblée. En avril 2018, 48 dispositifs font partie de la liste. Cette dernière est mise à jour régulièrement.

Le programme du ministère de l'éducation, Investing in Innovation (i3), est un autre exemple de subventions à plusieurs étages. Destiné à améliorer la réussite scolaire des étudiants, ce programme établi en 2010 offre trois types de subventions selon la force de la preuve d'efficacité à l'appui des projets : « scale-up » pour des subventions jusqu'à 20 millions de dollars pour les projets disposant de preuves solides d'efficacité ; « validation » pour des subventions jusqu'à 12 millions pour des projets ayant des preuves modérées d'efficacité ; « development » pour des subventions jusqu'à 3 millions pour des projets montrant des signes

⁷⁹ Feldman et Haskins, Evidence-Based Policymaking Collaborative (2016). Il s'agit du Social Innovation Fund (Corporation for National and Community Service) ; Investing in Innovation (i3) remplacé en 2015 par Education Innovation and Research (Department of Education) ; Supporting Effective Educator Development (Department of Education) ; Home Visiting (Department of Health and Human Services) ; Teen Pregnancy Prevention (Department of Health and Human Services) ; Development Innovation Ventures (U.S. Agency for International Development).

⁸⁰ Le gouvernement fédéral dépense environ 600 milliards de dollars chaque année en subventions pour les États et les collectivités locales (voir le rapport du GAO déjà cité, GAO-16-818).

⁸¹ Consolidated Appropriations Act (2016).

d'efficacité ou soutenus par une théorie d'action forte⁸². L'éligibilité des projets aux différentes subventions est vérifiée au moment de la sélection des offres (voir Encadré 3 et Annexe 6).

**Encadré 3 – Comment sont déterminés les dispositifs efficaces ? Et à quel moment ?
L'exemple du programme Investing in Innovation (i3) (ministère de l'éducation).**

Les niveaux de preuve sont définis précisément dans les documents officiels (le registre fédéral). Les niveaux de preuve élevés et modérés sont réservés aux projets soutenus par des évaluations utilisant des méthodes expérimentales ou quasi expérimentales. Schématiquement, pour un avoir un niveau élevé de preuves d'efficacité, un projet doit être appuyé par :

- au moins une étude utilisant une méthode expérimentale (avec un niveau d'attrition faible) sur un échantillon large d'étudiants couvrant plusieurs sites (établissements, villes, États) ;
- ou par au moins deux études mobilisant une méthode expérimentale (avec un niveau d'attrition élevé) ou une méthode quasi expérimentale à partir d'un échantillon large d'étudiants couvrant plusieurs sites.

Pendant la sélection des offres, l'organisme en charge de la recherche et de l'évaluation au ministère de l'éducation, Institute of Educational Sciences (IES), examine les études citées dans les propositions des candidats pour déterminer si elles vérifient les conditions exigées. Le non-respect de ces conditions entraîne *de facto* l'inéligibilité du projet.

Pay for Success

Pay for Success est un mécanisme de financement de programmes. 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, et un évaluateur indépendant qui détermine si le programme atteint ses objectifs. Il est originaire du Royaume-Uni où il a été mis en place pour la première fois en 2010 (Social Impact Bonds, SIB) et s'est développé depuis dans plusieurs pays.

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 (c'est-à-dire s'il atteint ses objectifs), 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 est trop faible au regard du coût, un tiers intervient parfois pour se porter garant et supporter une partie du coût en cas d'échec du programme.

Ce mécanisme 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 (voir Encadré 4 pour l'exemple du programme ABLE lancé par la ville de New York). Par ailleurs, il

⁸² La théorie d'action fait référence à un cadre conceptuel qui identifie les différentes composantes du projet ainsi que les relations possibles entre ces composantes et les variables d'intérêt.

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 programme^{83,84}.

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 récidive⁸⁵. Ils émanent de gouvernements locaux exclusivement (États, villes, comtés) (voir Tableau 1). 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 de dollars pour chaque⁸⁶).

Tableau 1 – Projets PFS sur la période 2012-2016

Location	Année de lancement	Domaine d'intervention	Investissement total (en millions de dollars)
Ville de New York (NY)	2012	Récidive	9,6
Comté de Salt Lake (Utah)	2013	Éducation de la petite enfance	7,0
New York (villes de New York et Rochester)	2013	Récidive	13,5
Massachusetts (Boston, Chelsea et Springfield)	2014	Justice pour mineurs	21,7
Chicago (Illinois)	2014	Éducation de la petite enfance	16,7
Comté de Cuyahoga (Ohio)	2014	Protection de l'enfance	4
Massachusetts	2014	Sans-abrisme	3,5
Comté de Santa Clara (Californie)	2015	Sans-abrisme	6,9
Denver (Colorado)	2016	Sans-abrisme	8,7
Caroline du Sud	2016	Santé de l'enfant et de la mère	17,5
Connecticut	2016	Petite enfance	12,5

Source : Milner, *Evidence-Based Policymaking Collaborative* (2016)

Le premier projet de ce type aux États-Unis a été lancé en septembre 2012 par la ville de New York, afin de réduire la récidive des jeunes détenus de la prison de Rikers Island. Le programme, appelé Adolescent Behavioral Learning Experience (ABLE), s'appuie sur une forme de thérapie comportementale et cognitive (TCC) connue sous le nom de Moral Reconciliation Therapy (MRT). Cette dernière avait été choisie essentiellement pour deux

⁸³ GAO (2015).

⁸⁴ Non Profit Finance Fund (2016).

⁸⁵ 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](#).

⁸⁶ 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 de dollars 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.

raisons : d'une part parce qu'elle avait fait la preuve de son efficacité, toutefois dans d'autres contextes que la prison ; d'autre part parce qu'elle était assez souple pour être adaptée au contexte spécifique de la prison (voir Encadré 4 pour plus de détails)⁸⁷.

Ce programme était financé par Goldman Sachs à hauteur de 9,6 millions de dollars. La ville de New York n'était tenue d'effectuer aucun paiement si le taux de récidive était abaissé de moins de 8,5 % grâce au programme. En revanche, à partir de ce taux, elle devait rembourser un montant compris entre 4,8 et 11,7 millions de dollars selon la baisse du taux de récidive. Dans ce schéma, l'investisseur pouvait récupérer sa mise initiale pour une réduction du taux de récidive au moins égale à 10 %, et pouvait espérer un rendement maximal de 2,1 millions de dollars pour une réduction du taux de récidive de 20 % au minimum. Au vu des évaluations disponibles avant la mise en place du programme, il était vraisemblable d'attendre une réduction du taux de récidive d'environ 11 % du fait de ce programme, soit un rendement de près de 0,5 million de dollars pour l'investisseur (voir Encadré 4). Toutefois, les résultats d'évaluation obtenus à l'aide d'une méthode quasi expérimentale⁸⁸ ont montré que le projet n'avait pas eu d'effet sur le taux de récidive au terme de la première année. Sur la base de ces résultats, Goldman Sachs a décidé de mettre fin au projet, et la ville de New York n'a eu à procéder à aucun paiement (voir Annexe 5).

**Encadré 4 – Comment sont utilisées les évaluations dans les dispositifs PFS ?
L'exemple du programme ABLE (Adolescent Behavioral Learning Experience)
de la ville de New York**

***Le programme ABLE était appuyé par diverses études attestant de l'efficacité
des thérapies cognitives et comportementales***

L'impact estimé du programme ABLE était fondé sur une méta-analyse de 58 évaluations de différents modèles de thérapie comportementale et cognitive (TCC) s'adressant à des délinquants criminels dans différents contextes. Cette méta-analyse publiée par Landenberger et Lipsey (2005) montrait que tous les modèles de TCC permettaient de réduire le taux de récidive dès lors qu'ils étaient mis en œuvre comme prévu initialement (de l'ordre de 25 % en moyenne), et que la Moral Reconciliation Therapy (MRT) conduisait à des résultats aussi favorables que les autres thérapies.

La thérapie MRT fut finalement retenue pour le programme ABLE car elle était la plus étudiée de toutes les interventions et parce qu'elle pouvait s'ajuster aux conditions uniques du milieu carcéral (caractérisé par un taux élevé de turnover notamment).

***Les résultats des évaluations permettaient de montrer qu'il était raisonnable d'attendre
une réduction de 11 % du taux de récidive***

Pour estimer l'effet probable du programme ABLE, MDRC a tenu compte du fait que certains bénéficiaires ne suivraient pas entièrement le programme car ils seraient sortis de prison avant la fin du programme. Les données disponibles permettaient de distinguer quatre groupes de jeunes éligibles selon le temps passé dans le programme. Le premier groupe (33 %) comprend les jeunes restant en prison suffisamment longtemps pour suivre l'intégralité du programme (soit au moins 10 semaines). Le second groupe (9,3 % des jeunes) comprend ceux qui suivent en grande partie le programme (entre 6 et 10 semaines), alors que le troisième groupe (18,8 % des

⁸⁷ Rudd, Nicoletti, Misner et Bonsu (2013).

⁸⁸ Vera Institute of Justice (2015).

jeunes) est composé de ceux qui le suivent en petite partie seulement (entre 3 et 6 semaines). Enfin, le quatrième groupe (38,9 % des jeunes) comprend ceux qui restent en prison trop peu de temps pour que le programme ait un quelconque effet sur eux. MDRC a fait ensuite logiquement l'hypothèse que l'effet du programme décroissait entre ces groupes : il est de 25 % pour le premier groupe, 16,6 % pour le second, 8,3 % pour le troisième et 0 % pour le quatrième groupe. Sous ces hypothèses, l'effet moyen estimé du programme ABLE est de 11,4 %. Selon les termes du contrat, cette réduction permettait à l'investisseur de récupérer sa mise initiale avec un intérêt de 0,48 million de dollars.

Références : Rudd, Nicoletti, Misner et Bonsu (2013) ; Landenberger et Lipsey (2005).

Pour terminer, l'Encadré 5 indique quelques évaluations d'impact qui ont eu un écho important dans la presse nationale américaine ou qui ont pu influencer dans une certaine mesure la décision publique. Ces exemples ont été mentionnés par les personnes interrogées lors des entretiens. Ils visent à illustrer cette question, mais ne permettent pas de tirer des enseignements plus généraux sur l'influence des évaluations d'impact sur la décision.

Encadré 5 – Contributions au débat et influence de quelques évaluations d'impact emblématiques

Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV)

Le programme MIECHV vise à donner aux femmes enceintes et aux familles les plus à risques les moyens et les compétences pour élever dans de bonnes conditions leur enfant de la naissance à l'entrée en Kindergarten (grande section de maternelle). Il s'appuie sur des visites de travailleurs sociaux et autres professionnels de la petite enfance au domicile des familles. Le ministère de la santé (Department of Health and Human Services), l'Administration for Children and Families (ACF) et la Health Resources and Services Administration (HRSA), octroient des financements aux États, aux territoires et aux communautés tribales pour mettre en place des dispositifs de visite à domicile appuyés par des preuves solides d'efficacité (sur la base de critères définis par le ministère de la santé) et répondant aux besoins de leur population. Les candidats peuvent choisir l'un des dix-huit dispositifs sélectionnés par le ministère de la santé (Department of Health and Human Services) à partir d'une revue de la littérature régulièrement mise à jour de l'efficacité des dispositifs existants ([Home Visiting Evidence of Effectiveness](#), HomeVEE).

Promulgué par la loi Patient Protection and Affordable Act de 2010, le programme MIECHV a d'abord été autorisé et financé pour cinq ans de 2010 à 2014 pour 1,5 milliard de dollars (avec une montée en puissance des crédits sur la période, de 100 millions de dollars en 2010 à 400 millions de dollars à partir de 2013). Selon la loi, 75 % des crédits doivent être alloués à des dispositifs considérés comme efficaces selon les critères établis par le ministère de la santé et identifiés dans HomeVEE, tandis que les 25 % restants peuvent être octroyés à des dispositifs prometteurs (ces derniers doivent être rigoureusement évalués). La loi précise également que 3 % des crédits sont réservés aux activités d'assistance technique, de recherche et d'évaluation (notamment pour l'évaluation nationale du programme, appelée MIHOPE).

Appuyé par de nombreuses évaluations montrant des résultats positifs, le programme a été prolongé par plusieurs lois successives, d'abord en 2015 (Protecting Access to Medicare Act de 2014 pour 400 millions de dollars), puis pour deux ans en 2016 et en 2017 (Medicare Access and Children's Health Insurance Program Reauthorization Act de 2015 pour 400 millions par année) et enfin pour cinq ans de 2018 à 2022 (Bipartisan Budget Act de 2018 pour 400 millions par

année) (<https://fas.org/sgp/crs/misc/IF10595.pdf> ; <https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting-overview>).

Moving to Opportunity

Lancée par le ministère du logement (Department of Housing and Urban Development, HUD) en 1994 dans cinq grandes villes (Baltimore, Boston, Chicago, Los Angeles, New York), l'expérimentation Moving to Opportunity (MTO) porte sur des familles à bas revenus vivant dans des quartiers très pauvres. L'objectif est de les aider à se reloger dans de meilleurs quartiers (via une aide financière et une aide à la recherche de logement) pour améliorer leur bien-être, leur santé, les performances scolaires des enfants ainsi que l'emploi et les revenus des parents et des enfants à l'âge adulte. Au total, 4 600 familles sont entrées dans le dispositif expérimental entre septembre 1994 et août 1998. Par tirage au sort, elles ont été affectées au groupe de contrôle et aux deux groupes de traitement (dont l'un est contraint à court terme sur le lieu de relocalisation). Les résultats ont été mesurés à différents horizons (à court terme, à moyen terme entre quatre et sept ans après le tirage au sort et enfin à long terme entre 10 et 15 ans après le tirage au sort). Ils sont issus d'un ensemble d'évaluations conduites par des chercheurs universitaires (www.nber.org/mtopublic/).

Les résultats ont montré que les familles déménageant dans des quartiers moins pauvres sont plus heureuses, en meilleure santé et se sentent davantage en sécurité et que les enfants âgés de moins de 13 ans lors du déménagement sont ensuite davantage susceptibles d'aller à l'université, d'avoir des revenus plus élevés, de vivre à leur tour dans de meilleurs quartiers et ont moins de risques d'appartenir à une famille monoparentale. En revanche, pour les enfants âgés de 13 ans ou plus lors du déménagement, les résultats sont légèrement négatifs à long terme. Enfin, le dispositif n'a aucun effet sur l'emploi et les revenus des personnes qui étaient adultes au moment du déménagement. Ces résultats tendent à montrer qu'un meilleur ciblage du dispositif sur les familles avec des jeunes enfants serait plus efficace (J-PAL Policy Briefcase, 2015).

Ce dispositif et les différentes évaluations réalisées ont été régulièrement commentés dans la presse nationale américaine sur une période de 20 ans environ (depuis 1994 où l'expérimentation a été mise en place et jusqu'en 2015 où les dernières évaluations de long terme ont été publiées) (www.nber.org/mtopublic/news.htm). Aujourd'hui, le sujet est toujours d'actualité car la ségrégation résidentielle reposant sur le revenu et la richesse est en augmentation aux États-Unis. Dans la continuité de cette expérimentation, un nouveau projet, Creating Moves to Opportunity (CMTO), a été lancé par différents acteurs, chercheurs (dont J-PAL et MDRC), autorités administratives, associations et fondations, pour identifier les voies et les moyens d'inciter les bénéficiaires du Housing Choice Voucher Program, principal dispositif d'aide au logement dans le secteur privé pour les familles à bas revenu, à s'installer dans de meilleurs quartiers (qualité de l'information dispensée lors de la recherche de logements, accompagnement global à la mobilité, incitations pour les propriétaires et les locataires, aides financières couvrant les dépenses liées au déménagement, etc.). (www.povertyactionlab.org/na/cmto).

Oregon Health Insurance Experiment

En 2008, l'État de l'Oregon a étendu le bénéfice de son dispositif de santé Medicaid, appelé Oregon Health Plan Standard, à des adultes à bas revenus âgés de 19 à 64 ans ne bénéficiant pas d'une assurance santé publique. Ce dispositif n'incluait plus de nouveaux bénéficiaires depuis 2004 pour des raisons budgétaires. Anticipant un excès de demande par rapport au nombre de places disponibles, l'État de l'Oregon a réalisé un tirage au sort des personnes pouvant bénéficier du dispositif. Les effets du dispositif ont été évalués à un horizon de deux ans en comparant les résultats des personnes tirées au sort (29 834 individus éligibles) à ceux des personnes qui ne l'ont pas été (45 088 individus) à partir à la fois de données administratives et de données d'enquêtes. Cette expérimentation a permis d'apporter des réponses à des questions non tranchées par la théorie économique et très polémiques, par exemple sur les effets de la

couverture santé sur l'accès aux soins ou sur la santé ou encore sur le recours aux services d'urgence des hôpitaux. Elle a donné lieu à de nombreuses évaluations conduites par des chercheurs universitaires (www.nber.org/oregon).

Les résultats ont montré que l'extension de la couverture santé à de nouveaux bénéficiaires dans l'État de l'Oregon leur a permis d'augmenter le recours aux soins de santé (nombre d'hospitalisations, recours aux services des urgences, consultations en ville), de réduire les contraintes financières liées à des dépenses de santé importantes, de diminuer la prévalence de dépression, mais n'a eu aucun effet sur la santé physique (cholestérol, pression artérielle ou risque cardio-vasculaire) ou sur l'emploi et les salaires (J-PAL Policy Briefcase, 2014).

Ces résultats ont été largement repris et commentés dans la presse nationale aux États-Unis (www.nber.org/oregon/7.media.html) en particulier ceux sur l'augmentation du recours aux services des urgences des hôpitaux. Selon un argument souvent avancé en faveur du dispositif, Medicaid devait en effet permettre de désengorger les services des urgences des hôpitaux en réorientant les patients vers les soins primaires.

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'impact ?
 - *Les exigences en termes d'évaluation sont-elles proportionnées à l'importance de la politique publique (par exemple, en termes budgétaires) ? Ces dispositions 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.

Annexe 2

Dates clés de l'évaluation

1921	Création du General Accounting Office (GAO) , dépendant exclusivement du Congrès (institution homologue à la Cour des Comptes).
1946	Création du Congressional Research Service (CRS) , le centre de recherche de la Library of Congress.
1972	Création du Office Technology Assessment (OTA) qui sera supprimé en 1995
Début des années 1960	Mise en place du Planning Programming Budgeting System (PPBS) , équivalent de la rationalisation des choix budgétaires en France, dans toutes les agences fédérales. L'évaluation est utilisée comme outil de contrôle et d'optimisation des dépenses publiques. Le PPBS est abandonné dans les années 1970.
1970	Le Legislative Reorganization Act de 1970 officialise la mission d'évaluation des politiques et programmes fédéraux du GAO . Cette mission lui avait été confiée précédemment par des amendements en 1967 à l'Economic Opportunity Act qui lui demandèrent d'évaluer l'efficacité et l'efficience des programmes de lutte contre la pauvreté mis en œuvre par l'administration Johnson.
1974	Création du Congressional Budget Office (CBO) , organisation non partisane, qui fournit des informations aux membres du Congrès sur le coût des mesures budgétaires et de leurs alternatives possibles.
1981	Les agences fédérales doivent évaluer les coûts et les bénéfices liés aux mesures réglementaires relevant de leurs attributions, sous le contrôle technique de l'OMB, selon une décision présidentielle (executive order 12 291).
1993	The Government Performance and Results Act (GPRA) demande aux décideurs publics plus d'informations objectives sur l'efficacité et l'efficience des programmes fédéraux et des dépenses.
2004	Mise en place par l'OMB d'un outil de revue des programmes fédéraux (Program Assessment Rating Tool, PART). Cet outil les classe en cinq catégories : efficace, modérément efficace, acceptable, résultats non démontrés et inefficace.
2004	Le General Accounting Office devient le Government Accountability Office . Il supprime sa division spécialisée dans l'évaluation des programmes. Sa mission d'évaluation est recentrée sur le contrôle et la surveillance des activités d'évaluation des agences fédérales.
2010	Le GPRA Modernization Act (GPRAMA) demande que les vingt-quatre agences fédérales et l'OMB établissent des objectifs prioritaires, les suivent trimestriellement, et rendent publics sur un site gouvernemental les progrès et stratégies adoptées pour renforcer la performance.
2011	Lancement de la Voluntary Fiscal Year Evaluation Initiative sous l'administration Obama, destinée à « inciter les agences, sur la base du volontariat, à soumettre des informations additionnelles sur la performance des programmes, et à prévoir et demander des budgets dédiés à la conduite d'évaluations prioritaires » (Deloitte, 2013).

Annexe 3

Principaux acteurs de l'évaluation

Type	Exemples	Date de création	Statut	Missions et types d'évaluation	Secteurs
Organes du Parlement	GAO	1921	Agence indépendante relevant du pouvoir législatif	Contrôle, audit, évaluation, analyse de politiques	Multisecteurs
	CBO	1974	Agence fédérale relevant du pouvoir législatif	Prévisions économiques et budgétaires ; estimations de coût de propositions de loi	Santé, sécurité nationale, travail, fiscalité, énergie et macroéconomie
	CRS	1946	Agence fédérale relevant du pouvoir législatif	Analyse des politiques publiques	Multisecteurs
Organes de l'exécutif	Services d'études des ministères :		Intégrés à chaque ministère	Recherche et évaluation	Thématique du ministère
	<i>Institute of Education Sciences (IES)</i>	2002	<i>Department of Education</i>	<i>Quatre centres, dont le National Center for Educational Evaluation and Regional Assistance (NCEE)</i>	<i>Éducation</i>
	<i>Office of Planning, Research and Evaluation (OPRE)</i>		<i>Administration for Children and Families (ACF) Department of Health and Human Services (HHS)</i>		<i>Famille, enfants</i>
	<i>Office of Policy Development and Research (PD&R)</i>		<i>Department of Housing and Urban Development (HUD)</i>		<i>Logement et développement urbain</i>
	<i>National Institute of Justice (NIJ)</i>		<i>Department of Justice (DOJ)</i>		<i>Justice</i>
	<i>Chief Evaluation Office</i>	2010	<i>Department of Labor (DOL)</i>		<i>Travail</i>
Organismes publics	Washington State Institute for Public Policy (WSIPP)	1983	État de Washington	Analyses coûts-bénéfices, méta-analyses, revues de la littérature académique, synthèses rigoureuses des recherches conduites par WSIPP	Éducation, justice criminelle, questions sociales, travail, législation
Sociétés savantes	American Evaluation Association (AEA)	1986	Association professionnelle internationale	Améliorer les pratiques et les méthodes d'évaluation ; accroître l'utilisation des évaluations ; promouvoir l'évaluation en tant que profession ; soutenir la contribution de l'évaluation à la création de connaissances	Multisecteurs (40 groupes thématiques)
	Federal Evaluators	1999	Association informelle de responsables de l'évaluation au niveau	Partager des informations sur les méthodes, les pratiques et les politiques	Multisecteurs

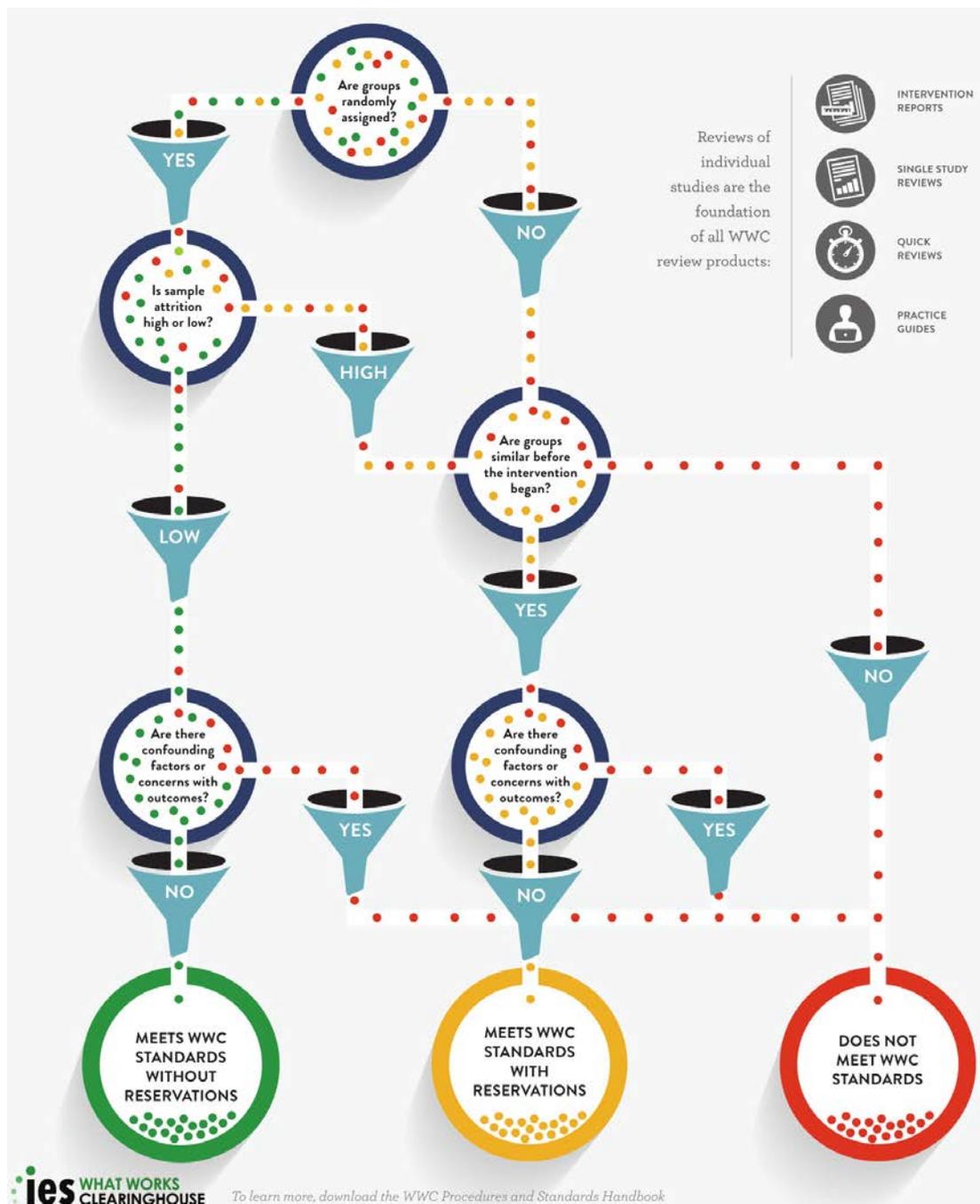
			fédéral (branches exécutive et législative)		
Organismes de recherche indépendants	NBER	1920	Organisation à but non lucratif et indépendante Financement par des subventions de recherche et des contrats, des dons de particuliers, de fondations et de sociétés et des revenus de placement	Recherche économique Publication des NBER working papers Organisation des NBER Summer Institute meetings, des rencontres entre des chercheurs Un réseau de plus de 1 400 chercheurs affiliés	Domaines très variés, regroupés en vingt programmes et quinze groupes de travail
	MDRC	1974	Organisation à but non lucratif et non partisane En 2015-2016 : financement à 70 % par les gouvernements au niveau fédéral, des États et des villes et à 18 % par des fondations, des sociétés et des individus.	Design d'interventions, assistance technique pour leur mise en œuvre et évaluation Plus de 250 experts et 80 projets en cours	Éducation et politiques sociales
	Mathematica Policy Research	1968	Organisation détenue entièrement par ses employés	Recherche, collecte et analyse de données, évaluation. Quatre centres associés, dont le Center for Improving Research Evidence (CIRE) Plus de 1200 experts	Politiques sociales, environnement, aides au développement
	Vera Institute for Justice	1961	Organisation à but non lucratif et indépendante En 2016 : financement à 86 % par les gouvernements aux niveaux fédéral, des États et des villes et à 9 % par des fondations	Recherche et analyses ; pilotage de projets. Environ 60 projets en cours dans 47 États.	Justice
	W.E. Upjohn Institute	1932	Organisation à but non lucratif, non partisane et indépendante	Analyses et évaluation ; études économiques ; administration et gestion de dispositifs publics Environ 100 employés	Travail, emploi, éducation, formation, développement économique et questions internationales
	The Pew Charitable Trusts	1948	Organisation à but non lucratif	Recherche et analyses	Multisecteurs : environnement ; économie et finance ; santé ; éducation ; arts et culture ; etc.
	Think tanks	RAND Corporation	1948	Organisation à but non lucratif En 2017 : financement à 82 % par les gouvernements au niveau fédéral, des États et des villes	Recherche et analyses Publication du <i>RAND Journal of Economics</i> , une revue à comité de lecture Environ 1 850 employés et 1 900 projets dont 650 nouveaux en 2017
Brookings Institution		1916	Organisation à but non lucratif et non partisane	Recherche, analyses et recommandations de	Domaines variés, regroupés en cinq

			Financement par des subventions de recherche et des contrats, des dons de particuliers, de fondations et de sociétés et des revenus de placement	politiques Environ 300 experts	programmes de recherche : Economic Studies, Global Economy and Development, Foreign Policy, Governance Studies, Metropolitan Policy Program
	Urban Institute	1968	Organisation à but non lucratif En 2016 : financement à 43 % par les gouvernements aux niveaux fédéral, des États et des villes et à 41 % par des fondations	Recherche et analyses ; data science ; recommandations de politiques Environ 450 experts et 200 projets en cours	Politiques économiques et sociales
	American Enterprise Institute for Public Policy Research (AEI)	1943	Organisation à but non lucratif et non partisane Financement par des dons de particuliers, de sociétés et de fondations et par des revenus de placement. L'organisation ne passe pas de contrat de recherche, et n'accepte pas de subvention du gouvernement.	Recherche sur l'économie mondiale, la politique étrangère des États-Unis et la sécurité internationale ainsi que sur les questions nationales (économiques et sociales)	Généraliste
Universités	Harvard University	1636	Université privée	Formation et recherche	Tous secteurs
	University of Chicago	1890	Université privée	Formation et recherche	Tous secteurs
	University of Michigan	1817	Université publique	Formation et recherche	Tous secteurs
	Massachusetts Institute of Technology	1861	Université privée	Formation et recherche	Tous secteurs
	Pennsylvania University	1740	Université privée	Formation et recherche	Tous secteurs
	University of California, Berkeley	1868	Université publique	Formation et recherche	Tous secteurs
	Columbia University	1754	Université privée	Formation et recherche	Tous secteurs
	Cornell University	1865	Université privée	Formation et recherche	Tous secteurs
	University of Maryland	1856	Université publique	Formation et recherche	Tous secteurs
	Northwestern University	1851	Université privée	Formation et recherche	Tous secteurs
	Yale University	1701	Université privée	Formation et recherche	Tous secteurs
	University of Wisconsin-Madison	1848	Université publique	Formation et recherche	Tous secteurs

	University of California, Los Angeles (UCLA)	1919	Université publique	Formation et recherche	Tous secteurs
	New York University	1831	Université privée	Formation et recherche	Tous secteurs
Associations, fondations, organismes de bienfaisance	MacArthur Foundation	1975	Fondation philanthropique privée	Elle soutient financièrement des organisations aux États-Unis (dans tous les États) et à l'étranger (via l'octroi de subventions) ou des individus (MacArthur Fellowship)	Multisecteurs
	Laura and John Arnold Foundation	2008	Fondation philanthropique privée	Financement de projets et d'initiatives ; de recherches et d'évaluations	Multisecteurs : justice criminelle, éducation ; santé ; finances publiques ; recherche ; politiques fondées sur la preuve
	Robin Hood Foundation	1988	Fondation philanthropique privée	Financements de projets et d'initiatives	Lutte contre la pauvreté à New York
	Robert Wood Johnson Foundation	1972	Fondation philanthropique privée	Financement de projets et d'initiatives ; de recherches et d'évaluations	Santé
	Boomborg Philanthropies	2006	Fondation philanthropique privée	Financement de projets et d'initiatives	Multisecteurs : santé publique ; environnement ; éducation ; innovation gouvernementale ; arts.
	Ford Foundation	1936	Fondation philanthropique privée	Financement de projets et d'initiatives	Inégalités
	Rockefeller Foundation	1913	Fondation philanthropique privée	Financements de projets et d'initiatives	Multisecteurs
Alliances et réseaux	Evidence-Based Policymaking Collaborative			Regroupement de chercheurs de plusieurs organisations (Urban Institute, Brookings Institution, American Enterprise Institute, Pew-MacArthur Results First Initiative) pour développer des ressources favorisant l'élaboration de politiques fondées sur la preuve au niveau fédéral	Politiques fondées sur la preuve
Alliances et réseaux	Abdul Latif Jameel Poverty Action Lab (J-Pal)	2003	Centre de recherche	Réseau de chercheurs en économie dans le monde entier spécialisés dans la conduite d'évaluations d'impact randomisées	Politiques de lutte contre la pauvreté

Annexe 4 Les critères de qualité méthodologique du What Works Clearinghouse

Figure A4-1 – Les standards du WWC



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 tests et de comparaison et les facteurs confondants) pour faire partie de la base.

Source : WWC, How the WWC rates a study ?

Annexe 5

Les Clearinghouses

WWC et RF Clearinghouse Database

Il existe plusieurs Clearinghouses, ayant chacune adopté son propre système de classement des dispositifs. Ainsi, le What Works Clearinghouse (WWC, Department of Education) distingue six niveaux de preuves (voir Tableau A5-1), qui supposent chacun le respect d'un certain nombre de conditions définies précisément dans les manuels des procédures du WWC⁸⁹.

Tableau A5-1 – Les six niveaux de preuves d'efficacité définis dans le WWC

Niveau de preuves		
Positif	++	Preuves fortes d'un effet positif
potentiellement positif	+	Preuves d'un effet positif
avec des effets mixtes	+-	Preuves d'effets de sens contraires
sans effet perceptible	0	Absence de preuve d'un quelconque effet, positif ou négatif
potentiellement négatif	-	Preuves d'un effet négatif
Négatif	--	Preuves fortes d'un effet négatif

Source : [What Works Clearinghouse Procedures Handbooks \(version 4.0\)](#), p. 30

La Figure A5-1 montre à partir d'un exemple de dispositif le type d'informations produit par l'outil WWC. Il s'agit ici d'un programme national financé par le gouvernement fédéral, appelé Head Start, qui s'adresse à des enfants de la naissance à l'âge de 5 ans provenant principalement de familles à faible revenu et qui vise à mieux les préparer à l'école. Il propose différents services aux enfants et à leurs familles, en matière notamment d'éducation, de santé et de nutrition. C'est l'un des 84 dispositifs recensés dans la thématique de la petite enfance du WWC (Early Childhood). L'outil WWC indique que le dispositif Head Start a des effets potentiellement positifs sur la réussite en lecture mais aucun effet perceptible sur la réussite en mathématiques et le développement socio-affectif chez les enfants de 3 et 4 ans. Ce résultat repose sur une seule étude respectant les critères de qualité méthodologique du WWC sur un total de 40 études passées en revue (selon une dernière mise à jour datant de juillet 2015).

Les résultats sont présentés sous la forme d'un indice qui indique la progression qu'aurait connue un enfant « moyen » du groupe de comparaison s'il avait bénéficié du programme : un indice de 13 signifie qu'il serait passé du 50^e au 63^e percentile de la distribution des scores en lecture (ou pour le dire autrement sur un groupe de 100 élèves, il aurait gagné 13 places au classement en passant de la 50^e à la 37^e place).

⁸⁹ Voir [WWC Procedures Handbook \(version 4.0\)](#).

Figure A5-1 – L'exemple du programme Head Start à partir de la base WWC

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	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.

Source : *What Works Clearinghouse. Topics: Early Childhood. Résultats disponibles à l'adresse suivante <https://ies.ed.gov/ncee/wwc/EvidenceSnapshot/636> ng*

La base de données, **Results First Clearinghouse Database**, développée par Pew Charitable Trusts et MacArthur Foundation, rassemble l'ensemble des études collectées dans neuf Clearinghouses (dont la WWC)⁹⁰ et offre un système de classement commun, distinguant cinq niveaux de preuves d'efficacité (voir Figure A5-2). La base de données établit des correspondances pour passer des échelles de notation propres aux neuf Clearinghouses à

⁹⁰ La base de données RF Clearinghouse Database comprend neuf Clearinghouses : Blueprints for Healthy Youth Development (Blueprints) ; California Evidence-Based Clearinghouse for Child Welfare (CEBC) ; The U.S. Department of Justice's CrimeSolutions.gov ; The U.S. Substance Abuse and Mental Health Services Administration's National Registry of Evidence-based Programs and Practices (NREPP)* ; The U.S. Department of Health and Human Services' Research-Tested Intervention Programs (RTIP) ; The Laura and John Arnold Foundation's Social Programs That Work ; The U.S. Department of Health and Human Services' Teen Pregnancy Prevention Evidence Review (TPP Evidence Review) ; The University of Wisconsin Population Health Institute and Robert Wood Johnson Foundation's County Health Rankings and Roadmaps What Works for Health ; The U.S. Department of Education's What Works Clearinghouse (WWC).

*NREPP a cessé ses opérations en janvier 2018.

l'échelle de notation commune^{91,92}. Par exemple, pour la base WWC, la correspondance est quasi immédiate et ne nécessite que le regroupement de deux niveaux en un seul (potentiellement négatif et négatif).

Figure A5-2 – Les cinq niveaux de preuves d'efficacité de RF Clearinghouse Database



Source : www.pewtrusts.org/en/research-and-analysis/data-visualizations/2015/results-first-clearinghouse-database

La figure A5-3 illustre l'information restituée pour le programme Head Start. La base de données RF Clearinghouse renvoie vers les sites des Clearinghouses pour des informations détaillées relatives aux dispositifs et à leurs résultats (en cliquant sur *Learn more*).

Figure A5-3 – L'exemple du Programme Head Start à partir de la base RF Clearinghouse

Head Start - Early Childhood Education



Clearinghouse: ●

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.

[Read less](#)

Clearinghouse:

WWC

Settings:

School

Clearinghouse rating:

Potentially positive effects

Ages:

Grades PK

Outcomes:

Mathematics achievement

Reading achievement

Social-emotional development

Target populations:

Not specified

Note : selon la base RF Clearinghouse, le programme Head Start est appuyé par des preuves élevées (mais pas les plus fortes) d'efficacité. Comme la base d'origine WWC note les performances du programme Head Start sur chacune des trois variables d'intérêt (réussite en mathématiques, réussite en lecture et développement socio-affectif), la base RF retient la note la plus élevée des trois pour effectuer la conversion dans l'échelle unique de notation.

Source : www.pewtrusts.org/en/research-and-analysis/data-visualizations/2015/results-first-clearinghouse-database

⁹¹ Results First Clearinghouse Database, Technical Appendix.

⁹² Selon une analyse de la base RF réalisée en 2015, près des deux tiers des dispositifs ne sont présents que dans une seule Clearinghouse. Pour le tiers restant (présent dans plusieurs Clearinghouses), la base RF Clearinghouse attribue plusieurs notes, qui dans 90 % des cas sont cohérentes entre elles. Les incohérences (10 % des cas) peuvent être dues à des différences dans les méthodes employées (par exemple, dans les études recensées) ou à des différences dans les variables d'intérêt étudiées (par exemple, l'achèvement de la scolarité pour l'un et les comportements sexuels risqués pour l'autre). Pour plus de détails, voir le guide [Results First Clearinghouse Database](#).

Annexe 6 Tiered evidence grants

Examen de deux programmes

1. Teen Pregnancy Prevention Program (TPP), ministère de la santé

Le ministère de la santé (Department of Health and Human Services, HHS) a établi en 2010 un programme, Teen Pregnancy Prevention (TPP), dont l'objectif est de diminuer les grossesses juvéniles. Ce programme propose deux types de subventions : une subvention de niveau 1 pour des projets dont l'efficacité a été démontrée au moyen d'évaluations rigoureuses menées antérieurement et une subvention de niveau 2 pour des projets innovants qu'il s'agit de tester et d'évaluer. Pour ce programme, le ministère doit allouer 75 % des crédits affectés aux programmes au financement de projets relevant du premier niveau et 25 % au financement de projets relevant du deuxième niveau (selon une obligation législative)⁹³.

Le programme Maternal Infant and Early Childhood Home Visiting, établi par le ministère de la santé la même année, offre les mêmes caractéristiques que le programme TPP : il présente une même structure de subventions à deux niveaux, et est assorti d'une même obligation d'allocation des crédits entre les deux types de subventions. Ce programme vise à améliorer la santé, le développement et le bien-être des jeunes enfants à risques et de leurs familles à partir de visites à domicile de professionnels et para-professionnels.

Programmes	Structure des programmes	Caractéristiques principales
<p>Teen Pregnancy Prevention (TPP)</p> <p>Établi en 2010 par le ministère de la santé (HHS), ce programme cherche à prévenir les grossesses juvéniles.</p>	<p>1^{er} niveau (« Tier 1 ») : il s'agit de répliquer, sur d'autres territoires ou sur une population élargie, des dispositifs dont l'efficacité est avérée</p>	<p>Candidats éligibles : organisations à but non lucratif, États et collectivités locales, universités, etc.</p> <p>Durée de financement : 5 ans</p> <p>Fonds versés sur 2010-2015 : 619,5 millions de dollars</p>
<p>Maternal Infant and Early Childhood Home Visiting</p> <p>Établi en 2010 par le ministère de la santé (HHS), ce programme vise à améliorer la santé, le développement et le bien-être des jeunes enfants à risques et de leurs familles.</p>	<p>2^e niveau (« Tier 2 ») : il s'agit de tester et d'évaluer des dispositifs innovants</p>	<p>Candidats éligibles : États, communautés tribales éligibles, organisations à but non lucratif</p> <p>Durée de financement : 2-3 ans (États) et 5-6 ans (communautés tribales)</p> <p>Fonds versés sur 2010-2015 : 1,7 milliard de dollars</p>

Source : GAO (2016)

⁹³ Consolidated Appropriations Act (2016).

Les dispositifs efficaces sont identifiés préalablement à la diffusion des avis de subventionnement par une revue des études disponibles

Pour ces deux programmes, le ministère de la santé a choisi d'identifier, préalablement à la diffusion des avis de subventionnement, les projets/dispositifs pouvant être considérés comme efficaces et relevant donc du premier niveau de subventionnement. Ce travail d'identification a été réalisé par Mathematica Policy Research, qui a passé en revue les études disponibles pour ce type d'interventions. Il est actualisé régulièrement pour prendre en compte les résultats des nouvelles études.

Dans le cas du programme de prévention des grossesses juvéniles, la dernière mise à jour de la revue (avril 2018) a permis d'identifier 48 dispositifs dont l'efficacité est attestée par des évaluations rigoureuses⁹⁴. Les résultats sont présentés de manière synthétique et pédagogique dans un tableau final où figurent en ligne les dispositifs et en colonne les cinq principales variables d'intérêt sur lesquelles ils peuvent jouer. Dans chaque case du tableau figure une bulle plus ou moins grosse selon le nombre d'études à l'appui. Elle peut prendre par ailleurs quatre couleurs différentes selon le sens des effets : elle est verte lorsque les effets sur la variable d'intérêt sont uniformément positifs ; jaune quand ils sont mixtes (positifs, nuls ou négatifs) ; elle est grise quand ils sont uniformément nuls et rouge quand ils sont uniformément négatifs (voir tableau A6-1 ci-dessous). Le tableau est mis à jour à chaque actualisation de la revue de la littérature.

Tableau A6-1 – Les quatre niveaux de preuves définis dans TPP Evidence Review

POSITIVE IMPACTS		Preuves d'effets uniformément positifs, quels que soient les indicateurs de résultats utilisés, les échantillons mobilisés (complets ou par sous-groupes) et/ou les études considérées
MIXED IMPACTS		Preuves d'effets variables, positifs, nuls et /ou négatifs selon les indicateurs de résultats utilisés, les échantillons mobilisés (complets ou par sous-groupes) et/ou les études considérées
NULL IMPACTS		Preuves d'effets uniformément nuls, quels que soient les indicateurs de résultats utilisés, les échantillons mobilisés (complets ou par sous-groupes) et/ou les études considérées
NEGATIVE IMPACTS		Preuves d'effets uniformément négatifs, quels que soient les indicateurs de résultats utilisés, les échantillons mobilisés (complets ou par sous-groupes) et/ou les études considérées

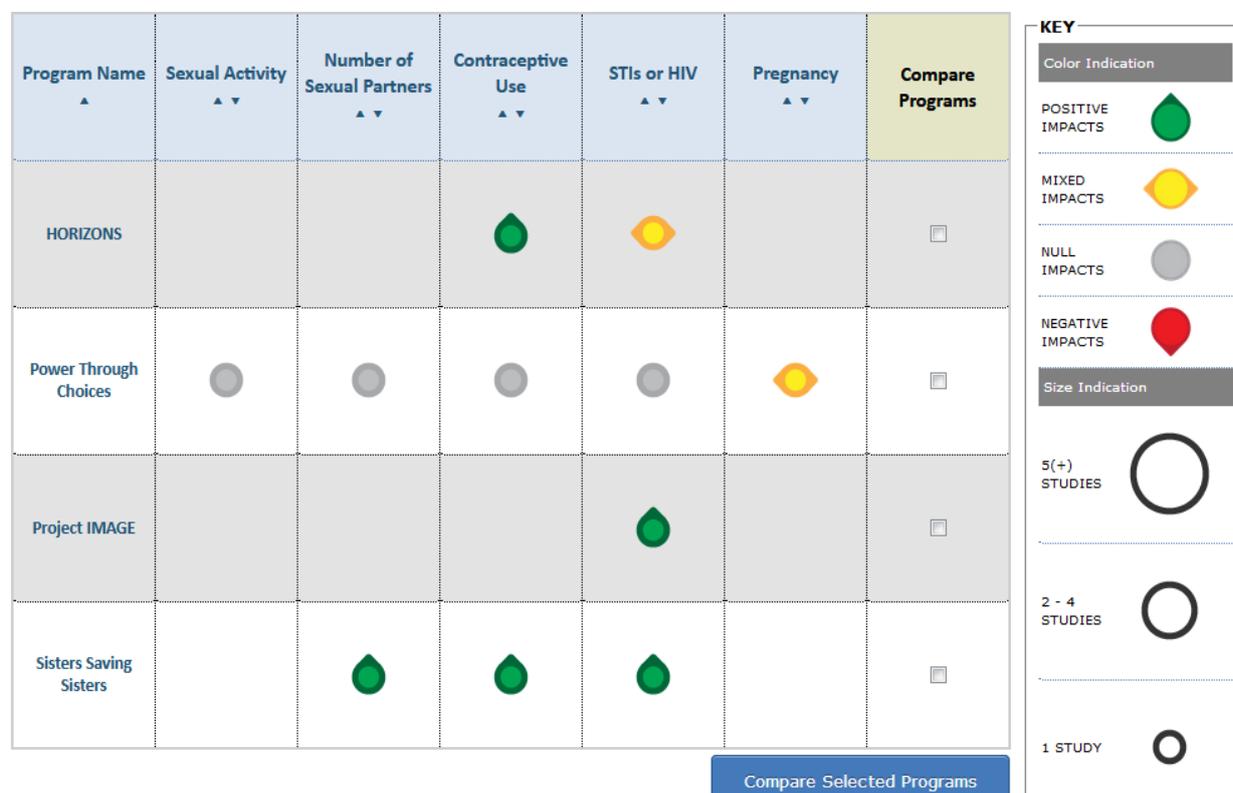
Source : Mathematica Policy Research (2016)

La figure A6-1 suivante est une extraction de la base de données TPP Evidence Review. Cette dernière permet de sélectionner les dispositifs selon plusieurs critères : le type de dispositif, la population ciblée et le lieu de mise en œuvre. Pour cette extraction, nous choisissons de retenir uniquement les dispositifs portant sur des jeunes afro-américains âgés de 18 à 19 ans. La base de données a permis d'identifier quatre dispositifs répondant à ces critères (indiqués en ligne). Elle indique que les variables d'intérêt les plus étudiées sont l'utilisation de contraceptifs et les maladies sexuellement transmissibles. Aucun de ces dispositifs ne conduit à des résultats uniformément négatifs sur l'une ou l'autre des

⁹⁴ Mathematica Policy Research (2018).

variables d'intérêt. Trois projets, Horizons, Project Image et Sisters Saving Sisters, trouvent des effets uniformément positifs sur au moins une des cinq variables d'intérêt. Chaque projet ne s'appuie cependant que sur les résultats d'une seule étude.

Figure A6-1 – L'exemple de quatre dispositifs figurant dans TPP Evidence Review



Note : le niveau de preuve d'efficacité associé aux dispositifs est évalué séparément pour chacune des cinq variables d'intérêt : l'activité sexuelle ; le nombre de partenaires sexuels ; l'utilisation de contraceptifs ; les maladies sexuellement transmissibles ; les grossesses. La taille des bulles indique le nombre d'études à l'appui : 5 études et plus pour la plus grosse bulle (non représentée ici), 2 à 4 études pour la bulle de taille intermédiaire (non représentée ici) et 1 étude pour la plus petite bulle.

Source : <https://tppevidencereview.aspe.hhs.gov/FindAProgram.aspx>. L'extraction concerne les dispositifs ciblant les jeunes afro-américains âgés de 18 à 19 ans.

Lorsqu'ils candidatent pour une subvention de niveau 1 (voir tableau), les candidats doivent sélectionner un dispositif dans la liste de ceux identifiés comme efficaces, selon les besoins de la population auxquels ils s'intéressent. Pour faire ce choix, le ministère de la santé met à disposition sur son site internet de nombreuses informations sur les dispositifs eux-mêmes (qu'il est possible de comparer, voir dernière colonne du tableau) et sur leurs résultats (à un niveau détaillé).

Les dispositifs efficaces doivent être appuyés par des études qui remplissent plusieurs critères

Dans le cas du programme de prévention des grossesses juvéniles, ces dispositifs sont appuyés par des études qui remplissent trois critères⁹⁵. Le premier critère concerne **l'éligibilité**. Les études doivent porter sur des programmes visant des jeunes américains âgés au plus de 19 ans, et dont l'objectif est de réduire le taux de grossesse juvénile, les maladies sexuellement transmissibles, ou les pratiques sexuelles à risques. Elles doivent également utiliser des méthodes expérimentales (RCT) ou quasi expérimentales. Elles doivent par ailleurs mesurer les effets du programme sur au moins une des cinq variables d'intérêt identifiées comme prioritaires : l'activité sexuelle ; le nombre de partenaires sexuels ; l'utilisation de contraceptifs ; les maladies sexuellement transmissibles ; les grossesses. Enfin, pour faire partie des programmes éligibles, on doit pouvoir compter au moins une étude conduite au cours des vingt dernières années.

Le second critère porte sur la **qualité méthodologique** des études. Les études éligibles doivent avoir un niveau de qualité méthodologique élevé ou modéré, celles ayant un niveau faible étant écartées. Ce niveau de qualité reflète la robustesse des résultats et donc le niveau de confiance qu'il est possible de leur accorder. Plus il est élevé, moins les résultats risquent d'être biaisés. Cette appréciation du niveau de qualité est faite par Mathematica à partir d'une version modifiée de l'outil développé pour élaborer le What Works Clearinghouse du ministère de l'éducation. Elle est établie à partir de l'examen de cinq critères techniques susceptibles d'influencer la robustesse des résultats : le design de l'étude (méthodes expérimentales ou quasi expérimentales) ; l'attrition (le fait que des individus sortent de l'échantillon au cours du temps) ; la similarité des groupes traités et de contrôle avant l'intervention ; la réaffectation d'individus de l'échantillon d'un groupe à l'autre ; la présence de facteurs confondants rendant impossible l'identification des effets du dispositif (et pouvant être liés au design ou au mode de collecte des données)⁹⁶ (voir le tableau A6-2 ci-dessous pour une présentation synthétique des conditions requises par niveau de qualité).

Le troisième et dernier critère porte sur **les preuves d'efficacité**. Il est examiné uniquement pour les études ayant un niveau de qualité élevé ou modéré. Un programme peut être considéré comme efficace s'il existe au moins une étude concluant à des résultats positifs statistiquement significatifs pour au moins une des cinq variables d'intérêt, qu'ils soient obtenus sur l'échantillon complet ou sur des sous-groupes (définis selon leur âge ou leur expérience sexuelle initiale). Cette étude ne doit pas montrer d'effet négatif statistiquement significatif (sur aucune des cinq variables d'intérêt).

⁹⁵ Mathematica Policy Research (2016).

⁹⁶ Cette condition permet de garantir que l'effet obtenu n'est pas dû à la catégorie particulière d'individus considérée. Par exemple, dans un design où seuls les élèves d'une école bénéficieraient du dispositif, il ne serait pas possible de savoir si l'effet est attribuable au dispositif ou aux caractéristiques de cette école.

Tableau A6-2 – Synthèse des conditions requises par niveau de qualité

Critères techniques	Qualité élevée	Qualité modérée	Qualité faible
1 - Design	RCT (expérimentations aléatoires contrôlées)	- Méthodes quasi expérimentales (avec groupe de contrôle) - RCT avec un niveau élevé d'attrition ou dont certains individus ont changé de groupe suite au tirage au sort	RCT ou méthodes quasi expérimentales ne respectant pas les critères d'un niveau de qualité élevé ou modéré
2 - Attrition	Respect des standards du What Works Clearinghouse	NA	Ne respecte pas les critères d'un niveau de qualité élevé ou modéré
3 - Équivalence initiale	Les différences initiales – statistiquement significatives – entre les deux groupes doivent être contrôlées	L'étude doit montrer que les deux groupes sont similaires avant l'intervention	Ne respecte pas les critères d'un niveau de qualité élevé ou modéré
4 - Réaffectation	Pas de réaffectation possible des individus entre les groupes après le tirage au sort	NA	Ne respecte pas les critères d'un niveau de qualité élevé ou modéré
5 - Facteurs confondants	Au moins deux catégories d'individus dans les groupes et pas de différence systématique dans le mode de collecte des données	Au moins deux catégories d'individus dans les groupes et pas de différence systématique dans le mode de collecte des données	Ne respecte pas les critères d'un niveau de qualité élevé ou modéré

Source : *Mathematica Policy Research (2016)*

Les projets financés doivent ensuite être évalués

Les projets subventionnés doivent être tous évalués, quel que soit le niveau de subventionnement (de niveau 1 ou de niveau 2). L'évaluation peut être financée par la subvention reçue dans la limite d'un plafond. Par exemple, pour le programme de prévention des grossesses juvéniles (TPP), le budget consacré à l'évaluation ne doit pas dépasser 10 % du montant total de la subvention reçue pour les projets bénéficiant d'une subvention de niveau 1 (tier 1).

En 2016, 41 évaluations des projets financés dans le cadre du programme TPP ont été publiées sur le site du ministère de la santé (HHS) : 19 évaluations correspondent à des projets subventionnés de niveau 1 (tier 1) et 22 à des projets subventionnés de niveau 2. Par ailleurs, parmi ces évaluations, 21 évaluations ont été publiées dans un supplément de la revue scientifique *The American Journal of Public Health* (septembre 2016)⁹⁷. Ces évaluations ont permis de dresser un premier bilan à la fois sur la qualité des évaluations réalisées et sur les premiers résultats du programme. Il en ressort que sur les 41 évaluations réalisées, 13

⁹⁷ Farb et Margolis (2016).

évaluations ont été jugées invalides (7 évaluations de projets de niveau 1 et 6 évaluations de projets de niveau 2, soit environ 30 % des évaluations réalisées), parce qu'elles avaient par exemple un niveau élevé d'attrition ou parce que les bénéficiaires du programme étaient trop jeunes pour avoir connu une expérience sexuelle avant la fin du programme.

Par ailleurs, les résultats des évaluations montrent des résultats modestement encourageants : environ un tiers des projets de niveau 1 (quatre évaluations sur douze évaluations valides) trouvent au moins un résultat positif sur l'une des variables d'intérêt considérées contre 50 % des projets de niveau 2 (huit évaluations sur 16 évaluations valides). Ces résultats posent question sur les critères choisis par le ministère de la santé en collaboration avec Mathematica Policy Research pour déterminer les interventions efficaces. En effet, il est surprenant de constater que les projets ayant un niveau 2 de subventionnement obtiennent plus souvent un effet positif que les projets ayant un niveau 1 de subventionnement, alors qu'ils sont nouveaux et qu'ils n'ont pas encore été rigoureusement évalués. En particulier, certains dispositifs considérés comme efficaces par le ministère de la santé (dans la liste des 48 interventions identifiées, voir ci-dessus) sont anciens (ils ont pu être mis en place une décennie ou deux avant) et n'ont pas été réévalués récemment. D'autres dispositifs reposent sur les résultats d'une seule évaluation montrant des effets positifs sur l'une des variables d'intérêt, ce qui peut paraître insuffisant. Selon des réflexions en cours, la preuve d'efficacité pourrait être renforcée en exigeant qu'au moins deux études conduisent à des résultats favorables (la seconde répliquant la première en suivant les mêmes procédures)⁹⁸.

2. Investing in Innovation (i3), ministère de l'éducation

Le ministère de l'éducation a mis en place en 2010 un programme qui vise à améliorer la réussite scolaire à l'université et à accroître le nombre d'étudiants, appelé Investing in Innovation (i3). Il propose trois types de subventions. Promulgué en 2015 dans le cadre du Every Students Succeeds Act, le programme Education Innovation and Research (EIR) remplace le programme i3.

Programme	Structure du programme	Caractéristiques principales
<p>Investing in Innovation (i3)</p> <p>Établi en 2010 par le ministère de l'éducation, il vise à améliorer la réussite scolaire et à accroître le nombre des élèves.</p> <p>Ce programme a été remplacé en 2015 par un autre programme présentant une architecture similaire à trois étages (Education Innovation and Research, EIR).</p>	<p>1^{er} niveau (« scale-up ») jusqu'à 20 millions de dollars : il s'agit de dispositifs dont l'efficacité est avérée</p> <p>2^e niveau (« validation ») jusqu'à 12 millions de dollars : il s'agit de dispositifs dont l'efficacité est moins établie que le 1^{er} niveau</p> <p>3^e niveau (« development ») jusqu'à 3 millions de dollars : il s'agit de dispositifs prometteurs mais dont l'efficacité est incertaine</p>	<p>Candidats éligibles : établissements scolaires et organisations à but non lucratif.</p> <p>Durée de financement : 3-5 ans</p> <p>Fonds versés sur 2010-2015 : 1,3 milliard de dollars</p>

Source : GAO (2016)

⁹⁸ Haskins et Joo (2017), Evidence-Based Policymaking Collaborative.

Les niveaux de preuve sont définis précisément dans la procédure officielle de soumission

Les niveaux de preuve sont précisément définis par le ministère de l'éducation dans les documents officiels (registre fédéral^{99,100}), ses procédures ou guides. Le niveau élevé de preuve d'efficacité exigé pour les subventions de niveau 1 (« scale-up ») signifie que l'une des deux options présentées dans le tableau A6-3 doit être vérifiée.

Tableau A6-3 – Critères d'un niveau élevé de preuves d'efficacité pour les subventions « scale-up »

Critères	Option 1	Option 2
Nombre d'études	Au moins 1 étude	Au moins 2 études
Résultats sur une variable d'intérêt pertinente	Impact positif statistiquement significatif (sans effet négatif)	
Standards des What Works Clearinghouse (WWC)	Les critères du WWC sont remplis <i>sans</i> réserve (RCT)	Les critères du WWC sont remplis <i>avec</i> réserve (RCT ou méthodes quasi expérimentales)
Taille de l'échantillon	Échantillon large	
Nombre de sites	Échantillon multi-sites	
Similarité des populations	Les populations – de l'étude et ciblée par le projet – doivent se recouper	

Note : les standards établis par le What Works Clearinghouse sont ceux figurant dans les manuels [des procédures et des standards](#). Un échantillon large signifie qu'il comprend au moins 350 étudiants ou au moins 50 groupes de classes ou d'écoles contenant chacun au moins 10 étudiants. Un échantillon multi-sites signifie qu'il contient plus d'un site, le site pouvant désigner un établissement scolaire, une municipalité ou un État.

Source : <https://innovation.ed.gov/files/2016/06/i3fundwebinar.pdf>

De la même façon, le niveau modéré de preuves d'efficacité exigé pour les subventions de niveau 2 (« validation ») requiert que l'une des deux options du tableau A6-4 soit vérifiée.

Tableau A6-4 – Critères d'un niveau modéré de preuves d'efficacité pour les subventions « validation »

Critères	Option 1	Option 2
Nombre d'études	Au moins 1 étude	Au moins 1 étude
Résultats sur une variable d'intérêt pertinente	Impact positif statistiquement significatif (sans effet négatif)	
Standards des What Works Clearinghouse (WWC)	Les critères du WWC sont remplis <u>sans</u> réserve (RCT)	Les critères du WWC sont remplis <u>avec</u> réserve (RCT ou méthodes quasi expérimentales)
Taille de l'échantillon	NA	Échantillon large
Nombre de sites	NA	Échantillon multi-sites
Similarité des populations	Les populations – de l'étude et ciblée par le projet – doivent se recouper	

Source : <https://innovation.ed.gov/files/2016/06/i3fundwebinar.pdf>

⁹⁹ *Federal Register*, vol. 81, n° 94, Monday, May 16, Notices.

¹⁰⁰ *Federal Register*, vol. 81, n° 79, Monday, April 25, Notices.

Par rapport au tableau précédent, l'option 1 ne comporte plus de condition sur la taille de l'échantillon ou sur le nombre de sites, tandis que l'option 2 ne requiert qu'une étude au minimum contre deux précédemment.

Enfin, le niveau le plus faible de preuves d'efficacité demandé pour les subventions de niveau 3 (« development ») requiert que le projet soit appuyé par une théorie d'action forte¹⁰¹ ou par des études indiquant des signes d'efficacité (*Evidence of Promise*). Dans ce dernier cas, l'une des trois options du tableau A6-5 doit être vérifiée.

Tableau A6-5 – Critères d'un niveau faible de preuves d'efficacité (Evidence of Promise) pour les subventions « development »

Critères	Option 1	Option 2	Option 3
Nombre d'études	Au moins 1 étude	Au moins 1 étude	Au moins 1 étude
Résultats sur une variable d'intérêt pertinente	Impact ou corrélation positif, statistiquement significatif ou d'ampleur importante (au moins 0,25 de déviation standard)		
Standards des What Works Clearinghouse (WWC)	NA Étude établissant des corrélations avec des contrôles pour les biais de sélection	Les critères du WWC sont remplis <u>avec</u> réserve (RCT ou méthodes quasi expérimentales)	Les critères du WWC sont remplis <u>sans</u> réserve (RCT)

Source : <https://innovation.ed.gov/files/2016/06/i3fundwebinar.pdf>

Les projets soumis doivent être appuyés par des études respectant le niveau de preuve demandé

Pour être éligibles à ces subventions, les projets soumis doivent satisfaire les niveaux de preuve demandés. Le non-respect de ces conditions entraîne l'inéligibilité et cela quelle que soit la note obtenue à l'issue de la procédure d'examen.

Les propositions soumises doivent mentionner à l'appui de leurs projets jusqu'à quatre études pour les subventions de type « scale-up » et jusqu'à trois études pour les subventions de type « validation ». Ces études, qui doivent figurer dans une annexe spécifique (l'annexe D), font l'objet d'une revue par l'administration en charge du programme, the Department's Office of Innovation and Improvement (OII), en lien avec l'Institute of Education Sciences (IES). Cette revue permet de vérifier que les études citées sont pertinentes au regard du projet (elles portent sur un dispositif proche) et respectent les conditions exigées sur les niveaux de preuve d'efficacité (voir ci-dessus). Les candidats doivent s'assurer que les études citées sont publiques et indiquer où elles sont disponibles (liens internet ou autres indications).

Les projets retenus devront être ensuite évalués

Les candidats doivent prévoir une évaluation indépendante de leur projet, c'est-à-dire conçue et réalisée indépendamment d'eux mais en coordination avec eux (voir registre fédéral). Le

¹⁰¹ La théorie d'action fait référence à un cadre conceptuel qui identifie les différentes composantes du projet ainsi que les relations possibles entre ces composantes et les variables d'intérêt.

plan d'évaluation doit être détaillé dans les offres des candidats. Sa qualité est évaluée et compte pour 20 % de la note finale attribuée au projet¹⁰².

Les méthodes employées pour l'évaluation doivent ainsi permettre de produire des preuves rigoureuses sur l'efficacité des dispositifs, c'est-à-dire qui respectent les standards du What Works Center Clearinghouse¹⁰³. Il s'agit donc exclusivement de méthodes expérimentales et quasi expérimentales.

Les ressources financières allouées à l'évaluation doivent également être suffisantes pour mettre en œuvre le plan d'évaluation envisagé. Mais contrairement au programme de prévention des grossesses juvéniles (TPP) du ministère de la santé, le programme Investing in Innovation (i3) du ministère de l'éducation ne donne pas d'indication sur le montant de la subvention pouvant être alloué au financement de l'évaluation.

Par ailleurs, les candidats s'engagent, s'ils sont retenus, à mettre à jour régulièrement le plan d'évaluation (d'abord dans les cent jours suivant l'attribution des subventions puis annuellement). Enfin, ils devront rendre publics les résultats des évaluations financées par eux et mettre à disposition de chercheurs tiers les données ayant servi aux évaluations tout en garantissant les conditions de sécurité des données confidentielles (pour les subventions de type « scale-up » et « validation » uniquement).

¹⁰² Plusieurs critères sont examinés pour juger de la qualité du plan d'évaluation. Nous n'en détaillons ici que deux (la force de la preuve et le coût financier des évaluations).

¹⁰³ Les critères du WWC doivent être remplis sans réserve pour les subventions de type « scale-up » et « validation » et avec réserve pour les subventions de type « development ».

Annexe 7 Pay for Success

Examen du Projet Rikers Island de la ville de New York

Aux États-Unis, le premier projet de type Pay for Success (PFS) a été lancé en septembre 2012 par la ville de New York¹⁰⁴. L'administration locale a vu dans ce mode de financement une façon de soutenir un programme innovant dans une période budgétaire contrainte¹⁰⁵. Ce programme était destiné à réduire la récidive des jeunes détenus (âgés de 16 à 18 ans) de la prison de Rikers Island qui était très forte (presque un jeune détenu sur deux avait un risque de retourner en prison dans l'année suivant sa libération).

Le programme, appelé Adolescent Behavioral Learning Experience (ABLE), s'appuie sur une forme de thérapie comportementale et cognitive (TCC) connue sous le nom de Moral Reconciliation Therapy (MRT) qui vise à améliorer les compétences sociales, la responsabilité personnelle et la prise de décision. Cette forme de thérapie avait été choisie essentiellement pour deux raisons : d'une part parce qu'elle avait démontré la preuve de son efficacité toutefois dans d'autres contextes que la prison ; d'autre part parce qu'elle était suffisamment souple pour être adaptée au contexte spécifique de la prison¹⁰⁶.

Ce programme a été administré par deux associations, Osborne Association et Friends of Island Academy. Il a été financé par Goldman Sachs à hauteur de 9,6 millions de dollars. Pour diminuer le risque encouru par l'investisseur, Bloomberg Philanthropies a servi de garant, en apportant une garantie de prêt de 7,2 millions de dollars (soit 75 % du montant total investi)¹⁰⁷. La ville de New York devait effectuer des paiements dont le montant était calculé en fonction de la baisse du taux de récidive due au programme (mesuré en nombre de jours en prison). La ville n'était tenue d'effectuer aucun paiement si le taux de récidive était abaissé de moins de 8,5 % grâce au programme. En revanche, à partir de ce taux, elle devait rembourser un montant compris entre 4,8 et 11,7 millions de dollars selon la baisse du taux de récidive. Dans ce schéma, l'investisseur pouvait récupérer sa mise initiale pour une réduction du taux de récidive au moins égale à 10 %, et pouvait espérer un rendement

¹⁰⁴ Pour plus de détails sur ce projet, voir le site www.payforsuccess.org/project/nyc-able-project-incarcerated-youth.

¹⁰⁵ Ce projet s'inscrivait dans une initiative beaucoup plus large à l'adresse des jeunes hommes, Young Men's Initiative, lancée en août 2011 par l'administration du maire de New York. Elle visait à améliorer la vie des hommes noirs et latinos, confrontés plus souvent que leurs homologues blancs et asiatiques, à des problèmes de pauvreté, de chômage et de violence. Comme les fonds disponibles pour cette initiative étaient globalement insuffisants par rapport aux propositions jugées intéressantes, l'administration locale a décidé de mettre en place ce mode de financement pour l'un des projets. Le choix s'est finalement porté sur celui de Rikers Island (parmi 40 projets possibles) pour diverses raisons tenant à ses caractéristiques (il était notamment appuyé par des études attestant de son efficacité ; il était également susceptible de générer des économies de coût importantes pour la ville) et une volonté politique de mettre en place un programme pour réduire la récidive des jeunes détenus de la prison de Rikers Island.

¹⁰⁶ Rudd, Nicoletti, Misner et Bonsu (2013).

¹⁰⁷ La garantie de prêt apportée par Bloomberg Philanthropies ne supprimait pas totalement le risque encouru par l'investisseur mais le réduisait substantiellement (la perte encourue par l'investisseur n'étant plus dans ce cas que de 2,4 millions de dollars au terme de la quatrième année de mise en œuvre du programme).

maximal de 2,1 millions de dollars pour une réduction du taux de récidive de 20 % au minimum. La ville pour sa part pouvait espérer générer des économies de coût comprises entre 1 million de dollars et 20,5 millions de dollars selon la réduction du taux de récidive (voir tableau A7-1). MDRC jouait le rôle d'intermédiaire et Vera Institute of Justice celui d'évaluateur indépendant (voir figure A7-1).

Les résultats d'évaluation obtenus à l'aide d'une méthode quasi expérimentale¹⁰⁸ ont montré que le projet n'avait pas eu d'effet sur le taux de récidive au terme de la première année. Suite à cette évaluation, Goldman Sachs a décidé de mettre fin au projet, et la ville de New York n'a eu à procéder à aucun paiement.

Tableau A7-1 – Les paiements et les économies de coût associés au programme de Rikers Island

Réduction du taux de récidive	Paiement de la ville de New York (en millions de dollars)	Économies de coût à long terme générées par le programme (en milliers de dollars)
≥ 20 %	11,712	20,5
≥ 16 %	10,944	11,7
≥ 13 %	10,368	7,2
≥ 12,5 %	10,272	6,4
≥ 12,0 %	10,176	5,6
≥ 11,0 %	10,080	1,7
≥ 10,0 %	9,600	≥ 1,0
≥ 8,5 %	4,800	≥ 1,0

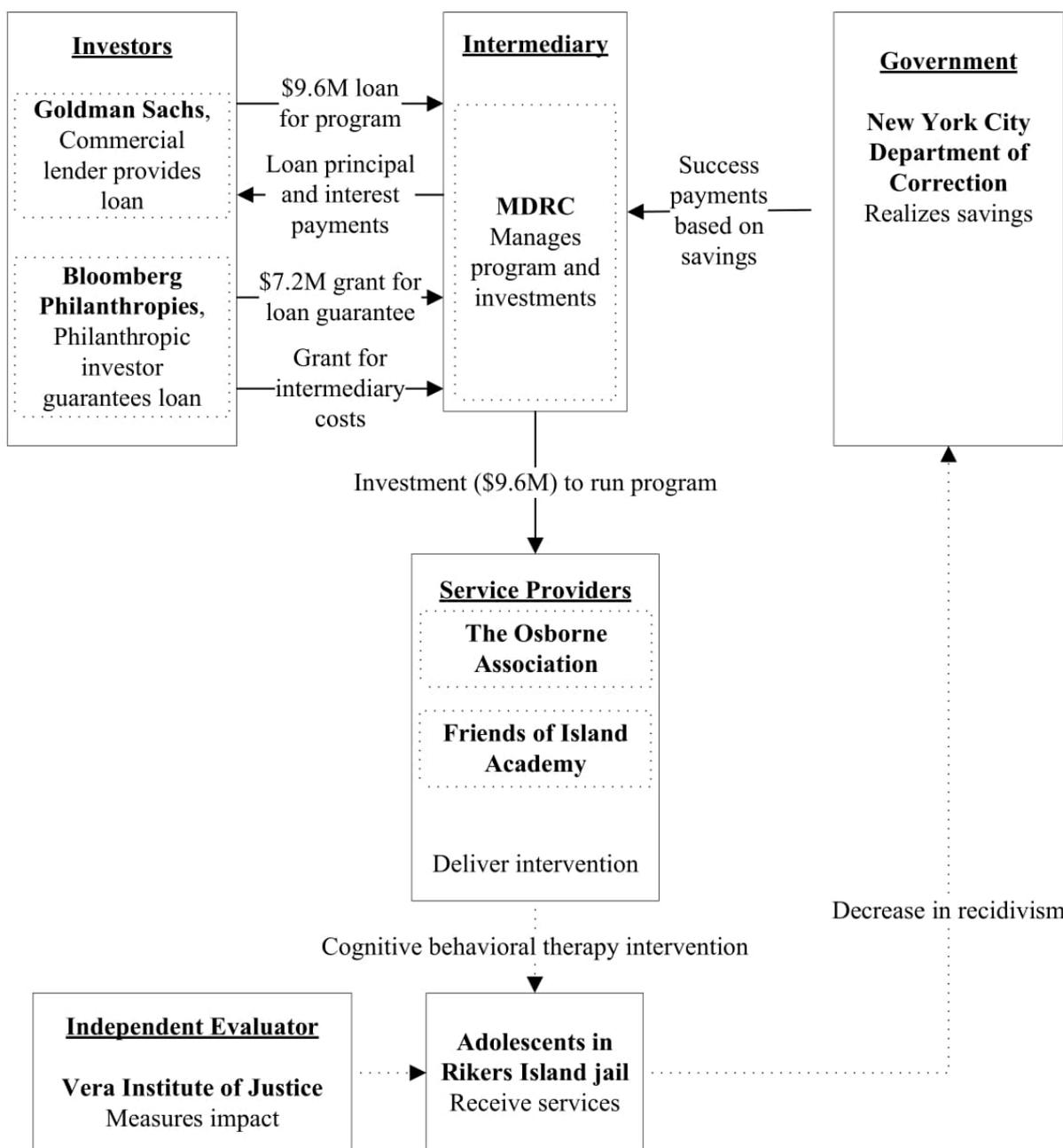
Note : il s'agit des économies de coût nettes c'est-à-dire après le paiement à MDRC et des coûts de mise en œuvre du programme sur deux années supplémentaires (années 5 et 6). Elles sont uniquement liées à la diminution de la population carcérale, et ne tiennent pas compte des bénéfices pour les bénéficiaires, les victimes, les familles, etc.

Source : Fact Sheet: the NYC ABLÉ Project for Incarcerated Youth – America's First Social Impact Bond.

Disponible à l'adresse suivante : www.payforsuccess.org/sites/default/files/resource-files/nyc_sib_fact_sheet_0.pdf.

¹⁰⁸ Vera Institute of Justice (2015).

Figure A7-1 – Structure du projet Pay for Success ABE



Note : cette figure illustre les relations entre les parties prenantes pour le projet ABE. Les lignes pleines indiquent les flux financiers, tandis que les lignes en pointillé indiquent des activités ou des services.

Source : Rudd, Nicoletti, Misner et Bonsu (2013).

Références et sources

Entretiens

Les entretiens ont été menés par l'auteur. Certains d'entre eux ont été réalisés avec l'aide de Sandra Debu et de Chloé Zapha (Département Économie, France Stratégie).

Mary Ann Bates, Executive Director, J-PAL North America, *entretien réalisé le 30 janvier 2019.*

Dorian Carloni, Analyst in the Tax Analysis Division, U.S. Congressional Budget Office (CBO), *entretien réalisé le 12 juillet 2018.*

Diana Epstein, Evidence Team Lead, Office of Management and Budget (OMB), *entretien réalisé le 7 février 2019.*

Molly Erwin, Chief Evaluation Officer, U.S. Department of Labor, *entretien réalisé le 9 janvier 2019.*

Sebastien Gay, Assistant Director Financial Analysis Division, U.S. Congressional Budget Office (CBO), *entretien réalisé le 24 janvier 2019.*

Jon Baron, Vice President of Evidence-Based Policy, Jon Ventures, *entretien réalisé le 12 février 2019.*

Naomi Goldstein, Deputy Assistant Secretary for Planning, Research, and Evaluation Administration for Children and Families, U.S. Department of Health and Human Services et **Emily Schmitt**, Deputy Director of OPRE, *entretien réalisé le 4 mars 2019.*

Matthew Soldner, Commissioner of the National Center for Education Evaluation and Regional Assistance (NCEE), one of the four centers of the Institute of Education Sciences, U.S. Department of Education, *entretien réalisé le 6 mars 2019.*

Demetra Nightingale, Institute Fellow at the Urban Institute and Chief Evaluation Officer at the U.S. Department of Labor from 2011-2016, *entretien réalisé le 6 mars 2019.*

Kelly Bidwell, Director, Office of Evaluation Sciences (OES), General Services Administration, *entretien réalisé le 27 mars 2019.*

Ouvrages, articles et rapports consultés

American Evaluation Association :

[Guiding Principles for Evaluators](#), June 2018.

[An Evaluation Roadmap For A More Effective Government](#), octobre 2013.

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.

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

Department of Health and Human Services, Office of Planning, Administration for Children and Families (ACF), Office of Planning, Research, and Evaluation (OPRE) :

Duggan A., Portilla X. A., Filene J. H., Crowne S. S., Hill C. J., Lee H. et Knox V. (2018), *Implementation of Evidence-Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Program Evaluation*, [OPRE Report 2018-76A](#), Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, octobre.

Michalopoulos C., Faucetta K., Hill C. J., Portilla X. A., Burrell L., Lee H., Duggan A. et Knox V. (2019), *Impacts on Family Outcomes of Evidence-Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Program Evaluation*, [OPRE Report 2019-07](#), Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, janvier.

Michalopoulos C., Lee H., Duggan A., Lundquist E., Tso A., Crowne S. S., Burrell L., Somers (J.), Filene (J. H.) et Knox (V.) (2015), *Evaluation Early Findings on the Maternal, Infant, and Early Childhood Home Visiting Program*, A report to Congress, [OPRE Report 2015-11](#), Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, janvier.

Department of Education, Institute of Education Sciences (IES), National Center for Education Evaluation and Regional Assistance (NCEE) :

Betts J., Dynarski M. et Feldman J. (2016), *Evaluation of the DC Opportunity Scholarship Program: Features of Schools in DC*, [NCEE Evaluation Brief](#), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education, août.

Dynarski M., Betts J. et Feldman J. (2016), *Applying to the DC Opportunity Scholarship Program: How Do Parents Rate Their Children's Current Schools at Time of Application and What Do They Want in New Schools?*, [NCEE Evaluation Brief](#), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education, avril.

Dynarski M., Rui N., Webber A. et Gutmann B. (2017), *Evaluation of the DC Opportunity Scholarship Program: Impacts After One Year*, [NCEE Report 2017-4022](#), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education, juin.

Dynarski M., Rui N., Webber A. et Gutmann B. (2018), *Evaluation of the DC Opportunity Scholarship Program: Impacts After Two Years*, [NCEE Report 2018-4010](#), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education, mai.

Feldman J., Lucas-McLean J., Gutmann B., Dynarski M. et Betts J. (2014), *Evaluation of the DC Opportunity Scholarship Program: An Early Look at Applicants and Participating Schools Under the SOAR Act*, [NCEE Report 2015-4000](#), Washington, DC:

National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education, octobre 2014.

Evidence-Based Policy making Collaborative :

- Schupmann W. et Fudge K. (2018), [Chief Evaluation Officers](#), Toolkit, juin.
- Nightingale D.S., Fudge K. et Schupmann W. (2018), [Evidence Toolkit: Learning Agendas](#), Research Report, mars.
- Feldman A. et Haskins R. (2016), [Tiered-Evidence Grantmaking](#), Toolkit, septembre.
- Milner J. (2016), [Pay for Success](#), Toolkit, septembre.
- Haskins R. et Joo N. (2017), [Tiered Evidence: What Happens when Evidence-Based Teen Pregnancy Programs Are Scaled-Up to New Sites?](#), *Policy Brief*, mai.
- Farb A.F. et Margolis A.L. (2016), *The Teen Pregnancy Prevention Program (2010-2015): Synthesis of Impact Findings*, *American Journal of Public Health*, 106 (S1): S9-S15.
- Feldman A. (2017), [Strengthening Results-Focused Government: Strategies to Build on Bipartisan Progress in Evidence-Based Policy](#), Brookings.
- Furubo J. E., Rist R. C. et Sandahl (R.) (2002), *International Atlas of Evaluation*, Transaction Publishers.

Government Accountability Office (GAO) :

- Performance Measurement and Evaluation*, [GAO-11-646SP](#), mai 2011.
- Designing Evaluations*, [GAO-12-208G](#), janvier 2012.
- Pay For Success, Collaboration among Federal Agencies Would Be Helpful as Governments Explore New Financing Mechanisms*, [GAO-15-646](#), septembre 2015.
- Tiered Evidence Grants, Opportunities Exist to Share Lessons from Early Implementation and Inform Future Federal Efforts*, [GAO-16-818](#), septembre 2016.
- Program Evaluation, Annual Agency-Wide Plans Could Enhance Leadership Support for Program Evaluations*, [GAO-17-743](#), septembre 2017.
- Interagency Council on Evaluation Policy, *Top-5 list of issues and solutions related to Federal evaluation activity*, Memorandum to the Commission on Evidence-based Policy, novembre 2016.

JPAL :

- Insuring the Uninsured*, [J-PAL Policy Briefcase](#), janvier 2014.
- Moving to Opportunity*, [J-PAL Policy Briefcase](#), mai 2015.
- Landenberger N. A. et Lipsey M. W. (2005), *The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment*, *Journal of experimental criminology*, 1(4), p. 451-476.
- Lenihan A. (2013), [Lessons from abroad: International approaches to promoting evidence-based social policy](#), Alliance for Useful Evidence, juillet.

Mathematica Policy Research :

- [Updates findings from the HHS Teen Pregnancy Prevention Evidence Review : August 2015 through October 2016](#), avril 2018.

Identifying Programs that Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors, [Review Protocol Version 5.0](#), Updated findings, avril 2016.

D'Amico R., Dunham K., Chavoya-Perez V., Kogan D., Mack M., Negoita M., Paprocki A., McConnell S. et Rosenberg L. (2015), *Providing Public Workforce Services to Job Seekers: Implementation Findings on the WIA Adult and Dislocated Worker Programs*, [Rapport](#), novembre.

McConnell S., Fortson K., Rotz D., Schochet P., Burkander P., Rosenberg L., Mastri A. et D'Amico R. (2016), *Do Job Seekers Benefit from Services Provided by the WIA Adult and Dislocated Worker Programs? Impacts at 15 Months*, [Rapport](#), mai.

Fortson K., Rotz D., Burkander P., Mastri A., Schochet P., Rosenberg L., McConnell S. et D'Amico R. (2017), *Providing Public Workforce Services to Job Seekers: 30-Month Impact Findings on the WIA Adult and Dislocated Worker Programs*, [Rapport](#), mai.

National Academies of Sciences, Engineering, and Medicine, [Principles and Practices for Federal Program Evaluation: Proceedings of a Workshop](#), The National Academies Press, 2017.

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

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

Office of Management and Budget (OMB), [Overview of Federal Evidence-Building Efforts](#), White Paper, 15 juillet, 2016.

Perret B. (2008), *L'évaluation des politiques publiques*, Paris: La Découverte.

Pew-Mac Arthur First Initiative, [How States engage in Evidence-Based Policymaking, a national assessment?](#), rapport, janvier 2017.

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

Rudd T., Nicoletti E., Misner K. et Bonsu J. (2013), [Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond](#), MDRC, décembre.

Vera Institute of Justice, [Impact Evaluation of the Adolescent Behavioral Learning Experience \(ABLE\) Program at Rikers Island](#), Summary of Findings, juillet 2015.

Yarbrough D. B., Shulha L. M., Hopson R. K. et Caruthers F. A. (2011), [The program evaluation standards: A guide for evaluators and evaluation users \(3rd ed.\)](#), Thousand Oaks, CA: Sage.

Textes réglementaires

Social Security Act, Section 115.

Consolidated Appropriations Act, 2016, Division H, title II, [Public Law 114-113](#), 18 décembre 2015, 129 Stat. 2616.

Consolidated Appropriations Act, 2017, Division H, Title I, Section 107, [Public Law 115-31](#), 5 mai 2017, 131 Stat. 135.

Federal Register, vol. 81, n° 94, Monday, May 16, Notices. [Department of Education. Notice Inviting Applications \(NIAs\), Investing in Innovation Fund—Scale-Up Grants.](#)

Federal Register, vol. 81, n° 94, Monday, May 16, Notices. [Department of Education. Notice Inviting Applications \(NIAs\), Investing in Innovation Fund—Validation Grants.](#)

Federal Register, vol. 81, n° 79, Monday, April 25, Notices. [Department of Education. Notice Inviting Applications \(NIAs\), Investing in Innovation Fund—Development Grants.](#)

Foundations for Evidence-Based Policymaking Act of 2018, [texte](#)

Patient Protection and Affordable Care Act of 2010, [texte](#)

Office of Management and Budget (OMB) :

Building and using evidence to improve government effectiveness, Budget of the U.S. Government, [Analytical Perspectives](#), chapter 6, Fiscal Year 2019.

Budget of the U.S. Government, [An American Budget - President's Budget](#), Fiscal Year 2019.

Budget of the U.S. 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.

Monitoring and Evaluation Guidelines for Federal Departments and Agencies that Administer United States Foreign Assistance, M-18-04, Washington, D.C., 11 janvier 2018.

Scholarships for Opportunity and Results (SOAR) Act of 2011, [texte](#).

Principaux sites internet consultés

White House

www.whitehouse.gov/omb/

<https://obamawhitehouse.archives.gov/omb/evidence>

<https://obamawhitehouse.archives.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american>

<https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/CoT-SBST-Charter-0315-SIGNED.pdf>

www.opm.gov/policy-data-oversight/senior-executive-service/executive-core-qualifications/#url=Overview

www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/

Agences fédérales

U.S. Department of Labor

www.dol.gov/asp/evaluation/

www.dol.gov/asp/evaluation/EvaluationPolicy.htm

www.dol.gov/asp/evaluation/CurrentStudies.htm

www.dol.gov/asp/evaluation/CompletedStudies.htm

www.dol.gov/asp/evaluation/completed-studies/CEO-FY-2018-Evaluation-Plan.pdf

<https://clear.dol.gov/>

<https://fr.slideshare.net/OECD-GOV/molly-irwin-evidence-informed-policy-making-27-june-2017>

www.dol.gov/asp/evaluation/currentstudies/Labor-Research-and-Evaluation-Grants.htm

U.S. Department of Education

<https://ies.ed.gov>

<https://legcounsel.house.gov/Comps/Education%20Sciences%20Reform%20Act%20Of%202002.pdf>

https://ies.ed.gov/director/sro/peer_review/report_review.asp

<https://ies.ed.gov/director/board/index.asp>

<https://ies.ed.gov/pdf/CommonGuidelines.pdf>

<https://ies.ed.gov/ncee/wwc/>

<https://ies.ed.gov/ncee/wwc/WhatWeDo>

<https://ies.ed.gov/ncee/wwc/Handbooks>

https://ies.ed.gov/ncee/pubs/PooledEval_FY2018/pdf/PooledEval_FY2018_to_FY2019.pdf

https://ies.ed.gov/ncee/projects/pdf/IESEvaluationPrinciplesandPractices_011117.pdf

<https://ies.ed.gov/ncee/projects/evaluation/index.asp>

https://ies.ed.gov/ncee/projects/evaluation/evaluations_filter.asp

<https://ies.ed.gov/ncee/pubs/>

https://ies.ed.gov/ncee/projects/evaluation/choice_impactmagnet.asp

<https://innovation.ed.gov/files/2016/06/i3fundwebinar.pdf>

U.S. Department of Health and Human Services

www.acf.hhs.gov/opre

www.acf.hhs.gov/opre/resource/acf-evaluation-policy

www.acf.hhs.gov/sites/default/files/opre/acf_common_framework_for_research_and_evaluation_v02_a.pdf

www.acf.hhs.gov/opre/resource-library/search?sort=recent#?type%5B5548%5D=5548&sort=recent&ajax=1

www.acf.hhs.gov/opre/resource-library/search?sort=recent#?type%5B3076%5D=3076&sort=recent&ajax=1

www.acf.hhs.gov/opre/resource/effects-four-responsible-fatherhood-programs-low-income-fathers-evidence-parents-and-children-together-evaluation

<https://homvee.acf.hhs.gov/Default.aspx>

<https://tppevidencereview.aspe.hhs.gov/>

<https://tppevidencereview.aspe.hhs.gov/FindAProgram.aspx>

<https://rtips.cancer.gov/rtips/index.do>

<https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting-overview>

Autres ministères

<https://oes.gsa.gov>

<https://oes.gsa.gov/work>

https://oes.gsa.gov/assets/files/GSA_OES_SolicitationFY19v2.pdf

<https://oes.gsa.gov/projects/quetiapine-prescribing/>

www.wsipp.wa.gov

www.huduser.gov/portal/research/home.html

www.huduser.gov/portal/about/pdr_roadmap.html

www.huduser.gov/portal/evidence.html

www.govinfo.gov/content/pkg/FR-2016-12-06/pdf/2016-29215.pdf

<https://usaidlearninglab.org/evaluation>

www3.epa.gov/evaluate/Im-training/ [ou www.epa.gov/evaluate](http://www.epa.gov/evaluate)

www.nationalservice.gov/resources/evaluation

www.cdc.gov/eval/resources/index.htm

www.crimesolutions.gov

<https://clinicaltrials.gov>

Congressional Budget Office

www.cbo.gov/about/10-things-to-know

General Accountability Office

www.gao.gov/about/gao_at_a_glance_2017_english.pdf

J-Pal

www.povertyactionlab.org/na

www.povertyactionlab.org/stateandlocal/aboutus

www.povertyactionlab.org/evaluations

www.povertyactionlab.org/na/cmto

MDRC

www.mdrc.org/projects

www.mdrc.org/method/methodology

www.mdrc.org/investing-future#overlay-context=about/about-mdrc-overview-0

Mathematica Policy Research

www.mathematica-mpr.com/our-publications-and-projects/projects#q1=&q2=&keyword=&p=1&sortBy=date&sortdir=desc&facets=&datefacets=

<https://cire.mathematica-mpr.com/about-cire>

<https://cire.mathematica-mpr.com/about-cire>

Mac Arthur Foundation

www.macfound.org/learning/?q=&type=assess#learning-search

www.pewtrusts.org/en/projects/pew-macarthur-results-first-initiative

www.pewtrusts.org/~media/data-visualizations/interactives/2018/rfcd/download/results_first_clearinghouse_database_technical_appendix.pdf

www.pewtrusts.org/~media/assets/2015/06/results_first_clearinghouse_database_user_guide.pdf

www.pewtrusts.org/en/research-and-analysis/data-visualizations/2015/results-first-clearinghouse-database

Laura and John Arnold Foundation (LJAF)

www.arnoldventures.org/

<http://evidencebasedprograms.org>

www.straighttalkonevidence.org

<https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/Key-Items-to-Get-Right-When-Conducting-Randomized-Controlled-Trials-of-Social-Programs.pdf>

<https://craftmediabucket.s3.amazonaws.com/uploads/Grants-for-Randomized-Controlled-Trials-Administered-by-the-LJAF-Evidence-Based-Policy-Team-2019.01.10.pdf>

Evidence-Based Policy Collaborative

www.evidencecollaborative.org

American Evaluation Association

www.eval.org/p/cm/ld/fid=4

Universités

www.aisp.upenn.edu/

www.capolicylab.org

<https://urbanlabs.uchicago.edu/>

<https://leo.nd.edu/about/>

Nonprofit Finance Fund

www.payforsuccess.org/projects/comparison

www.payforsuccess.org/project/nyc-able-project-incarcerated-youth

www.payforsuccess.org/resource/fact-sheet-nyc-able-project-incarcerated-youth

www.payforsuccess.org/sites/default/files/resource-files/nyc_sib_fact_sheet_0.pdf

Autres

www.socialscienceregistry.org

<https://osf.io>

<https://sreereg.icpsr.umich.edu/pages/about.php>

www.3ieimpact.org/evidence-hub/impact-evaluation-repository

www.nber.org/mtopublic/

www.nber.org/mtopublic/news.htm

www.nber.org/oregon/

www.nber.org/oregon/7.media.html

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.