

Dispositifs d'évaluation des politiques publiques et des programmes : connaissances de base, choix des méthodes, sociogramme des acteurs et études de cas

Maurice Baslé, Jean Michel Josselin, Benoît Le Maux

▶ **To cite this version:**

Maurice Baslé, Jean Michel Josselin, Benoît Le Maux. Dispositifs d'évaluation des politiques publiques et des programmes : connaissances de base, choix des méthodes, sociogramme des acteurs et études de cas. In Rapport à l'Assemblée nationale : Etude pour le Comité d'évaluation et de contrôle des politiques publiques de l'Assemblée Nationale, 2018. halshs-01962582

HAL Id: halshs-01962582
https://shs.hal.science/halshs-01962582

Submitted on 20 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dispositifs d'évaluation des politiques publiques et des programmes : connaissances de base, choix des méthodes, sociogramme des acteurs et études de cas

Etude pour le Comité d'évaluation et de contrôle des politiques publiques de l'Assemblée Nationale

13 février 2018

Maurice BASLÉ, professeur émérite des universités en sciences économiques

Jean-Michel JOSSELIN, professeur des universités en sciences économiques

Benoît LE MAUX, maître de conférences (HDR) en sciences économiques

Université de Rennes 1

Centre de Recherche en Économie et Management (UMR CNRS 6211)

maurice.basle@univ-rennes1.fr
jean-michel.josselin@univ-rennes1.fr
benoit.le-maux@univ-rennes1.fr

SOMMAIRE

RESUME **3**

1 UNE REVUE DES CONNAISSANCES DU DOMAINE DE L'EVALUATION DES POLITIQUES PUBLIQUES ET DES PROGRAMMES (EPPP) **4**

1.1 STRATEGIE DE RECHERCHE DOCUMENTAIRE 4

1.2 DEFINITIONS, CRITERES, ETAPES ET PRINCIPES 7

1.3 DES EVALUATIONS DISTINGUEES SELON DES CRITERES DE PERTINENCE, EFFICACITE, EFFICIENCE ET COHERENCE. 9

1.4 LES ETAPES DE L'EVALUATION 11

1.5 LE RESPECT DE PRINCIPES ETHIQUES ET DEONTOLOGIQUES 13

2 DES CONTROVERSES SUR LES METHODES ET UN MODELE D'AIDE A LA DECISION **15**

2.1 LE CHOIX DES METHODES ET LE DEGRE DE CONFIANCE DANS LES RESULTATS DE L'EVALUATION 15

2.2 DES DONNEES ET DES METRIQUES 19

2.3 DES INCERTITUDES INEVITABLES 20

2.4 POUR Y VOIR CLAIR : UN ARBRE DE DECISION ET DES TABLEAUX COMPARATIFS 21

3 SOCIOGRAMME DES ACTEURS DE L'EVALUATION EN FRANCE **28**

3.1 L'EVALUATION DES POLITIQUES PUBLIQUES EN FRANCE 28

3.2 LES COMMANDITAIRES EN EPPP 30

3.3 LES PRESTATAIRES EN EPPP 34

3.4 BENEFICIAIRES, NON-BENEFICIAIRES ET CONTRIBUABLES : EVALUATION, DEMOCRATIE ET COMMUNICATION 36

4 DES ETUDES DE CAS POUR ILLUSTRER LES PRATIQUES DE L'EVALUATION EN FRANCE **38**

4.1 ÉTUDE QUALITATIVE : EVALUATION DU PROGRAMME DE REUSSITE EDUCATIVE 38

4.2 ÉVALUATION PARTICIPATIVE : POLITIQUES DE DEPLACEMENT DE NUIT DANS L'AGGLOMERATION NANTAISE 41

4.3 ÉVALUATION EXPERIMENTALE : DISPOSITIFS D'AIDE A L'ACCES A L'EMPLOI DES JEUNES DIPLOMES 44

4.4 ÉVALUATION QUASI-EXPERIMENTALE : LES EFFETS DU RECRUTEMENT EN CONTRAT AIDE SUR LA TRAJECTOIRE PROFESSIONNELLE 46

4.5 ANALYSE D'IMPACT BUDGETAIRE : GENESE ET CONSTRUCTION DU GUIDE METHODOLOGIQUE DE LA HAUTE AUTORITE DE SANTE SUR L'IMPACT BUDGETAIRE DES MEDICAMENTS INNOVANTS ET ONEREUX 49

4.6 ANALYSE COUT-BENEFICE : EVALUATION SOCIO-ECONOMIQUE DES INVESTISSEMENTS PUBLICS DU COMMISSARIAT GENERAL A L'INVESTISSEMENT 52

ANNEXE 1. APERÇU DES SITES DE RESSOURCES ET EXTRACTION DE LEURS CONTENUS **57**

ANNEXE 2. DISCUSSION SUR LES CRITERES DE COMPARAISON DES METHODES **60**

ANNEXE 3. RESSOURCES BIBLIOGRAPHIQUES **62**

LISTE DES FIGURES **63**

LISTE DES TABLEAUX **63**

LISTE DES ENCADRES **63**

Résumé

Le présent rapport vise à couvrir le champ des connaissances utiles aujourd'hui en évaluation des politiques publiques et des programmes (EPPP), à rendre compte des différentes méthodes existantes et de leurs forces et faiblesses, à présenter les principaux acteurs de l'évaluation en France et, enfin, à fournir quelques cas pratiques d'EPPP en France. Le rapport est d'abord fondé sur une revue documentaire à large spectre dans le temps et dans l'espace. L'EPPP est une activité aux origines lointaines pour certains champs de politiques publiques (éducation, aide sociale, santé), une activité qui a pris de nouvelles dimensions à partir des années 1970 tant en France que dans les grandes organisations internationales dont la Commission européenne. Le champ des publications est dominé par les revues et les ouvrages publiés en anglais. La revue documentaire dans ce champ des EPPP bute sur son manque de structuration tant au plan des publications de la recherche (plusieurs disciplines sont concernées), qu'au plan professionnel (grande diversité et éparpillement des pratiques des commanditaires et des prestataires).

Le rapport est d'abord une tentative de revue des connaissances en EPPP que le lecteur doit maîtriser (partie 1). Un consensus existe sur les définitions de l'activité d'EPPP, sur les critères de pertinence, efficacité, efficience et cohérence qui sont utilisés et portés dans les cahiers des charges (ou termes de référence). Le consensus existe aussi sur la description des étapes nécessaires à la production de bonnes EPPP et sur le nécessaire respect de principes éthiques et déontologiques.

Le rapport fournit ensuite une revue des méthodes (partie 2). La technicité est ici élevée. En outre, le choix des outils et des méthodes suscite des controverses. Il est exposé que les choix de méthode ne sont jamais neutres car ils influencent le degré de confiance dans les résultats des EPPP. Pour que le lecteur soit éclairé, un arbre de décision entre les différentes méthodes est ici proposé. L'arbre de décision publié dans ce rapport doit permettre à un commanditaire d'EPPP d'opter pour telle ou telle méthode en fonction des qualités qui sont requises : ici, la complétude, la simplicité, la transparence et la transférabilité des résultats. Une discussion pourrait conduire à augmenter le nombre de ces qualités.

Le rapport tient ensuite compte du fait qu'en France, l'institutionnalisation des EPPP n'est pas encore achevée et qu'en conséquence les acteurs sont encore multiples et hétérogènes. Il contient un premier état des lieux à travers un sociogramme (partie 3).

Enfin, afin de maitriser la réalité des activités d'EPPP en France, six études de cas sont présentées dans le rapport (partie 4). Elles sont analysées (commanditaires, objectifs, méthodes, résultats, recommandations) et discutées. Le lecteur pourra se forger à leur lecture une compréhension d'ensemble du domaine des EPPP.

1 Une revue des connaissances du domaine de l'évaluation des politiques publiques et des programmes (EPPP)

1.1 Stratégie de recherche documentaire

L'évaluation des politiques publiques est une prestation intellectuelle de connaissance de l'action publique. Ce caractère « cognitif » est reconnu tant en France qu'à l'étranger. La prestation évaluative mobilise des moyens humains et matériels spécifiques et permanents tant au service de la commande qu'à celui de l'usage des rapports et du suivi des recommandations.

Compte tenu de son champ, les politiques publiques et leurs programmes, la prestation d'EPPP obéit à un ensemble de règles strictes qui doivent permettre de conjuguer rigueur scientifique dans l'aide à la décision publique, mais aussi opérationnalité (finalité gestionnaire), pédagogie (finalité d'apprentissage et de mobilisation des agents publics), éthique (finalité de compte-rendu), et surtout enrichissement du débat politique : ouverture au débat d'experts, au débat citoyen et préparation à l'animation des controverses (finalité informative et finalité démocratique).

Une revue des connaissances d'une manière générale[1] et *a fortiori* pour un domaine comme l'Evaluation des politiques publiques et des programmes doit permettre à tout un chacun d'avancer dans le repérage des dimensions du thème (dimension socio-politique, dimension économique et de gestion, dimension de l'activité d'évaluation elle-même comme pratique à la fois scientifique et professionnelle). La revue des connaissances permet également de sélectionner des concepts et mots-clés qui permettent à tout lecteur curieux de réinterroger les différentes bases de connaissance ensuite.

Une revue des connaissances peut être orientée à des fins d'amélioration du savoir scientifique sur l'EPPP, à des fins de constitution d'une base de ressources et de composants pour l'enseignement, ou à des fins plus professionnelles d'amélioration de la mise en œuvre des activités d'EPPP soit comme commanditaire soit comme prestataire.

[1]Voir par exemple les références suivantes :

- Université de Bretagne-Loire (2018) *Elaborer une stratégie de recherche d'information*, https://guides-formadoct.u-bretagneloire.fr/recherchedocumentaire.
- Goarin CD, Giraud E, Mannina B, Quoniam L (1999). La méthodologie et la stratégie de recherche d'information à valeur-ajoutée sur Internet. *Le micro-bulletin thématique du CNRS* (L'information scientifique et technique et l'outil Internet, expériences, recherches et enjeux pour les professionnels de l'IST) : 47-67.
- Hart C (2009) *Doing a Literature Review*. Sage editions.
- Machi, LA et McEvoy BT (2016). *The Literature Review: Six Steps to Success*. Troisième édition.

Encadré 1. Démarche de recherche documentaire en EPPP

Pour la recherche d'ouvrages (français et étrangers) et de thèses en français, il est possible de consulter le Catalogue du système universitaire de documentation (Sudoc). Celui-ci est réalisé par les bibliothèques et centres de documentation de l'enseignement supérieur et de la recherche. Ses 12 millions de notices bibliographiques décrivent tous les types de documents (livres, thèses, revues, ressources électroniques, documents audiovisuels, etc.) ainsi que les collections de revues et journaux d'environ 2 000 établissements documentaires hors enseignement supérieur (bibliothèques municipales, centres de documentation, etc.).

En outre, pour la recherche d'articles en français ou en anglais, le moteur de recherche Isidore (http://www.rechercheisidore.fr) moissonne les données de nombreux portails documentaires dont Cairn, HAL, etc. Il est aussi possible d'utiliser le moteur Business Source Complete ainsi que les ressources des catalogues des grands éditeurs scientifiques (Elsevier, Sage, Springer, etc.).

Enfin, des remontées d'informations assez hétéroclites mais substantielles seront obtenues avec du requêtage avec les moteurs de recherche de type Google ou plus spécifiquement WorldwideScience :

https://worldwidescience.org/wws/desktop/en/results.html.

Toute stratégie de recherche documentaire commence par l'identification de mots-clés. A titre indicatif, voici une première liste sommaire en français et en anglais que le lecteur peut utiliser :

- Évaluation de politiques publiques et de programmes publics, Suivi et évaluation de programmes publics, Évaluation de performance publique, Nouvelle performance publique et évaluation, Indicateurs d'évaluation de politiques et de programmes, Management public et évaluation, Connaissance et action publique, Analyse coût-bénéfice, Analyse coût-efficacité, Évaluation expérimentale, Évaluation quasi-expérimentale, Évaluation participative.
- Evaluation of public policies and programs, Monitoring and evaluation of public programs, Public performance evaluation, New public performance and evaluation, Policy and program evaluation indicators, Public management and evaluation, Knowledge and public action, Cost-benefit evaluation, Experimental evaluation, Quasi-experimental evaluation, Participatory evaluation.

Cette liste de mots clés peut être complétée grâce à différents ouvrages ou publications à facteur d'impact élevé (voir annexe 1) et aux lexiques existants, par exemple le glossaire OCDE des principaux termes relatifs à l'évaluation : https://www.oecd.org/dac/evaluation/2754804.pdf.

Le domaine de l'EPP étant loin d'être un domaine de connaissances entièrement structuré, il ne faut pas s'attendre à faire émerger en une simple requête ou même en combinant plusieurs requêtes, pour chaque groupe de mots-clés, les principales publications, celles qui ont le meilleur facteur d'impact (grandes revues réputées, bien diffusées et grand nombre de citations). Malgré tout, une fois le domaine mieux cerné par la possibilité de l'aborder avec une liste de mots-clés, la recherche sur le plus grand ensemble de ressources documentaires (les catalogues, mais aussi les moteurs de recherche du Web, dictionnaires, encyclopédies, etc.) est faisable par tout un chacun (cf. encadré 1). Dans le cas de l'EPPP, le lecteur intéressé pourra extraire de longues listes de publications (rédigées en langue anglaise le plus souvent) et ensuite opérer différentes requêtes pour recouper les informations contenues dans les publications.

De cette revue de la littérature, il est possible de déduire certains éléments de consensus mais aussi l'existence de débats fondamentaux.

a) **Le consensus de la littérature sur le besoin d'une activité régulée d'évaluation des politiques publiques est large.** Dans les différentes pratiques nationales (au Nord ou au Sud) ou pour les organisations internationales, il s'agit d'estimer un résultat des politiques publiques et plus précisément des programmes qui leur sont associés. L'activité est encouragée : l'ONU a déclaré 2015 comme étant « l'année internationale de l'évaluation »[2].
b) **Des débats existent sur l'autorité de portage de l'activité d'EPPP :** Haute autorité indépendante, offices rattachés au parlement, agences rattachées aux exécutifs nationaux ou locaux ou de sécurité sociale. La recherche de l'indépendance du jugement évaluatif peut entraîner une juridisation de l'évaluation, sur le modèle du contrôle, par exemple, sur le modèle français de la Cour des comptes et des chambres régionales des comptes. La démocratie libérale peut être jugée suffisamment forte et ouverte à la pluralité des électeurs-citoyens pour qu'au contraire, des élus et des citoyens s'emparent du projet évaluatif. L'équilibre trouvé est variable selon les pays (Pour la France, voir chapitre 4 de l'ouvrage Baslé, M. 2008, cf. annexe 3). En France, le Conseil économique, social et environnemental a jugé, de par sa composition, qu'il était une autorité susceptible d'animer et de promouvoir une culture de l'évaluation des politiques publiques (rapport présenté par Nasser Mansouri-Guilani au nom de la délégation à la prospective et à l'évaluation des politiques publiques, avis adopté par le CESE le 8 septembre 2015).
c) **A ces débats sur la régulation des évaluations se superposent des débats plus fondamentaux :** la connaissance de l'action publique et de ses effets est, par nature, lourde

[2] Cf. http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-evaluant-ses-politiques-publiques/2015-annee-internationale-evaluation

d'enjeux politiques et il y a débat sur les rapports entre la science et les politiques publiques[3] [4]. Les angles d'attaque des sous-disciplines scientifiques sur les politiques publiques sont différents (les politiques publiques peuvent quelquefois être définies comme des annonces, ou comme des réalités budgétaires avec des réalisations effectives, ou encore comme des réalités managériales avec le souci de délivrance de services conformes à des instructions). Les querelles sur la maitrise et la portée des méthodes et des outils ne sont pas propres aux sciences de l'action publique. Certains regrettent, dans l'EPPP comme dans d'autres domaines, un recours trop massif à des approches qui seraient trop quantitatives (fondées sur des mesures traduisant mal la réalité des politiques publiques et proposent de se rapprocher davantage des opinions des bénéficiaires ou des non bénéficiaires (en recueillant des matériaux dits plus qualitatifs pouvant faire l'objet d'analyse sémantique). D'autres sont convaincus de la supériorité des seules méthodes d'expérimentation appuyées sur l'existence d'un contrefactuel (par exemple en comparant la situation des bénéficiaires de la politique expérimentée à celle de non-bénéficiaires susceptibles d'être ultérieurement concernés si cette politique est généralisée)[5].

1.2 Définitions, critères, étapes et principes

Deux constats peuvent être faits après l'étape de revue de la littérature (ouvrages, revues, sites web et moteurs de recherche usuels en science). D'une part, la convergence est réelle sur la définition de l'EPPP: il s'agit **d'estimer une valeur ajoutée par la politique publique**. D'autre part, les critères sont presque toujours les mêmes : il faut **mesurer la pertinence, l'efficacité, l'efficience et la cohérence des politiques publiques**.

Il faut cependant reconnaître que, dans la pratique, l'accent est porté en premier sur l'évaluation de l'efficacité (résultats, effets, impact) et n'est pas toujours mis sur l'efficience (manque d'évaluations des impacts budgétaires des programmes ou manque de comparaison des options à discuter, manque aussi de mise en relation de l'impact avec les coûts nécessaires à la réussite du programme) ou sur la cohérence.

Le consensus : évaluer, c'est donner une estimation de la valeur ajoutée à la société. Il suffit de retenir que dans « évaluation », il y a « valeur ». La valeur ajoutée sociétale[6] (que l'évaluation cherche à estimer) est créée par les services rendus grâce à la mise en œuvre des politiques publiques

[3] Jacobsson S et Vico EP (2014) The many ways of academic researchers: How is science made useful*? Science and public policy*, 41: 641-657. Oxford University Press.
[4] Baslé M (2010). *Connaissance et action publique.* Editions Economica. Paris.
[5] Bono PH et Trannoy A (2012) Causalité et contre-factualité dans l'évaluation économique des politiques publiques. *Revue Labyrinthe*, 2 : 35-53.
[6] Il ne s'agit pas simplement de la valeur ajoutée pour les administrations mais bien de la valeur ajoutée pour la société tout entière.

(une politique ou une mission se déclinant en programmes d'action publique). Un exemple peut-être la valeur ajoutée des enseignants par exemple dans une évaluation aux Etats-Unis[7]. Par extension, de nouvelles lois et règlementations peuvent également contribuer à créer de la valeur ajoutée collective. Ceci justifie l'évaluation d'impact des lois au même titre que l'évaluation d'impact des politiques publiques (comme par exemple des projets publics d'infrastructure, de santé, d'éducation, etc.).

Un constat émerge : l'oubli fréquent de l'évaluation de l'efficience alors qu'il faudrait donner une estimation de la valeur ajoutée sociale nette. Pour mesurer une valeur ajoutée sociétale, il ne suffit pas de considérer le coût de production des services rendus par la politique évaluée, même si cette convention est à juste titre utilisée pour les besoins des comptes nationaux. On ne saurait réduire l'action publique aux moyens qu'elle mobilise, sinon l'évaluation se ramènerait à la seule budgétisation de l'action. Cette approche comptable et faussement simple conduirait à maximiser la dépense en imaginant que le seul déploiement de moyens supplémentaires suffirait à atteindre l'objectif fixé par la politique. Il faut en fait évaluer les services rendus nets des moyens engagés pour la mise en œuvre de la politique (valeur ajoutée sociale nette).

Un autre constat émerge : l'oubli fréquent d'une approche comparative des options concurrentes pour la réalisation de la politique publique. Beaucoup d'évaluations, en pratique ne sont pas comparatives. Or l'on sait que l'estimation d'une seule voie d'action ne suffit pas. En France, le cas du projet d'aéroport de Notre dame des landes est typique de cet oubli du choix à poser entre plusieurs options de projets (un nouvel aéroport, un ancien aéroport agrandi ou bien une troisième solution de transport). Toute politique publique donnée (par exemple l'amélioration du transport urbain) doit en effet se décliner en plusieurs options ou stratégies possibles (y compris éventuellement le *statu quo*), explorées et discutées avec les parties prenantes et l'estimation doit être comparative.

Dans un environnement où les ressources sont contraintes, les bénéfices d'une option ont non seulement une contrepartie budgétaire (les moyens engagés) mais également une contrepartie économique, c'est-à-dire la renonciation à la valeur ajoutée sociétale qu'aurait générée la meilleure des options délaissées : c'est le coût d'opportunité. Par exemple, si le *statu quo* en matière de transport urbain est un réseau de bus classique, les options alternatives envisagées peuvent être le tramway, le métro, ou encore des bus à énergie propre, etc. Toutes les options ou stratégies (dont le *statu quo*) devront être évaluées par rapport à toutes les autres. D'autres coûts d'opportunité sont également à prendre en compte, mais plus difficiles à appréhender : il faudrait idéalement tenir compte du fait que l'action publique mobilise des ressources prélevées sur l'activité économique privée (par exemple via l'impôt). Ces ressources auraient pu être utilisées autrement par les agents privés (pour la

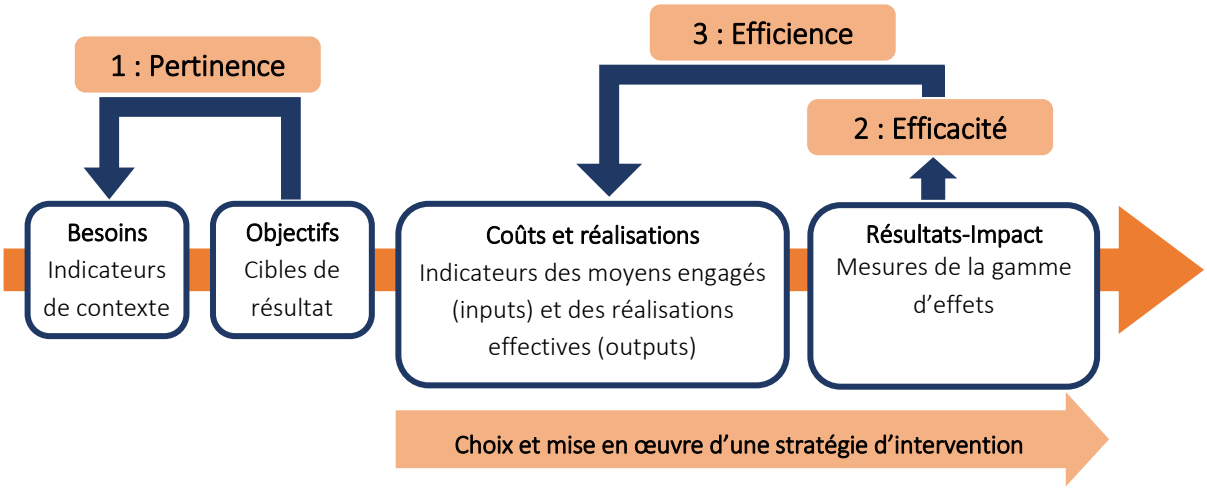
[7] Isenberg E et Hock H (2012) Measuring School and Teacher Value Added in *DC, 2011-2012 School Year. Final Report. ERIC Institute of Education Sciences*. Mathematica Policy Research.

consommation et l'investissement privé) ou par l'action publique dans d'autres domaines de politique publique (par exemple, les ressources consacrées au transport n'iront pas à l'éducation).

1.3 Des évaluations distinguées selon des critères de pertinence, efficacité, efficience et cohérence.

Une production de valeur ajoutée sociétale nette par une politique publique peut être scénarisée, comme toute activité productive, par une chaîne de valeur « inputs-outputs-résultats-impact » (cf. figure 1).

Figure 1. Chaine « inputs-outputs-résultats-impact »



Le consensus dans la littérature est assez élevé pour indiquer que l'activité d'évaluation devrait théoriquement estimer à la fois la **pertinence** (politique appropriée aux besoins), l'**efficacité** (politique aboutissant au résultat visé), et l'**efficience** (politique à la fois efficace et économisant les moyens). Et, comme différentes activités publiques peuvent être conduites en même temps pour concourir à des impacts sociétaux donnés, on n'oubliera pas de rechercher également la **cohérence** d'ensemble des activités publiques. C'est le plus difficile à réaliser. Ainsi, le budget général de l'État 2018 compte 32 missions qui doivent être servies par 123 programmes. Les méthodes d'évaluation usuelles, se concentrant généralement sur une politique donnée, sont pour l'instant mal armées pour juger de cette cohérence. Le défi de la cohérence entre les programmes et les politiques publiques est toujours présent : il faudrait savoir éviter les doublons et redondances entre programmes et se mettre en position de renforcer leur complémentarité, La revue de la littérature montre que l'on traite très peu les questions de cohérence entre programmes et politiques publiques. Une exception dans l'évaluation des

fonds européens réalisée à l'initiative de la Commission européenne : la cohérence dite inter-fonds européens est recherchée[8].

Reprenons plus en détail l'analyse de la chaîne de valeur. Dans la lecture de la figure 1, le lecteur retrouve le premier critère de la pertinence en début de chaine. L'identification des besoins d'une politique est une phase essentielle puisqu'elle conditionne l'existence même de la politique publique. Cette phase, loin d'être triviale, nécessite d'être argumentée à l'aide d'indicateurs de contexte. Par exemple, si le décrochage scolaire apparaît à des élus intuitivement comme un problème susceptible de déclencher une politique publique, il n'en faut pas moins documenter, contextualiser et rationaliser cette intuition. On calculera des taux de décrochage par âge, zone géographique, origine socio-économique des parents, pour éventuellement exclure du ciblage de la politique des zones peu touchées par le phénomène. A cette démarche statistique pourra s'ajouter une approche qualitative (comme des entretiens semi-directifs), par exemple avec des élus ou acteurs locaux.

Si le besoin d'une politique est avéré, alors il faudra définir des objectifs à atteindre (c'est une forme de management stratégique par objectif), auxquels seront associés des indicateurs de résultats ciblés, par exemple un taux de décrochage diminué de moitié sur cinq ans parmi les élèves des écoles primaires, une réinsertion des décrocheurs en moins de deux ans, etc.

Ensuite, comme indiqué précédemment, la politique publique envisagée sera déclinée en options ou stratégies possibles d'intervention, mises en concurrence au regard de leurs coûts et de leurs effets. Ces stratégies doivent être des réponses différenciées aux besoins préalablement identifiés ainsi que des façons de répondre aux objectifs. Par exemple, les déclinaisons de la politique publique de lutte contre le décrochage scolaire peuvent être l'accompagnement individualisé hors du temps scolaire, la prise en charge dans une école dédiée aux enfants en difficulté, le renforcement des équipes pédagogiques dans le système scolaire usuel, etc. Selon le type de besoin identifié, il peut être nécessaire de s'assurer du respect du droit des personnes, du droit de l'environnement, des libertés individuelles de même que de la faisabilité administrative et technique des stratégies d'intervention. Cette identification des options ou stratégies d'intervention est une phase cruciale qui doit faire l'objet d'un débat le plus ouvert possible afin de minimiser le risque d'écarter des options dont il aurait fallu à tout le moins évaluer l'impact potentiel. L'expérience montre qu'une fois la machinerie de l'évaluation lancée, il est difficile d'intégrer au processus des stratégies initialement délaissées ou ignorées (pas de rattrapage d'options après coup).

Le deuxième critère, celui qui est au cœur de la plupart des activités d'évaluation, est celui de l'**efficacité** du programme : des réalisations effectives produisent des résultats et de l'impact, plus

[8]Bachtler, J (1997) The coherence of EU regional policy: contrasting perspectives on the structural funds in *Regional policy and development series*, 17. Kingsley, London.

généralement une gamme d'effets. A noter que les indicateurs de résultats choisis ne préjugent pas nécessairement de l'ensemble des effets de la politique. Par exemple, le décrochage a un impact sur la cohésion des familles, sur l'insertion des jeunes sur le marché du travail, sur la délinquance, etc. Mais il sera jugé opportun de focaliser l'évaluation sur les résultats dont on juge qu'ils doivent être atteints prioritairement.

Le troisième critère est celui de l'**efficience** du programme. Cette étude incluant la dimension de la performance (résultats par rapport aux coûts) ne devrait pas être aussi négligée. Toute intervention publique générant des coûts, il est naturel de précisément calculer ces derniers. La Déclaration des droits de l'homme et du citoyen de 1789 évoque déjà, dans ses articles 14 et 15, le droit des citoyens et de la société de contrôler les finances publiques. Le périmètre de calcul des coûts doit être en cohérence avec celui de la politique ou du programme mis en œuvre. Le calcul doit également s'appuyer sur un système d'information et de suivi des réalisations et de leurs coûts qui est adapté au contexte de l'évaluation. Cette exigence d'exactitude permet de rapporter les effets précédemment évoqués aux moyens engagés pour les obtenir. A ce titre, le critère de l'efficience établit la relation entre le coût d'une option et son impact (sa gamme d'effets). Comme pour l'efficacité, il existe plusieurs méthodes d'évaluation de l'efficience, qui seront examinées par la suite.

1.4 Les étapes de l'évaluation

L'activité évaluative est, selon la littérature, organisée comme n'importe quelle activité de prestation intellectuelle. Se succèdent des phases de diagnostic, de faisabilité, d'opportunité politique, de rédaction du cahier des charges, de mise en marché etc…

La première étape est celle du **diagnostic** des besoins de la politique si elle existe déjà au moins en partie. Il convient ici de rassembler les informations existantes, d'en collecter d'autres, en particulier de suivi de gestion, de faire le point des parties prenantes (« avec qui ? »), de la population cible (« pour qui ? »), etc.

La deuxième étape est celle de l'appréciation de **la faisabilité** de l'évaluation : celle-ci étant censée apporter une mesure de la valeur ajoutée de la politique pour la collectivité, il faut s'assurer qu'elle produira des résultats susceptibles d'être mesurés. Si la politique publique est énoncée de manière trop vaste ou trop vague, elle ne pourra être un objet d'évaluation.

Vient ensuite l'appréciation de **l'opportunité politique** de l'évaluation. C'est une étape délicate : l'évaluation peut répondre à une urgence, à un débat citoyen, à un problème lancinant et récurrent, à de nouvelles circonstances vis-à-vis desquelles il faut réagir, à l'anticipation de problèmes futurs, etc. Il faut rappeler que, quoiqu'il en soit, l'évaluation prendra du temps et mobilisera des compétences, des énergies et des budgets, qu'elle ne fournira jamais de réponses rapides et simplistes à des problèmes

complexes. Toutes les réponses ne seront pas définitives. Certaines devront être accompagnées d'un indice de confiance. La qualité du débat démocratique est aussi au prix de cette exigence de rigueur.

L'étape de la **rédaction du cahier des charges** de l'évaluation est cruciale pour le commanditaire puisqu'elle va fournir le cadre de travail du prestataire. Cette rédaction demande des moyens humains permanents et une expertise importante de la part des services du commanditaire qui sont en charge de cette mission (par exemple les services d'études et de recherche d'un ministère ou encore les services d'évaluation d'une assemblée). Outre l'exposé des besoins et donc de la question posée, le cahier des charges doit décrire précisément les objectifs de la politique à évaluer et les relier aux effets attendus. Cela dit, autant la description des objectifs doit être précise et circonscrite, autant une certaine latitude doit être laissée à l'évaluateur dans sa recherche et son appréciation des effets (attendus et inattendus) de la politique. De même, **le choix de la méthode** de mesure de l'efficacité puis de l'efficience (s'il est pour cette dernière décidé d'aller « au bout » du processus évaluatif) doit être une codécision du commanditaire et de l'évaluateur. **Le temps du débat** qui consiste à discuter du choix des méthodes (par exemple en utilisant un arbre de décision comme celui proposé plus loin) peut s'avérer très fructueux et éviter de s'engager dans des modes d'évaluation par exemple trop coûteux, trop longs à mettre en œuvre, susceptibles de fournir des indicateurs de résultats inadaptés ou trop incertains.

Le recensement des données et connaissances existantes sur le sujet est une aide précieuse. Le cahier des charges doit en faire le recensement mais aussi prévoir la récolte des données de suivi ou de *monitoring* (moyens engagés, données sur les bénéficiaires, indicateurs de résultats, etc.) qui seront mobilisées dans la démarche évaluative.

La mise en place d'un comité de suivi du marché de prestations intellectuelles est obligatoire. L'instance ou comité de suivi comprend *a minima* le commanditaire et les parties prenantes de la politique publique examinée.

Suivent alors **la mise en marché et le choix du prestataire de l'évaluation**. Une des difficultés de cette étape est de trouver un équilibre entre une commande publique précise (et donc d'autant plus intelligible pour le prestataire) et une suffisante ouverture du cahier des charges qui laisse des marges de manœuvre au prestataire pour adapter méthodes et mesures aux circonstances et au budget de l'activité d'évaluation.

Le respect des temps de l'évaluation doit être assuré. Selon les méthodes utilisées, l'horizon de la fin de l'évaluation peut être éloigné et il convient d'en tenir compte, de même que du temps de mise en œuvre des recommandations susceptibles d'en découler. Au plan calendaire, à noter que l'évaluation peut se faire en amont, avant la mise en place éventuelle de la politique (par exemple quant à l'opportunité d'un dépistage de masse en cancérologie), ou en aval, après sa mise en place (par exemple, les résultats d'un programme de rénovation urbaine). Cette mise en place peut s'accompagner d'un suivi « chemin faisant », des moyens effectivement engagés et des résultats d'étape (par exemple

l'évaluation des effets économiques et sociaux des ordonnances du 22 septembre 2017 relatives à la loi travail). La démarche doit prévoir un temps du bilan et un temps d'utilisation des résultats de l'évaluation. Le bilan est nécessaire à deux titres. D'une part, dès lors que les commanditaires ont décidé la mise en place de la stratégie jugée la plus efficace ou efficiente, il convient également d'en accompagner la mise en œuvre : trop de recommandations de rapports évaluatifs restent lettre morte. Il en résulte un gaspillage de fonds publics et une perte de chance (d'opportunité) pour les citoyens qui auraient pu en bénéficier. D'autre part, l'évaluation des politiques publiques reste un processus où toutes les parties prenantes « apprennent en faisant ». Il convient donc de conduire le changement en capitalisant cet apprentissage, en le diffusant dans toutes les sphères de l'action publique. Des domaines pionniers comme celui de la santé ont à cet égard une forme d'exemplarité qui peut s'avérer utile dans d'autres champs d'intervention.

L'étape de la pédagogie de l'évaluation et de ses résultats est cruciale également : sa qualité sera un gage de la pérennité de la démarche évaluative. Dans tous les cas, l'évaluation produit des informations et des connaissances qu'il faut rendre intelligibles tout en conservant leur précision, leur rigueur scientifique et éventuellement leur incertitude.

Une fois la mesure de l'efficacité ou de l'efficience obtenue se pose **la question de l'éventuel transfert des résultats** à l'ensemble des bénéficiaires potentiels, que ce soit dans le temps ou sur d'autres périmètres territoriaux. Une évaluation est très rarement menée à l'échelle de toute la population concernée. Même si les résultats sont obtenus sur des échantillons qui sont parfois très larges (par exemple, les bases de données en santé permettent désormais de travailler sur plusieurs dizaines de milliers de patients), il faut se pencher sur les conditions de leur généralisation avant la décision de mise en œuvre des recommandations qu'ils portent. Ceci vaut également pour des différences de contextes territoriaux (métropole, monde rural par exemple).

L'étape de la capitalisation des résultats est très utile à long terme. La capitalisation fait évoluer le choix des méthodes ainsi que les méthodes elles-mêmes. L'évaluation de l'évaluation des politiques publiques n'est pas un simple élément de langage, c'est aussi une façon intelligente d'assurer en pratique continue, ouverte et intelligente l'amélioration de l'action publique.

1.5 Le respect de principes éthiques et déontologiques

Puisque l'évaluation prépare une forme de jugement de l'action publique, la dimension éthique et des références-métiers dans la production des connaissances évaluatives doivent être présentes à chacune des étapes.

C'est le cas tout d'abord lors de l'identification des besoins, lesquels peuvent venir d'impératifs moraux ou politiques, par exemple pour corriger des situations individuelles jugées moralement inacceptables (situations d'insalubrité, de précarité, etc.).

La définition des objectifs des évaluations doit aussi être confrontée à **des principes éthiques souvent encadrés par la loi** (protection des données personnelles, droits des enfants, respect des libertés fondamentales, etc.) ou par un comité d'éthique qui peut être constitué pour la circonstance de la politique publique évaluée. De même, les effets à mesurer et les indicateurs de résultat qui y sont associés doivent être appréciés sous l'angle éthique. Les politiques publiques ont en effet souvent des objectifs de justice et de solidarité. Toutefois, il n'est pas dans la mission du prestataire de définir les principes éthiques ou de justice. Le prestataire doit travailler conformément au droit existant (par exemple en matière de droit de la protection des données personnelles). Il se peut que telle ou telle méthode d'évaluation soit susceptible de heurter un principe moral comme l'égalité des citoyens. Un exemple usuel est celui de l'expérimentation où un groupe dit « de traitement » reçoit une intervention (supposément bénéfique) tandis qu'il est comparé à un autre groupe dit « de contrôle » ou « groupe témoin » qui n'en bénéficie pas. L'éventuelle perte d'opportunité ou perte de chance que subit le groupe de contrôle (si tant est que l'intervention s'avère effectivement bénéfique) doit pouvoir être évaluée par le comité d'éthique de l'évaluation. Dans certains cas, cependant, le cahier des charges peut demander à l'évaluateur le respect d'un guide méthodologique de l'appréciation des dimensions éthiques de l'évaluation, comme par exemple la Haute Autorité de santé depuis 2013[9] .

Des principes déontologiques de connaissance minimum d'un domaine peuvent être requis. Professionnellement, certains pays ont pensé à un « ordre professionnel » des évaluateurs (Etats-Unis). Une déclaration de l'American Evaluation Association (AEA) sur la compétence en évaluation a ainsi été approuvée le 22 avril 2011. Sans aller jusqu'à l'idée d'un Ordre, la charte de la Société française de l´évaluation fournit des principes dont peuvent se réclamer les participants à l'évaluation des politiques publiques et qui peuvent être recommandés par les commanditaires.[10]

Plus spécifiquement, universitaires et chercheurs participant à l'activité évaluative continuent de respecter leur charte éthique (**charte nationale de déontologie des métiers de la recherche**).

Enfin, l'évaluateur doit naturellement respecter les principes déontologiques usuels dans la production de connaissances. La condition *sine qua non* est **l'absence de conflits d'intérêt**, déclaration qui devra être contrôlée par le commanditaire.

[9] Haute Autorité de Santé (2013) *L'évaluation des aspects éthiques à la HAS. Guide méthodologique.*
[10] http://www.sfe-asso.fr/evaluation/charte-evaluation

2 Des controverses sur les méthodes et un modèle d'aide à la décision

2.1 Le choix des méthodes et le degré de confiance dans les résultats de l'évaluation

Le choix des méthodes conditionne la qualité des résultats évaluatifs et permet à divers degrés de satisfaire les buts des évaluations : pertinence, cohérence et tout particulièrement efficacité et efficience.

Prenons comme exemple très simple l'image d'une politique publique qui vise à relier deux points du territoire (par exemple un bassin d'habitation et un bassin d'emploi) séparés par un obstacle (par exemple une rivière ou un axe de circulation dense telle une autoroute). La situation actuelle (le *statu quo*) consiste à emprunter un ouvrage d'art existant mais éloigné et encombré (par exemple un pont). Dans cet exemple simplifié, les options techniquement réalisables sont un autre pont ou un tunnel, plus proches. La politique publique peut donc être mise en œuvre par diverses stratégies : le *statu quo*, la construction d'un nouveau pont et le creusement d'un tunnel. On suppose pour simplifier que ces trois stratégies d'intervention sont mutuellement exclusives. Mettons l'accent sur l'efficacité et l'efficience.

La question posée de l'efficacité de la politique évaluée consiste à vérifier si la stratégie d'intervention mise en œuvre produit les effets attendus (dans notre exemple, un gain de temps ou de sécurité) ou d'autres effets (nuisances sonores, impact environnemental). En compilant la littérature, il apparaît que les méthodes d'évaluation de l'efficacité incluent l'évaluation qualitative, l'évaluation participative, l'enquête de satisfaction, l'expérimentation, la quasi-expérimentation, la monétisation de l'impact, la « méthode des effets ».

La question peut être aussi celle de la sélection des stratégies efficientes. Elle complète les résultats d'une évaluation de l'efficacité en introduisant la dimension « coût et budget » des politiques publiques. Elle est mobilisée pour comparer des stratégies d'intervention en concurrence. Par exemple, dans le cas le plus simple où deux stratégies d'intervention produisent des effets similaires (même efficacité), on privilégiera la stratégie d'intervention la moins coûteuse (de ce fait la plus efficiente). L'analyse devient naturellement plus complexe lorsque les stratégies produisent des effets différents. En compilant la littérature, il apparaît que les méthodes d'évaluation de l'efficience incluent l'analyse financière, l'analyse des coûts, l'analyse d'impact budgétaire, l'analyse coût-bénéfice, l'analyse coût-efficacité, l'analyse multicritère.

Notons que les deux types de questions évaluatives évoqués ci-dessus sont complémentaires. Premièrement, évaluer une politique publique sur la base de ses seuls effets sera insuffisant. Se pose en effet la question de l'opportunité de l'utilisation des ressources publiques rares (cf. la notion de coût d'opportunité évoquée au début de ce rapport) et donc de l'efficience des politiques publiques. Savoir si telle ou telle politique a atteint ses objectifs (efficacité) est bien sûr primordial, mais ce qui compte

au final est de savoir si, l'ayant sélectionnée, l'on aurait pu atteindre ces mêmes objectifs avec des stratégies d'intervention moins coûteuses (efficience). Deuxièmement, choisir une stratégie d'intervention sur la base de son efficience ne sera évidemment possible que si l'ensemble des effets a été correctement identifié. A titre illustratif, nous pouvons citer l'analyse coût-efficacité, méthode de sélection fréquemment utilisée en évaluation médico-économique pour comparer des stratégies d'intervention substituables. Cette méthode se fonde sur les résultats d'une ou plusieurs expérimentations ou quasi-expérimentations préalables (évaluation d'impact) pour calculer des ratios différentiels coût-résultat (par exemple le coût supplémentaire par année de vie gagnée en passant d'une stratégie à une autre). De manière similaire, l'analyse coût-bénéfice, méthode fréquemment employée en matière d'investissements en infrastructures publiques, se fonde sur une monétisation préalable des effets (choix d'une valeur monétaire pour le temps, la vie humaine, etc.) pour calculer des indicateurs de performance.

Dans beaucoup de cas, quel que soit le choix des méthodes et des outils, l'activité évaluative peut conduire à des résultats fragiles. Pour cette raison, même si l'évaluation doit être un élément essentiel dans le choix des modalités de gouvernance, elle ne doit rester qu'un outil d'aide à la décision, et être ouverte à la controverse.

Prenons l'exemple des difficultés inhérentes aux méthodes d'évaluation de l'efficacité. La principale difficulté est d'isoler les effets générés par (imputables ou attribuables à…) l'intervention publique de ceux générés par d'autres causes possibles. Considérons le cas d'un programme de tutorat destinés aux élèves en échec scolaire. Pour mesurer l'impact de ce programme, nous pouvons essayer de comparer les notes moyennes des bénéficiaires du programme avant et après l'intervention et voir si les élèves ont amélioré ou non leurs résultats. Toutefois, cette approche ne sera robuste que si la méthode de notation reste inchangée sur la période analysée (enseignants identiques, examen vraiment similaire, etc.). De même, comparer les notes des bénéficiaires du programme (initialement en échec scolaire) avec celles des non-bénéficiaires (initialement sans difficulté scolaire) est susceptible d'introduire un autre type de biais puisque ces deux types d'élèves ne sont *a priori* pas nécessairement comparables au regard d'un grand nombre de dimensions (motivation, âge, genre, situation familiale, etc.). Dans le premier cas, c'est l'environnement changeant du programme qui peut rendre difficile l'exercice de l'évaluation, dans le deuxième cas, c'est le manque de comparabilité des individus (ou « biais de sélection ») qui affecte les résultats de l'évaluation. Pour réduire tout biais éventuel, il faudrait idéalement comparer la situation issue de la politique publique à la situation qui serait survenue si l'intervention n'avait pas eu lieu, situation hypothétique appelée « contrefactuel ».

Pour ce faire, **les méthodes expérimentales et quasi-expérimentales** approximent ce qui se serait passé en l'absence de la politique publique, notamment en cherchant un groupe de comparaison adéquate : on parle alors d'évaluation contrefactuelle. **L'expérimentation**, quand elle est possible, et quand elle ne conduit pas à trop formater l'intervention (pour en faire une expérience sous contrôle parfait mais s'éloignant de la vie réelle du programme), est une méthode d'évaluation d'impact

généralement considérée comme robuste. Inspirée des essais cliniques de la recherche médicale, l'approche consiste à répartir au hasard un échantillon d'individus entre un groupe de traitement (bénéficiaire de l'intervention) et de contrôle (non-bénéficiaire). La répartition au hasard est le cœur de la méthode puisqu'elle permet d'avoir des groupes comparables. L'effet de l'intervention se mesure alors directement à partir des différences observées entre les deux groupes pour une période donnée. Dans certains cas, d'autres biais peuvent toutefois apparaître, du fait même de la répartition au hasard, rendant difficile l'estimation de l'efficacité de l'intervention. Par exemple, les membres du groupe de contrôle, s'ils sont informés qu'ils sont exclus du programme, peuvent connaître une démotivation accrue. Les membres du groupe de traitement peuvent quant à eux très bien décider de ne pas suivre avec assiduité le programme évalué ou même l'abandonner.

L'évaluation peut se fonder sur une approche quasi-expérimentale lorsque cette répartition au hasard n'est pas possible pour les raisons évoquées ci-dessus, mais aussi pour des raisons éthiques, logistiques ou réglementaires. L'évaluation quasi-expérimentale évalue un programme en situation de vie réelle, directement lors de sa mise en œuvre sur la population cible. L'approche compare les bénéficiaires de l'intervention publique avec les non-bénéficiaires et, si des biais sont présents, utilise des méthodes spécifiques permettant de réduire les différences entre les groupes et d'isoler l'effet de l'intervention. Les méthodes quasi-expérimentales incluent (a) la méthode de la différence de différences (ou double différence) ; (b) la régression en discontinuité ; (c) l'appariement par score de propension ; (d) la méthode des variables instrumentales.

a) La plus simple des méthodes quasi-expérimentales est celle de **la différence de différences.** L'idée est de tenir compte dans un premier temps de la différence de résultats observée entre le groupe des bénéficiaires et le groupe des non-bénéficiaires avant intervention et ce, afin de quantifier le biais de sélection initial. Dans un deuxième temps, il s'agit d'examiner la différence observée entre les deux groupes après intervention, et de soustraire le biais de sélection préalablement calculé. Dans sa version la plus simple, cette méthode ne nécessite donc que quatre valeurs moyennes. Par exemple, les élèves bénéficiaires d'un programme de lutte contre l'échec scolaire ont une moyenne de groupe de 7/20 avant intervention, tandis que les élèves non-bénéficiaires ont une moyenne de groupe de 13/20. Une différence préalable de –6/20 existe donc entre ces deux groupes : c'est le biais de sélection initial. Après intervention, si les élèves bénéficiaires du programme ont une moyenne de 10/20, et les non-bénéficiaires de 14/20, l'effet estimé du programme de tutorat sera alors de (10–14)–(–6)=2 : les élèves ayant bénéficié du programme de tutorat améliorent leur moyenne de 2 points en moyenne, toutes choses égales par ailleurs.
b) **La méthode de régression en discontinuité** n'est quant à elle applicable que dans les situations où il existe un seuil d'éligibilité préalablement choisi et permettant de définir qui sera bénéficiaire de l'intervention. Par exemple, dans le cadre d'un programme de tutorat, ceux qui n'ont pas obtenu la moyenne de 10/20 seront éligibles à l'intervention. Si un tel seuil existe,

l'approche consiste ensuite à comparer les bénéficiaires et les non-bénéficiaires qui sont proches de ce seuil. En d'autres termes, les observations trop éloignées sont exclues de l'analyse. L'idée est qu'en examinant les observations situées à proximité du seuil, on devrait éliminer les biais de sélection. Pour reprendre l'exemple précédent, il s'agira de comparer les élèves ayant 8 ou 9 de moyenne ayant bénéficié du programme de tutorat, avec ceux ayant 10 ou 11 n'ayant pas bénéficié du programme afin d'éviter la comparaison d'élèves trop dissemblables. Pour mettre en œuvre cette méthode, il faut également s'assurer que la démarche évaluative n'influence pas elle-même les comportements (par exemple, ayant pris connaissance du seuil de 10/20, certains élèves peuvent essayer de travailler plus pour éviter d'avoir à participer au programme de tutorat).

c) **La méthode des scores de propension** consiste à apparier des bénéficiaires et des non-bénéficiaires qui ont des caractéristiques similaires afin de créer un groupe de traitement et un groupe de contrôle directement comparables. L'impact de l'intervention est alors calculé comme la différence de résultats observée après intervention entre les deux groupes sélectionnés. Pour sa mise en œuvre, il n'est pas nécessaire d'avoir des données sur l'indicateur de résultat avant intervention : cela peut se révéler être un atout si aucun système d'information n'existe au préalable. Toutefois, l'approche nécessite un ensemble de données conséquent sur les caractéristiques des bénéficiaires et des non-bénéficiaires, par exemple leur âge, leur genre, leur cursus scolaire, leur situation familiale, leur revenu si l'on évalue une politique d'emploi.

d) Enfin, **la méthode des variables instrumentales** est une approche statistique permettant de corriger les « phénomènes d'auto-sélection » qui adviennent lorsque certains bénéficiaires choisissent d'eux-mêmes de participer (ou non) au programme. Comprendre les raisons de cette participation est primordial puisque celle-ci peut être déterminée par des facteurs affectant également l'indicateur de résultat (on parle de « facteurs qui sèment la confusion » ou de « facteurs confondants »). Par exemple, en reprenant le cadre précédent, il est possible que seuls les élèves les plus motivés aient suivi le programme de tutorat, ce qui pourrait induire une surestimation de l'effet du programme : la différence de résultat entre le groupe des bénéficiaires et celui des non-bénéficiaires inclut (1) l'effet du programme mais aussi (2) l'impact de la motivation des élèves, facteur difficilement mesurable. En d'autres termes, l'évolution de la situation du bénéficiaire dépend certes de la politique mise en œuvre, mais également d'autres influences qui, si elles sont omises, brouillent celles qui sont spécifiques à la politique évaluée. Examiner l'appartenance des élèves à tel ou tel groupe n'est donc plus suffisant pour juger de l'effet de la politique. Pour résoudre ce problème, la méthode des variables instrumentales va dans un premier temps chercher d'autres variables (ou « instruments ») qui sont cette fois aisément mesurables et non liés aux facteurs qui sèment la confusion. Ces instruments vont servir à estimer et modéliser la probabilité de participer au programme. Par exemple, l'éloignement des élèves du lieu du tutorat (en km) peut s'avérer être un « instrument » pertinent : l'éloignement n'est lié en aucun cas à la motivation des élèves et peut expliquer en partie l'absentéisme de certains. Dans un deuxième temps, la

comparaison se fait sur la base des probabilités ainsi estimées (qui ne dépendent que de facteurs mesurables et « non-confondants ») et non plus sur l'appartenance des élèves à tel ou tel groupe (qui dépendait de facteurs non-mesurables et « confondants »). Pour conclure, la méthode dite des variables instrumentales part de l'idée que l'évaluation doit aussi tenir compte de certains facteurs inobservables et, en conséquence, identifier l'influence d'autres canaux (les instruments), permettant d'isoler l'effet de la politique.

Pour conclure, la mise en œuvre de telle ou telle méthode dépend souvent du contexte de la politique publique. Toutefois, un point commun est que la récolte de données sur les non-bénéficiaires est tout aussi importante que celles sur les non-bénéficiaires. En effet, les données sur les non-bénéficiaires utilisées par les approches expérimentales et quasi-expérimentales permettent d'approximer le contrefactuel et de mesurer l'effet de l'intervention.

2.2 Des données et des métriques

Comme indiqué plus haut, une réflexion sur la récolte des données doit se faire en amont, afin d'intégrer des informations pertinentes sur une population ou un échantillon de non-bénéficiaires dont les caractéristiques sont plus ou moins similaires à celles des bénéficiaires du programme évalué.

Plus généralement, la collecte des données constitue une activité indissociable de la démarche évaluative. En ce sens, l'accès à des données pertinentes est souvent considéré comme une difficulté par les évaluateurs. La disponibilité de données de suivi est donc une condition *sine qua non* à l'EPPP[11].

La question de la métrique (l'unité de mesure) utilisée constitue également un enjeu majeur. Les deux méthodes que sont l'expérimentation ou la quasi-expérimentation expriment les effets en unité « physique », c'est-à-dire dans leur unité d'origine (par exemple, un taux d'insertion pour une politique d'emploi, le nombre de vies sauvées pour un programme de santé). D'autres méthodes sont disponibles, permettant d'exprimer les effets soit en valeur monétaire (méthodes dites de monétisation avec, par exemple, un prix de référence de la vie humaine ou du temps perdu), soit de manière qualitative (enquête de satisfaction, étude qualitative, évaluation participative), soit plus spécifiquement par leurs effets sur la croissance économique ou l'emploi (méthode des effets).

Enfin, une autre difficulté est celle de la multiplicité des effets. Rares sont les politiques publiques ayant un effet unique. Les effets peuvent être ceux visés par la politique (par exemple, une meilleure régulation du trafic suite à une construction de voirie). La politique peut également engendrer d'autres

[11] Voir le séminaire commun SFE-Cour des comptes, octobre 2015, « Des données pour évaluer » : http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-evaluant-ses-politiques-publiques/des-donnees-pour-evaluer

effets sur la collectivité (pollution sonore, croissance économique). Ces effets peuvent être négatifs ou positifs, ou avoir un horizon temporel différent (à court, moyen ou long terme).

Comment traiter ces questions de métrique et de multiplicité des effets ? La première solution, proposée notamment dans la méthode d'analyse coût-bénéfice et la méthode multicritère, est de convertir les effets dans une métrique qui les rende comparables, puis de synthétiser l'information à travers un indicateur unique. Pour l'analyse coût-bénéfice, cette métrique est monétaire. Il s'agit d'attribuer une valeur en euro à chacun des effets, qu'ils soient économiques, sociaux ou environnementaux. Les effets négatifs auront une valeur monétaire négative, tandis que les effets positifs auront une valeur positive. Il s'agit ensuite d'additionner l'ensemble des effets pour calculer un indicateur de performance unique. De manière similaire, l'analyse multicritère compare les stratégies en rapportant chacun des effets sur une échelle allant par exemple de 0 à 100. Dans sa version la plus simple, la méthode calcule ensuite un indicateur composite sur la base de cette métrique. La seconde solution, que l'on retrouve en analyse coût-efficacité, consiste à centrer l'analyse sur ce qui est jugé comme étant l'impact principal. Dans le calcul médico-économique par exemple, les effets d'une politique de santé sont synthétisés par l'année de vie gagnée pondérée par sa qualité.

2.3 Des incertitudes inévitables

Apprécier le degré d'incertitude lié aux résultats obtenus est un des enjeux majeurs de l'évaluation des politiques publiques. Deux types d'incertitude sont à considérer :

- **Des incertitudes sur les paramètres.** Ici l'incertitude est quantifiable et liée à l'estimation plus ou moins précise des effets et des coûts (par exemple, l'échantillon examiné est-il représentatif de la population cible ? Quid de la qualité des données ?).
- **Des incertitudes structurelles.** Ici l'incertitude est non-quantifiable, difficile à cerner, et pourtant fondamentale. Par exemple : l'ensemble des effets est-il bien pris en compte ? Toutes les stratégies pertinentes sont-elles incluses dans l'analyse ? Les effets passent-ils par un réseau de chaînes de valeurs ajoutées et non une seule chaîne ? Les résultats dépendent-ils de l'horizon temporel analysé ?

L'approche usuelle pour tenir compte de l'incertitude sur les paramètres est le calcul d'intervalles de confiance. Cela revient à calculer une marge d'erreur autour des effets et coûts estimés. Par exemple, le coût d'un équipement productif estimé à 100 000 euros sera agrémenté d'un écart plus ou moins grand, par exemple ± 20 000 euros, selon le degré d'incertitude. Une fois ces écarts calculés pour l'ensemble des coûts et des effets, il est possible de les analyser simultanément afin d'obtenir une mesure globale de l'incertitude paramétrique, méthodologie connue sous le nom d'analyse de sensibilité probabiliste.

L'incertitude structurelle est quant à elle difficilement quantifiable. Nous ne parlerons pas ici des biais volontairement construits par les groupes d'intérêt dans le choix des questions évaluatives. Dans certains cas, le manque de données disponibles ou de connaissances factuelles force l'évaluateur à émettre des jugements de valeur sur la mesure de certaines dimensions de la politique (par exemple les effets à long terme du projet). La sélection de la stratégie d'intervention la plus efficiente peut évidemment fortement varier au gré des hypothèses émises. Il est donc essentiel de vérifier la robustesse de cette sélection. Pratiquement, puisque ces éléments d'incertitude sont difficiles à mesurer, l'approche consiste à analyser un ensemble de scénarios possibles, par exemple un scénario pessimiste opposé à un scénario optimiste.

La discussion autour de l'incertitude structurelle n'a pas qu'une dimension technique, elle permet notamment de s'interroger sur la pertinence de la population cible, de la gamme d'effets que l'on souhaite mettre à jour, etc. En outre, comme cela a déjà été évoqué, il convient d'avoir toujours à l'esprit que la stratégie la plus efficiente est peut-être celle qui n'a pas été incluse dans l'analyse.

2.4 Pour y voir clair : un arbre de décision et des tableaux comparatifs

Ici sont proposés un arbre de décision et des tableaux comparatifs innovants et originaux. Ils doivent permettre d'encadrer le débat et les controverses sur les performances absolues ou relatives de telle ou telle méthode dans un univers des acteurs des EPPP qui est riche de controverses en France.

L'arbre de décision (figure 2) vise à servir de table d'orientation entre les différentes méthodes susceptibles d'être mobilisées par les acteurs de l'évaluation. La représentation n'est pas exhaustive. Ont été retenus les outils majeurs de l'évaluation[12]. A des fins d'utilisation commode dans un raisonnement du type « *si le commanditaire souhaite telle ou telle qualité de résultats, alors il doit se tourner plutôt vers telle ou telle méthode* », cet arbre est complété par quatre tableaux croisés (tableaux 1 à 4).

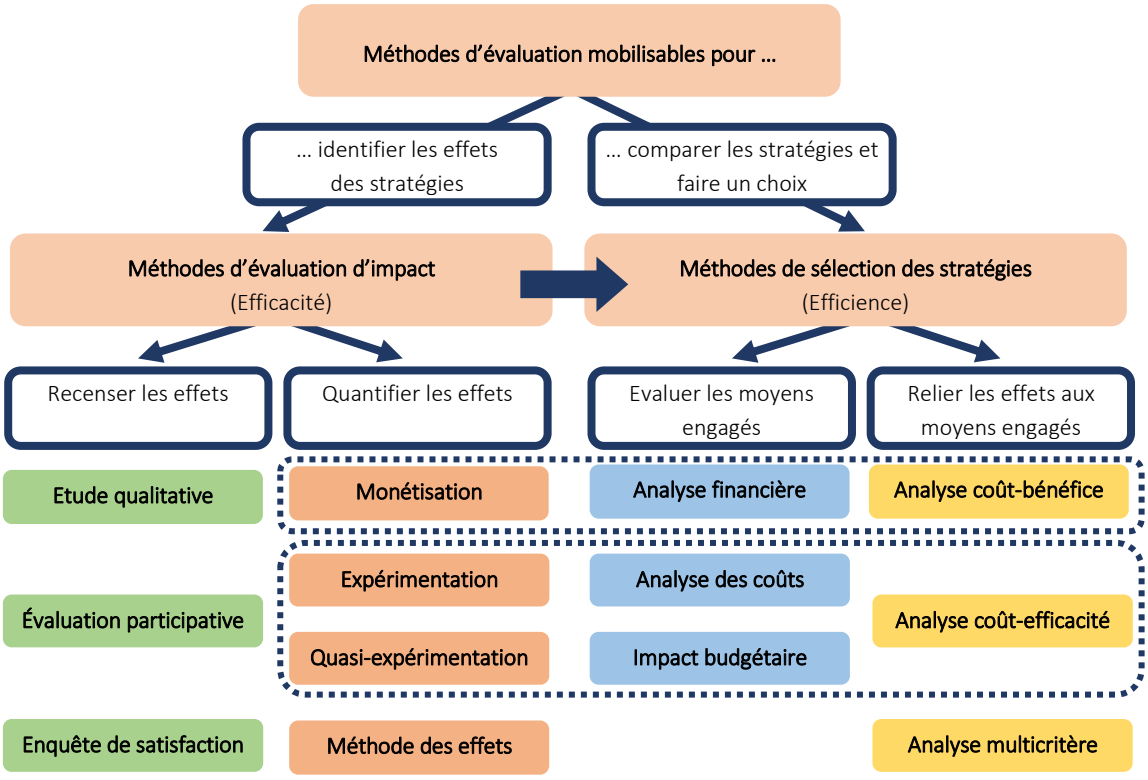
Les tableaux croisés permettent de comparer les méthodes d'évaluation de façon plus détaillée au vu des quatre critères retenus dans ce rapport. Ce sont des critères issus du travail scientifique sur la qualité des résultats des études évaluatives[13] : la complétude, la simplicité, la transparence, la transférabilité (voir encadré 2). Une discussion sur l'ajout d'autres critères, en particulier sur la

[12] Pour une liste plus longue, consulter le site « Better Evaluation » présenté dans l'annexe 1.
[13] Voir à la fois les critères du choix rationnel (complétude, simplicité) dans Mathieu F (2015), *Annuaire de philosophie des sciences* (https://www.thebookedition.com/fr/annuaire-de-philosophie-des-sciences-p-123501.html) et une ressource en évaluation qualitative Gohier C (2004) De la démarcation entre critères d'activité scientifique et d'ordre éthique en recherche interprétative, *Recherches qualitatives*, 24 : 3-17 (http://www.recherche-qualitative.qc.ca/documents/files/revue/edition_reguliere/numero24/24gohier.pdf).

contribution de l'EPPP à un meilleur fonctionnement de la démocratie, peut évidemment être menée (voir annexe 2).

La figure 2 permet de comprendre le cheminement vers l'une ou l'autre des méthodes d'évaluation. L'arbre de décision se lit en colonne et en ligne. En colonne, il montre en premier lieu le choix entre une ou plusieurs méthodes d'évaluation d'impact afin de recenser et de quantifier les effets des stratégies, et en vue d'évaluer l'efficacité. Ces méthodes permettent d'identifier les effets de manière qualitative (expression des effets par les acteurs et parties prenantes) ou de manière quantitative (mesures, monétaires ou non, de ces effets). En second lieu, en colonne toujours, il montre les choix à opérer en vue d'évaluer les moyens engagés (approche financière, par la comptabilité analytique ou par l'impact sur le budget dans lequel s'insère la politique publique). Enfin, en troisième lieu, l'arbre souligne le choix entre plusieurs méthodes permettant de relier les moyens aux résultats et ce, en vue d'évaluer l'efficience.

Figure 2. Arbre de décision des méthodes d'évaluation



Encadré 2. Critères de comparaison des méthodes

Complétude : la méthode permet une vision à la fois précise (capacité à mesurer) et exhaustive (capacité à rendre compte des diverses dimensions) de ce que l'on cherche à estimer (point de vue des parties prenantes, ensemble des moyens engagés ou des effets, etc.) ;

Simplicité : la méthode est aisée à mettre en œuvre (des points de vue techniques, éthiques, logistiques, budgétaires, etc.) ;

Transparence : les informations sur ce qui a été fait sont accessibles pour les différentes étapes du processus d'évaluation (hypothèses émises, protocole choisi, données de suivi et statistiques employées), les résultats sont donc reproductibles et vérifiables, difficilement manipulables ;

Transférabilité : les résultats sont généralisables à la population cible ou à d'autres populations susceptibles de recevoir l'intervention.

En ligne, l'arbre de décision décrit tout d'abord les choix entre l'analyse coût-bénéfice et l'analyse coût-efficacité. Par exemple, la monétisation des effets et l'analyse financière des politiques publiques et programmes constituent une étape préalable à l'analyse coût-bénéfice. L'analyse coût-efficacité utilisera des mesures de résultats, d'effets ou d'impact tirées de l'expérimentation ou de la quasi-expérimentation et effectuera une analyse préalable des coûts engendrés. L'analyse de ces coûts peut être complétée par une analyse d'impact budgétaire qui examine dans quelle mesure la mise en place d'une nouvelle stratégie d'intervention affecte le budget global d'un programme public. Au contraire, l'analyse multicritère peut s'appuyer sur tout ou partie des informations fournies par les autres méthodes. Elle compare des profils multidimensionnels de résultats et de coûts.

Les quatre tableaux comparatifs ci-dessous permettent de voir plus clair dans les choix d'évaluation à mettre en œuvre. Le tableau 1 compare les méthodes d'évaluation d'impact au plan du recensement des effets. Le tableau 2 compare les méthodes d'évaluation d'impact au plan de la quantification de ces effets. Le tableau 3 compare les méthodes de sélection des stratégies (analyse financière, analyse des coûts, analyse d'impact budgétaire) sur le plan des moyens engagés. Le tableau 4 compare les méthodes de sélection des stratégies sur le plan de l'efficience, c'est-à-dire de la liaison entre effets et moyens (analyse coût-bénéfice, analyse coût-efficacité, analyse multicritère).

Pour conclure, l'arbre de décision et les tableaux comparatifs doivent permettre au lecteur commanditaire d'évaluation d'établir, en fonction de ses objectifs évaluatifs, que telle ou telle

méthode est préférable ou même s'impose aux autres méthodes. Considérons par exemple un commanditaire souhaitant l'évaluation d'un programme spécifique et ayant pour objectif premier de recenser l'ensemble des effets : si la contrainte de temps s'avère forte, l'évaluation participative et l'enquête de satisfaction pourront être écartées au profit de l'étude qualitative (critère de simplicité). L'évaluation participative sera au contraire privilégiée si le commanditaire souhaite prendre le temps d'un dialogue approfondi entre les différentes parties prenantes (critère de transférabilité). Enfin, si le commanditaire souhaite avoir une vision plus précise des effets (critère de complétude), ou mettre en place une approche plus transparente (critère de transparence), il pourra se tourner alors vers l'enquête de satisfaction.

Tableau 1. Méthodes d'évaluation d'impact : recenser les effets

Méthode	Complétude	Simplicité	Transparence	Transférabilité
Etude qualitative	1☑ 2☐ 3☐	1☐ 2☐ 3☑	1☐ 2☑ 3☐	1☑ 2☐ 3☐
Réalisée avec des entretiens semi-directifs menés sous la forme de questionnaires. Les principaux acteurs et bénéficiaires du programme sont interrogés sur différentes dimensions de l'intervention. L'approche recense les différents points de vue et peut être complétée par une recherche documentaire ou une observation sur le terrain.	L'étude identifie les différents effets de l'intervention, sous réserve de l'exhaustivité du recensement des points de vue. Les effets sont listés mais ne sont pas quantifiés.	La récolte des données est simple mais longue et se fonde en conséquence sur un échantillon de taille relativement modeste (par exemple, moins de 50 entretiens). Les entretiens peuvent être enregistrés, puis retranscrits.	Les entretiens semi-directifs peuvent orienter le discours des personnes interrogées dans certaines directions et introduire des biais (par exemple un biais de complaisance vis-à-vis de l'enquêteur ou vis-à-vis des services évalués).	De par sa faible taille, l'échantillon peut ne pas être représentatif.
Évaluation participative	1☑ 2☐ 3☐	1☐ 2☑ 3☐	1☑ 2☐ 3☐	1☐ 2☐ 3☑
Réalisée avec la participation active des principales parties prenantes à l'exercice d'évaluation (réalisation d'entretiens individuels et collectifs, débats publics, etc.). L'approche favorise l'expression de la diversité des points de vue et l'adhésion des acteurs aux conclusions de l'évaluation.	Comme dans le cas de l'étude qualitative, on identifie les différents effets de l'intervention, sous réserve de l'exhaustivité du recensement des points de vue. Les effets sont listés mais ne sont pas quantifiés.	Le manque d'expériences de certains acteurs peut rendre la concertation difficile : possibilité de zones de conflits en amont de la décision.	L'évaluateur, lui-même « embarqué » ou partie prenante du processus, encadre et oriente le processus évaluatif, et joue un rôle de médiateur.	Le dialogue vise à établir un échange le plus large possible entre les groupes d'intérêts en présence.
Enquête de satisfaction	1☐ 2☑ 3☐	1☐ 2☑ 3☐	1☐ 2☑ 3☐	1☐ 2☐ 3☑
Réalisée par sondage sur la satisfaction des usagers d'une intervention publique donnée. Ceux-ci répondent à un questionnaire où ils sont amenés à se prononcer sur les conditions matérielles du service (accès à la prise en charge, horaires, délais, etc.) mais aussi sur ses aspects qualitatifs (accueil, qualité de l'écoute, etc.).	La satisfaction des usagers est appréciée sur une échelle de valeur (par exemple de 1 à 10). L'enquêté choisit un niveau correspondant à la qualité qu'il perçoit. Cette forme de mesure peut cependant masquer l'efficacité réelle de la politique.	L'approche se fonde sur des méthodes d'enquête (rédaction d'un questionnaire) et la constitution d'un échantillon suffisamment large (400 à 1 000 usagers).	Les protocoles d'enquête peuvent parfois introduire des biais (par exemple un biais de complaisance vis-à-vis de l'enquêteur ou vis-à-vis des services évalués).	De par sa taille, l'échantillon est représentatif.

Tableau 2. Méthodes d'évaluation d'impact : quantifier les effets

Méthode	Complétude	Simplicité	Transparence	Transférabilité
Monétisation	1☐ 2☑ 3☐	1☑ 2☐ 3☐	1☑ 2☐ 3☐	1☐ 2☑ 3☐
Estimation à partir d'un échantillon de la disposition à payer des usagers (ce qu'ils sont prêts à payer au maximum) pour une intervention donnée. On utilise des données d'enquêtes (questionnaires) ou des données d'observation[1].	Mesure en accord avec la théorie économique du consommateur[2] mais qui suppose que les usagers sont en mesure de percevoir et d'appréhender l'ensemble des effets attendus.	La récolte des données est complexe (procédures d'enquête spécifiques, ou données d'observation géo-localisées).	Les méthodes statistiques requièrent un haut niveau d'expertise et sont sujettes à débat (résultats d'estimation variables selon la méthode).	Permet de comparer des effets *a priori* non comparables et de généraliser les résultats. La monétisation de certains effets peut cependant faire débat (la valeur finie de la vie humaine par exemple).
Expérimentation	1☐ 2☐ 3☑	1☐ 2☑ 3☐	1☐ 2☐ 3☑	1☐ 2☑ 3☐
Répartition au hasard d'individus[3] entre un groupe de traitement (bénéficiaire de l'intervention) et de contrôle (non-bénéficiaire). L'effet de l'intervention se mesure directement à partir des différences observées entre les deux groupes.	Lorsque le protocole expérimental ne modifie pas les comportements (démotivation, assiduité), la répartition au hasard permet d'avoir des groupes de contrôle et de traitement directement comparables.	La méthode est bien balisée mais peut être longue ou onéreuse à mettre en œuvre, ou encore faire face à des contraintes d'ordre éthique, logistique, politique et réglementaire.	Les résultats sont aisément vérifiables. Méthode de référence en santé.	L'échantillon analysé peut avoir des caractéristiques en partie différentes de celles de la population cible.
Quasi-expérimentation	1☐ 2☑ 3☐	1☐ 2☐ 3☑	1☐ 2☑ 3☐	1☐ 2☐ 3☑
Comparaison d'un groupe d'individus[3] bénéficiaires de l'intervention avec un groupe de non-bénéficiaires. Ces deux groupes ne sont pas nécessairement comparables. L'approche peut utiliser diverses méthodes[4] permettant de réduire les différences entre les groupes et d'isoler l'effet de l'intervention.	L'approche compare des individus[3] n'ayant pas fait l'objet d'une répartition au hasard. De ce fait, les données comportent des biais qu'il est nécessaire de corriger.	L'approche peut se fonder sur un système d'information préexistant, mais l'extraction des données peut s'avérer complexe.	Nécessite l'usage de méthodes statistiques sophistiquées afin d'isoler l'effet de l'intervention par rapport aux autres causes possibles.	Permet d'évaluer un programme en situation de vie réelle, après sa mise en œuvre sur la population cible.
Méthode des effets	1☐ 2☑ 3☐	1☑ 2☐ 3☐	1☐ 2☑ 3☐	1☑ 2☐ 3☐
L'approche vise à estimer pour une zone géographique donnée la valeur ajoutée du projet et la manière dont elle se partage entre les différents agents économiques présents sur le territoire. Elle évalue donc les coûts et avantages directs et indirects de l'intervention sur l'ensemble des secteurs économiques locaux.	Repose sur le concept keynésien de *multiplicateur* selon lequel des dépenses initiales (fiancées par subventions extérieures) stimulent, sur une base territoriale donnée, l'activité économique dans ce territoire. L'approche ne permet pas de considérer des effets autres que pécuniaires.	Afin d'estimer les différents effets, l'approche nécessite un système comptable élaboré retraçant les flux monétaires entre un certain nombre de secteurs économiques.	La méthode est sophistiquée. Le choix du territoire analysé peut aussi influencer les résultats selon que la zone bénéficie ou non d'interdépendances sectorielles.	L'approche n'a de sens que pour une zone géographique donnée, afin de rendre comparables les effets multiplicateurs des différentes stratégies.

1 Par exemple l'impact de la politique sur le prix du foncier. Plus généralement, les méthodes de monétisation des effets incluent l'évaluation contingente, la méthode des choix discrets, la méthode des coûts de transport et la méthode des prix hédonistes.
2 Elle s'appuie sur la « disposition à payer », mesure de la satisfaction que les usagers peuvent retirer du programme évalué.
3 « Individu » est un terme générique désignant les bénéficiaires potentiels du programme public évalué. Il peut s'agir de personnes (usagers, patients, élèves, contribuables, etc.) ou d'entités spécifiques (entreprises, ménages, associations, collectivités, etc.).
4 Les méthodes quasi-expérimentales incluent la méthode de la différence de différences, l'appariement par score de propension, la régression en discontinuité, et la méthode des variables instrumentales (voir section 2.1).

Tableau 3. Méthodes de sélection des stratégies : évaluer les moyens

Méthode	Complétude	Simplicité	Transparence	Transférabilité
Analyse financière	1☐ 2☐ 3☑	1☐ 2☑ 3☐	1☐ 2☑ 3☐	1☐ 2☑ 3☐
Similaire à celle conduite lors du choix d'investissements privés. Calcule sur l'horizon temporel de la politique les coûts d'investissement, les coûts et revenus de fonctionnement ; analyse également les sources de financement.	Permet une vision précise de la soutenabilité financière de la stratégie, de la profitabilité financière de l'investissement, ainsi que de la profitabilité du capital investi. L'analyse couvre l'ensemble des flux financiers sur un horizon temporel large.	L'approche financière est bien balisée mais les projets complexes nécessitent une analyse assez sophistiquée.	L'interprétation des différents taux et ratios doit être menée avec précaution. Les choix méthodologiques (par exemple d'amortissement) doivent être explicites.	L'évaluateur peut être amené à émettre des jugements de valeur sur la mesure de certains flux financiers. L'examen de plusieurs scénarios peut être nécessaire.
Analyse des coûts	1☐ 2☑ 3☐	1☐ 2☑ 3☐	1☐ 2☐ 3☑	1☑ 2☐ 3☐
Détermine les coûts unitaires (par bénéficiaire) associés à chaque stratégie. Ils comprennent les coûts[1] directement associés à sa mise en œuvre et ceux liés aux évènements non intentionnels qu'elle génère.	L'identification de l'ensemble des coûts n'est pas toujours aisée, en particulier lorsqu'ils relèvent d'évènements non intentionnels.	Prévoir une comptabilité analytique de la stratégie facilite l'élaboration des coûts. Si un *micro-costing*[2] s'avère nécessaire, il peut être long et coûteux.	La démarche comptable qui préside à l'analyse des coûts est transparente et vérifiable.	L'analyse est très liée au périmètre de coût choisi. Le *micro-costing* s'appuie souvent sur des échantillons de taille réduite ce qui peut compromettre sa représentativité.
Analyse d'impact budgétaire	1☐ 2☑ 3☐	1☐ 2☑ 3☐	1☐ 2☐ 3☑	1☐ 2☐ 3☑
Examine dans quelle mesure la mise en place d'une nouvelle stratégie d'intervention affecte le budget global d'un programme préexistant. L'approche est adoptée par de nombreuses agences nationales d'évaluation des technologies de santé.	Complémentaire d'une analyse coût-efficacité, l'approche se focalise sur le budget pour un horizon de temps plutôt court.	Nécessite des données de prévision sur les dépenses, le nombre de bénéficiaires et leur répartition entre les différents services qui composent le programme.	Les différentes étapes de calcul, les hypothèses utilisées et les données mobilisées, ainsi que leurs sources, doivent être détaillées.	Se fonde sur une estimation de la part de la population susceptible de bénéficier de l'intervention.

1 Par exemple, dans une analyse médico-économique, le coût unitaire comprend le coût du traitement d'un patient « moyen » mais aussi le coût associé aux conséquences non intentionnelles que sont les effets indésirables dudit traitement.

2 Le *micro-costing* consiste à calculer le coût complet d'une action importante au sein de la stratégie. Par exemple, le suivi d'un demandeur d'emploi pourra être analysé au plus près, avec le calcul du temps passé avec son ou ses conseillers, le temps de traitement de son dossier par ces derniers, sur un horizon pertinent (par exemple six mois). Ces temps seront valorisés par le coût horaire du travail des conseillers. Le *micro-costing* est régulièrement utilisé dans l'évaluation des programmes de santé.

Tableau 4. Méthodes de sélection des stratégies : relier effets et moyens

Méthode	Complétude	Simplicité	Transparence	Transférabilité
Analyse coût-bénéfice	1☐ 2☐ 3☑	1☑ 2☐ 3☐	1☐ 2☑ 3☐	1☐ 2☑ 3☐
Evalue l'écart entre les effets négatifs (les coûts) et les effets positifs (les bénéfices) d'une intervention. L'approche se fonde sur les données d'une analyse financière préalable (investissement, recettes et coûts de fonctionnement) auxquelles se rajoutent les effets de l'intervention exprimés en équivalent monétaire.	Les bénéfices et les coûts sont mesurés dans la durée et agrégés à l'aide d'un taux d'actualisation[1] qui pondère la répartition des coûts et des effets dans le temps.	L'approche requiert des hypothèses sur la valeur monétaire des différents effets (monétisation).	Complexité engendrée par la combinaison des différents outils mobilisés. Les hypothèses et les résultats peuvent cependant être discutés avec les parties prenantes.	L'approche est tributaire des méthodes de monétisation et donc de leur capacité à rendre les résultats généralisables.
Analyse coût-efficacité	1☐ 2☑ 3☐	1☐ 2☑ 3☐	1☐ 2☐ 3☑	1☐ 2☑ 3☐
Les effets de l'intervention sont exprimés dans leur unité d'origine (comme le nombre d'emplois créés, nombre de vies sauvées, etc.). L'approche compare les différences de coût et d'efficacité d'une stratégie à une autre. C'est la méthode de référence en évaluation médico-économique.	Les effets et les coûts actualisés[1] sont mesurés sur toute la durée de l'impact de la politique. Lorsque les effets (positifs ou négatifs) ne peuvent être synthétisés (par exemple en années de vie pondérées par leur qualité), ils doivent être analysés séparément.	L'approche suppose que l'on sache faire une estimation quantitative des effets analysés (par expérimentation ou quasi-expérimentation).	La méthode est transparente, notamment lorsque les coûts et les effets sont aisément identifiables. Elle exprime pour le décideur le coût d'opportunité[2] qu'il y a à choisir une stratégie plutôt qu'une autre.	L'approche est tributaire des méthodes expérimentales, quasi-expérimentales et d'analyse des coûts, et donc de leur capacité à rendre les résultats généralisables.
Analyse multicritère	1☐ 2☐ 3☑	1☐ 2☐ 3☑	1☐ 2☑ 3☐	1☑ 2☐ 3☐
Evalue l'ensemble des effets d'une intervention sur la base de critères et de pondérations choisis en concertation avec les principaux acteurs et bénéficiaires du programme. Dans sa forme la plus simple, elle consiste en la création d'indicateurs composites mesurant la performance des stratégies évaluées.	Permet de prendre en compte la multiplicité des effets et des parties prenantes au projet, d'autant plus si la dimension des coûts est prise en compte en même temps que les autres dimensions.	Peut se fonder sur une estimation quantitative des effets analysés ou, si les données ne sont pas disponibles, sur une notation faites par des experts.	Le processus d'évaluation est ouvert à tous les acteurs. Certaines approches ont été outillées avec des logiciels permettant de faire des comparaisons de profils de résultats[3] et nécessitent un haut niveau d'expertise.	L'agrégation *ad hoc* des effets rend les résultats difficilement transférables (par exemple, les résultats dépendent fortement des pondérations choisies, et donc du contexte dans lequel celles-ci sont définies).

1 L'actualisation consiste à pondérer les données selon leur survenance dans le temps, et à ramener l'ensemble des coûts et des bénéfices à une même année de référence. L'approche se fonde sur l'idée qu'un bénéfice, ou qu'un coût, devrait avoir une pondération plus élevée si son impact est immédiat.
2 Le coût d'opportunité du choix d'une stratégie est la renonciation aux avantages de la stratégie non retenue (nets de ses coûts) la mieux placée en termes d'efficience après la stratégie retenue.
3 Les méthodes d'analyse multicritère incluent les analyses compensatoires, dont l'objet est la création d'un indicateur composite, et les analyses non compensatoires (Electre, Prométhée, etc.), beaucoup plus complexes, dont le but principal est de trier les stratégies par groupe.

3 Sociogramme des acteurs de l'évaluation en France

3.1 L'évaluation des politiques publiques en France

Aujourd'hui, en France, l'évaluation est connectée aux pratiques internationales. En même temps, elle s'inscrit toujours dans la tradition française de ses fondateurs et pionniers (tels Pierre Massé et Marcel Boiteux) et dans la lignée des travaux confiés par le Premier ministre au Commissariat général du Plan des années 1980 (Publication du rapport Deleau, Nioche, Penz, Poinsard, « Evaluer les politiques publiques »), travaux portés par le Conseil scientifique de l'évaluation (créé par le décret n° 90-82 du 22 janvier 1990 relatif à l'évaluation des politiques publiques). L'objectif officiel est alors « de rechercher si les moyens juridiques, administratifs ou financiers mis en œuvre permettent de produire les effets attendus de cette politique et d'atteindre les objectifs qui lui sont fixés ».[14]

L'activité évaluative en France a bénéficié depuis de nouveaux développements dans différents domaines : évaluation dans le domaine social ou urbain (comme la politique de la ville ou le suivi et l'évaluation de la rénovation urbaine) ; évaluation des politiques de l'emploi ; évaluation des programmes de développement territorial. Le domaine médico-économique (évaluation des programmes de santé) a également contribué de manière intensive à son essor. L'évaluation des fonds européens consacrés au développement socio-économique a été une étape importante d'appropriation des outils et des méthodes. Par exemple, le programme MEANS (Méthodes d'Evaluation des Actions de Nature Structurelle), créé dans les années 1990 et devenu ensuite le programme EVALSED (Evaluation of Socio-Economic Development). Il a fourni une méthodologie conciliant rigueur scientifique et spécificités propres aux interventions communautaires.[15]

L'expertise interne aux administrations (services d'études et de statistiques) est généralement mobilisée pour produire les évaluations mais la commande s'ouvre également de plus en plus aux chercheurs universitaires (publiant dans des revues scientifiques spécialisées) et aux cabinets-conseils par le moyen de marchés publics de prestations intellectuelles. Cette commande entend en outre s'inscrire dans la transparence nécessaire en démocratie et la modernisation de l'action publique (France Stratégie[16]). Même si certains ont pu parler de « retard français », l'évaluation est appelée à animer de manière croissante le débat public et en particulier le débat parlementaire en France, en particulier à l'occasion du vote de nouvelles lois (un premier exemple est fourni par la mise en place de l'évaluation des effets économiques et sociaux des ordonnances du 22 septembre 2017 relatives à

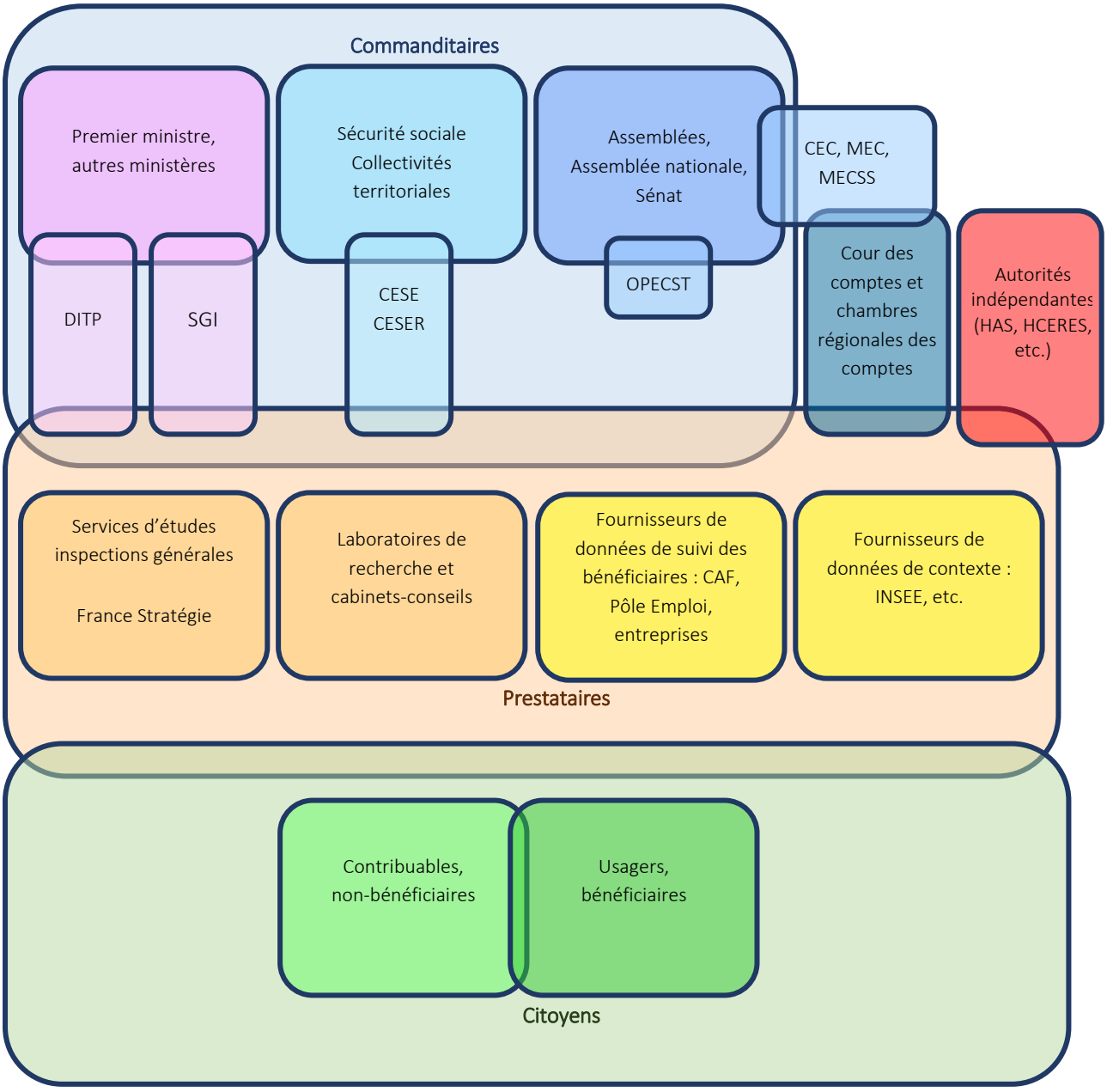
[14] Conseil scientifique de l'évaluation (1996) *Petit guide de l'évaluation des politiques publiques.* Paris : La Documentation française.

[15] Commission Européenne (2008) *EVALSED : the resource for the evaluation of socio-economic development*. DG Politique régionale.

[16] http://www.strategie.gouv.fr/thematiques/evaluation-de-politiques-publiques

la loi travail[17]). Fait marquant vu l'organisation des financements publics en France, l'évaluation a été développée également par les collectivités territoriales et dans le secteur de la santé.

Figure 3. Sociogramme simplifié des acteurs de l'évaluation en France (janvier 2018)



[17] http://www.strategie.gouv.fr/actualites/appel-idees-methode-devaluation-ordonnances-22-septembre-2017

Le sociogramme de la figure 3 offre une première présentation simplifiée (donc à enrichir) des différents acteurs et parties prenantes de la démarche évaluative en France. Par définition, un sociogramme décrit les relations qui les unissent ou au contraire des déconnexions qui les caractérisent. Il permet aussi d'identifier de manière simple et visuelle l'ensemble des protagonistes et parties prenantes de l'évaluation en France. L'outil est sommaire au sens où il ne remplace pas une représentation plus systémique des jeux d'acteurs. Il ne dit pas non plus quelles sont les capacités réelles des différents acteurs de l'EPPP. Les divers sigles utilisés à la figure 3 sont explicités ci-après. Enfin, ce sociogramme demanderait à être plus approfondi dans l'avenir lorsque les compétences de chacun des acteurs pourront être plus précisément cartographiées.

3.2 Les commanditaires en EPPP

Le Comité d'évaluation et de contrôle de l'Assemblée nationale. Créé par la réforme du Règlement du 27 mai 2009, le comité d'évaluation et de contrôle (CEC) permet à l'Assemblée nationale de mettre en œuvre la fonction d'évaluation qui lui est explicitement reconnue par l'article 24 de la Constitution. L'article 146-2 du Règlement de l'Assemblée nationale prévoit que le CEC, présidé de droit par le Président de l'Assemblée nationale, comprend également 36 membres désignés de manière à assurer une représentation proportionnelle des groupes politiques et une représentation équilibrée des commissions permanentes. Au titre de ses missions, dont le champ a été strictement délimité par le Conseil constitutionnel (décision n° 2009-581 DC du 25 juin 2009), le CEC assure les fonctions suivantes :

- assurer l'évaluation des politiques publiques transversales : le CEC, de sa propre initiative ou à la demande d'une commission permanente, évalue les politiques publiques portant sur des domaines de compétence plus larges que ceux d'une commission permanente. Chaque groupe peut obtenir de droit la réalisation d'une évaluation par session ordinaire (article 146-3 du Règlement) ; en outre, en application de L 132-6 du code des juridictions financières, le CEC peut demander l'assistance de la Cour des comptes ;
- être tenu informé des conclusions des missions d'information, qu'elles soient propres à une commission permanente, communes à plusieurs commissions permanentes ou créées par la Conférence des présidents (article 146-4 du Règlement) ;
- formuler des propositions pour l'ordre du jour de la semaine réservée par priorité au contrôle et à l'évaluation : en application de l'article 48 de la Constitution, le CEC peut « en particulier, proposer l'organisation en séance publique, de débats sans vote ou de séances de questions portant sur les conclusions de ses rapports ou sur celles des rapports des missions d'information » des commissions permanentes ou de la Conférence des présidents (article 146-7 du Règlement).

La mission d'évaluation et de contrôle (MEC). Cette mission a été mise en place au sein de la commission des finances de l'Assemblée nationale en février 1999, à la suite des conclusions du

groupe de travail sur le contrôle parlementaire et l'efficacité de la dépense publique qui fut l'initiateur de la loi organique relative aux lois de finances (LOLF). La MEC travaille en collaboration avec la Cour des comptes qui est informée préalablement du choix des thèmes retenus et dont des membres assistent à ses réunions ou sont entendus par elle. Un rapport demandé à la Cour en application de l'article 47-2 de la Constitution ou du 2° de l'article 58 de la LOLF peut constituer le point de départ de ses travaux.

La mission d'évaluation et de contrôle des lois de financement de la sécurité sociale (MECSS). Elle a été mise en place au sein de la commission chargée des affaires sociales en décembre 2004 conformément à l'article 38 de la loi n° 2004-810 du 13 août 2004 relative à l'assurance maladie. La loi organique n° 2005-881 du 2 août 2005 relative aux lois de financement de la sécurité sociale a prévu sa création, élargi son champ d'investigations et précisé ses pouvoirs aux articles L.O. 111-9, L.O. 111-9-1, L.O. 111-9-3 et L.O. 111-10 du code de la sécurité sociale. La MECSS permet de suivre de manière permanente l'application des lois de financement de la sécurité sociale et de procéder à l'évaluation de toute question relative aux finances de la sécurité sociale. Le choix des thèmes étudiés par la MECSS relève de la commission des affaires sociales, sur proposition de son bureau et après consultation de la Cour des comptes.

L'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST). Créé par la loi, l'Office est un organe d'information commun à l'Assemblée nationale et au Sénat. Composé de 18 députés et 18 sénateurs, il a pour mission, aux termes de la loi, « d'informer le Parlement des conséquences des choix de caractère scientifique et technologique afin, notamment, d'éclairer ses décisions ». Il permet ainsi au Parlement de disposer d'une expertise pour éclairer des choix politiques de long terme.

La Cour des comptes. Elle a pour mission principale de s'assurer du bon emploi de l'argent public et d'en informer les citoyens. Juridiction indépendante, elle se situe à équidistance du Parlement et du Gouvernement, qu'elle assiste l'un et l'autre, conformément à l'article 47-2 de la Constitution. La Cour assiste le Parlement et le Gouvernement dans l'évaluation des politiques publiques. Elle cherche à vérifier si les résultats d'une politique publique sont à la hauteur des objectifs fixés, et si les moyens budgétaires sont utilisés de manière efficace et efficiente. Le rôle de la Cour n'est pas de commenter les choix faits mais d'évaluer les conséquences et de formuler des recommandations pour atteindre les objectifs votés par le Parlement. Les pouvoirs publics peuvent ainsi fonder leurs décisions sur des analyses objectives.

DTIP. Le décret n° 2012-1199 du 30 octobre 2012 portant création du comité interministériel pour la modernisation de l'action publique est abrogé. Depuis les décrets du 20 novembre 2017, le secrétariat général pour la modernisation de l'action publique (SGMAP) laisse place à la direction interministérielle de la transformation publique (DITP) et à la direction interministérielle du numérique et du système d'information et de communication de l'Etat (DINSIC). La DITP est placée sous

l'autorité du ministre de l'Action et des Comptes publics, chargé de la réforme de l'Etat. La DINSIC est quant à elle placée sous l'autorité du ministre chargé du numérique. Dans le même temps, il est institué un comité interministériel de la transformation publique, chargé de définir la politique du Gouvernement dans le domaine de la transformation publique et de s'assurer de son application. Le comité interministériel de la transformation publique comprend le ministre chargé de la réforme de l'Etat, le ministre chargé du numérique et l'ensemble des membres du Gouvernement. Le comité interministériel est placé sous la présidence du Premier ministre ou par délégation de celui-ci sous la présidence du ministre chargé de la réforme de l'Etat.

France Stratégie. Héritier du Commissariat au Plan fondé en 1946, créé par décret le 22 avril 2013, modifié par décret le 24 mars 2017, le Commissariat général à la stratégie et à la prospective (CGSP) appelé France Stratégie participe et témoigne des évolutions et transformations de la société française depuis plus de soixante-dix ans. France Stratégie est un organisme de réflexion, d'expertise et de concertation, autonome, rattaché au Premier ministre. Chargée d'inventer les politiques publiques de demain et d'animer le débat public, France Stratégie a quatre missions :

- évaluer les politiques publiques : analyser ce qui fonctionne vraiment et produire un diagnostic rigoureux pour une action publique efficace, éclairée par la preuve ;
- anticiper les défis et les mutations : dessiner des trajectoires de moyen terme et des scénarios de long terme pour préparer la France au prévisible comme à l'inattendu ;
- débattre avec une pluralité de parties prenantes : dialoguer et concerter pour enrichir l'analyse des contributions du monde de la recherche, de la sphère publique, des partenaires sociaux et de la société civile ;
- proposer de nouvelles solutions : recommander des orientations et des réformes, produire ou appuyer des idées neuves pour renouveler le débat et l'action publics. France Stratégie a également la charge d'animer un réseau de sept Conseils et Centres, dont notamment le Conseil d'Orientation des Retraites, le Conseil d'Orientation pour l'Emploi, le Conseil d'Analyse Économique et le Centre d'Études Prospectives et d'Informations Internationales, le réseau Emplois Compétences. France Stratégie a publié en décembre 2017 la note d'étape concernant la présentation de la démarche d'Évaluation des ordonnances du 22 décembre 2017 relatives au dialogue social et aux relations de travail :

  > France Stratégie (2017) *Évaluation des ordonnances du 22 septembre 2017 relatives au dialogue social et aux relations de travail. Présentation de la démarche : mise en œuvre par le comité d'évaluation*.

Le CGI (Commissariat général à l'investissement) et le jury indépendant des programmes Investissements d'avenir. Le programme d'investissements d'avenir (PIA) a été doté de 35 milliards d'euros lors de son lancement en 2010 (PIA 1), puis de 12 milliards d'euros supplémentaires avec le PIA 2 mis en place en 2013. Le pilotage du PIA a été confié à un Commissariat général à

l'investissement (CGI) qui supervise l'action des opérateurs du programme, chargés de sa mise en œuvre opérationnelle en lien étroit avec les ministères concernés. Le CGI assure également une mission d'expertise des investissements publics, notamment par la tenue d'un inventaire permanent et l'organisation de contre-expertises indépendantes (exemple des contre-expertises des évaluations socio-économiques pour les transferts sur le Plateau de Saclay). Il s'est traduit par des appels à projets nationaux autour de thématiques d'avenir destinés à sélectionner des équipes d'excellence (biotechnologies, recherche fondamentale, réseaux électriques intelligents, etc.). En septembre 2015, Louis Schweitzer, commissaire général à l'investissement, a demandé à France Stratégie de constituer et d'animer un comité d'experts, chargé de l'examen à mi-parcours du PIA. Les projets ont été évalués par un jury international ou des experts indépendants réunissant des compétences de haut niveau chargé de sélectionner les projets innovants, à fort potentiel de croissance.

Le CGI a été transformé en Secrétariat général à l'investissement (SGI) par décret n° 2017-1706 du 18 décembre 2017 relatif au secrétaire général pour l'investissement. Il appuie et anime des dispositifs thématiques d'évaluation ministériels des investissements et il établit une synthèse des évaluations des programmes d'investissements à l'attention du Premier ministre et formule des propositions de réorientations des actions en tenant compte des résultats observés.

La Haute Autorité de Santé (HAS). Elle a été créée par la loi du 13 août 2004, relative à l'assurance maladie. Son mandat est de contribuer à la régulation du système de santé par l'amélioration de la qualité et de l'efficience. Ses missions sont définies aux articles 161-37 et suivants du code de la sécurité sociale. Elles peuvent être regroupées en deux activités principales : accréditation et certification des établissements sanitaires et sociaux ; évaluation et recommandation en santé publique. Pour ce qui est de sa mission d'évaluation, le décret n° 2012-1116 du 2 octobre 2012 requiert une évaluation médico-économique des produits de santé qui impactent significativement à la fois le service médical rendu et les dépenses de l'assurance maladie. Cette évaluation mobilise la Commission d'évaluation économique et de santé publique (CEESP) de la HAS : l'analyse coût-efficacité correspondante est fournie par le laboratoire pharmaceutique puis évaluée par la CEESP. En parallèle, la Commission de la transparence (CT) s'il s'agit d'un médicament, ou la Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS) dans le cas d'un dispositif médical, évalue l'amélioration du service médical rendu par le produit de santé évalué. Les deux évaluations CEESP/CT ou CEESP/CNEDiMTS sont ensuite transmises au Comité économique des produits de santé (CEPS) qui négocie le prix avec l'industriel. Le CEPS est un organisme interministériel placé sous l'autorité conjointe des ministres chargés de la santé, de la sécurité sociale et de l'économie.

Le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES). Le Haut Conseil est une autorité administrative indépendante. Il est chargé :

- d'évaluer les établissements d'enseignement supérieur et leurs regroupements, les organismes de recherche, les fondations de coopération scientifique et l'Agence nationale de la recherche ou, le cas échéant, de s'assurer de la qualité des évaluations conduites par d'autres instances ;
- d'évaluer les unités de recherche à la demande de l'établissement dont elles relèvent, en l'absence de validation des procédures d'évaluation ou en l'absence de décision de l'établissement dont relèvent ces unités de recourir à une autre instance ou, le cas échéant, de valider les procédures d'évaluation des unités de recherche par d'autres instances.

Il fonde son action, en ce qui concerne les critères d'évaluation, sur les principes d'objectivité, de transparence et d'égalité de traitement entre les structures examinées et, en ce qui concerne le choix des personnes chargées de l'évaluation, sur les principes d'expertise scientifique au meilleur niveau international, de neutralité et d'équilibre dans la représentation des thématiques et des opinions. Il veille à la prévention des conflits d'intérêts dans la constitution des comités d'experts chargés de conduire les évaluations. Il peut conduire directement des évaluations ou s'assurer de la qualité des évaluations réalisées par d'autres instances en validant les procédures retenues.

3.3 Les prestataires en EPPP

Pour produire les prestations intellectuelles en réponse aux cahiers des charges des commanditaires de toutes sortes (exécutif, législatif, collectivités territoriales, administration de santé, Commission européenne), les prestataires s'appuient sur leurs ressources en capital humain (cf. encadré 3) mais également sur les publications afférentes au domaine tant en français qu'en anglais (rapports antérieurs d'évaluations, outils et méthodes développés par les chercheurs spécialisés en évaluation de politiques publiques). Les prestataires doivent pouvoir analyser et traiter les ressources des services d'études statistiques, les données de suivi de gestion ouvertes ou rendues spécifiquement accessibles pour les besoins de l'évaluation, et les ressources des fournisseurs de données de contexte (de type INSEE).

Le champ des professionnels comprend les cabinets ayant un département secteur public et un sous-domaine Evaluation. Petits, moyens ou grands, ils sont pour la plupart réunis autour de la Société française de l'évaluation et, à l'échelle européenne, autour de la société européenne d'évaluation (https://www.europeanevaluation.org/). Ce champ des professionnels est structuré et adapté à différents types de commanditaires (Commission européenne, ministères, collectivités locales, autres commanditaires publics, etc.)[18]. **Dans les prestataires internes, la liste est longue des autorités qui peuvent aussi évaluer en direct,** par exemple :

[18] Voir : http://www.ladocumentationfrancaise.fr/dossiers/evaluation-politiques-publiques/acteurs-evaluation.shtml et Matyjasik N (2013) Des évaluations de politiques publiques et des consultants. Émergence d'un champ professionnel et segmentation des rôles. *Revue française d'administration publique*, 148 : 907-921.

- **la Haute Autorité de santé** évalue d'un point de vue médical et médico-économique les produits, actes, prestations et technologies de santé, en vue de leur admission au remboursement. Elle élabore des recommandations sur les stratégies de prise en charge ;
- **le HCERES** qui est un Haut Conseil qui a succédé à l'agence appelée AERES dans le domaine de l'Evaluation de la Recherche et de l'Enseignement Supérieur (voir le décret n° 2014-1365 du 14 novembre 2014 relatif à l'organisation et au fonctionnement du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur) ;
- **la Direction de l'évaluation et de prospective (DEP)** au sein du ministère de l'Education nationale, qui œuvre depuis 1987 sur un double registre : l'aide à la décision pour laquelle elle élabore des instruments de pilotage ; la contribution au débat public grâce à une politique de publication importante (voir par exemple la publication chaque année des indicateurs de résultats des lycées régulièrement repris dans la presse) ;
- **la DREES**, direction de la recherche, des études, de l'évaluation et de la statistique au sein du ministère de la Santé et de la Protection sociale ;
- **la DARES** ou direction de l'animation et de la recherche, des études et des statistiques au sein du ministère du Travail qui a répondu à la nécessité de mieux apprécier les effets des aides à l'emploi de plus en plus coûteuses.

Encadré 3. Les formations des prestataires en EPPP

Les formations académiques et professionnelles en évaluation de politiques publiques ont permis de produire les compétences du métier dont sont désormais porteuses des ressources humaines spécialisées de haut niveau tant dans les laboratoires de recherche que chez nombre de consultants. Sans vouloir être exhaustif, ont été particulièrement pionnières et en pointe depuis 1996 les formations de second cycle (DESS puis masters) à Sciences Po, à l'Université de Lyon, à l'Université de Rennes 1, puis à l'école d'Economie de Paris ou à l'école d'Economie de Toulouse (programmes Politiques publiques et développement). Les nouvelles habilitations confortent d'ailleurs ces sites ainsi que d'autres formations plus récentes comme à l'Université de Paris Dauphine (parcours Politiques Publiques et Opinion. Études, Enquêtes, Évaluations), à l'Université d'Aix-Marseille (Management et évaluation des projets et programmes publics), à l'Université de Picardie Jules Verne (Science politique : évaluation de politiques publiques), etc. On n'oubliera pas également le catalogue des formations continues désormais offertes par la Société française d'évaluation (http://www.sfe-asso.fr/formations-sfe/catalogue-formations-sfe), certaines Universités et écoles, ainsi que l'ENA : https://www.ena.fr/Formation-continue/Offre-formation-continue-2018/Gouverner/Evaluation-des-politiques-publiques-Session-1.

Hors de France, la société européenne d'évaluation a repéré, grâce à un travail réalisé et mis à jour en 2017 par l'Université de Berne , six formations-types de niveau master en évaluation : Anvers : Master in Development Evaluation and Management ; Berne : Certificate, Diploma and Master of Advanced Studies in Evaluation ; Huddersfield (UK): MSc Social Research and Evaluation (Formation à distance), Norwich (UK): MSc Impact Evaluation for International Development,

Oxford (UK): Evidence-Based Social Intervention and Policy Evaluation, Saarbrücken (Germany): Master of Evaluation (M.Eval).

Pour ce qui est des thèses en évaluation, à l'international, des programmes de doctorat pluridisciplinaire sont développés principalement aux Etats-Unis : par exemple, à l'école de politique publique et d'administration publique de la George Washington University (https://tspppa.gwu.edu/) ou à l'Université d'Etat du Michigan (http://progeval.msu.edu/) (voir le recensement établi par l'American Evaluation society : http://www.eval.org/p/cm/ld/fid=43). En France, même si les écoles doctorales sont généralement désormais pluridisciplinaires, les thèses de doctorat dans le champ de l'évaluation sont soutenues en économie ou en gestion ou encore en science politique, sociologie, etc.

3.4 Bénéficiaires, non-bénéficiaires et contribuables : évaluation, démocratie et communication

A la base du sociogramme des acteurs et parties prenantes, figurent des ni-ni (ni commanditaires, ni prestataires). On n'oubliera pas que tout édifice de connaissances évaluatives s'appuie sur le savoir et les comportements des bénéficiaires, des non-bénéficiaires, des contribuables qui renseignent et témoignent, des services administratifs qui mettent en œuvre, etc. Une distinction existe entre, d'un côté, les bénéficiaires directs de la politique publique évaluée et, de l'autre côté, les contribuables, pouvant financer tout ou partie du programme. Dans de nombreuses situations, ces deux populations sont très éloignées l'une de l'autre. C'est par exemple le cas lorsqu'un programme est entièrement financé par l'impôt, les bénéficiaires ne contribuant pas, ou très peu, au financement global de la politique. On peut citer ici l'exemple du Programme de réussite éducative (PRE), en partie financé par l'Etat, et ciblant les enfants entre deux et seize ans en situation de fragilité éducative au sein des quartiers prioritaires. Ce type de financement s'inscrit dans une démarche de solidarité ou de redistribution. Au contraire, lorsqu'il y a un péage, ticket modérateur ou reste à charge, les bénéficiaires seront sensibilisés non seulement aux avantages mais aussi au coût de la politique publique évaluée dont ils partageront une partie du fardeau du financement.

D'une manière générale, les non-bénéficiaires et aussi les contribuables participent de facto à certaines évaluations d'efficacité ou d'efficience. Leurs intérêts peuvent être représentés par exemple lors de la monétisation d'effets externes (retombées sur les non-bénéficiaires, par exemple les nuisances sonores d'un équipement qu'ils n'utilisent pas nécessairement) ou lors de l'introduction des coûts d'opportunité des fonds publics dans l'évaluation. Enfin, comme le note la section 2.1, la récolte de données sur les non-bénéficiaires peut aussi s'avérer primordiale si l'on cherche à évaluer l'impact d'une intervention *via* des méthodes contrefactuelles.

Les évaluations structurent et produisent de bases de données, des entretiens enregistrés ou retranscrits, des données d'enquête ou d'expérimentation, des comptes rendus de conférences de consensus. Ces « données pour évaluer »[19] ouvertes, peuvent de plus en plus être structurées et utilisées. Les évaluateurs devraient, dans un avenir proche, disposer de plateformes ou Datathèques-services pour tirer parti de l'intelligence contenue dans les masses de données de suivi.

[19] Des données pour évaluer. *Séminaire commun Cour des comptes et Société française d'évaluation*. Octobre 2015. http://www.modernisation.gouv.fr/actualites/actualites/seminaire-des-donnees-pour-evaluer-la-mobilisation-des-donnees-pour-le-suivi-et-levaluation-des-politiques-publiques

4 Des études de cas pour illustrer les pratiques de l'évaluation en France

Les six études de cas qui suivent ne sauraient rendre compte de toute la diversité des domaines d'action de l'évaluation des politiques publiques et des programmes en France. Toutefois, le présent rapport, en s'appuyant sur quelques-unes des méthodes présentées auparavant, a l'ambition de fournir des éclairages importants sur la variété des pratiques d'EPPP en France.

Dans ces six études de cas, il est visible immédiatement que les enjeux, nationaux, locaux, sectoriels portés par ces politiques sont loin d'être négligeables, ce qui renforce l'intérêt que les commanditaires et les citoyens devraient porter à l'utilisation des recommandations de l'évaluation.

Les études de cas sont présentées dans l'ordre adopté pour l'arbre de décision (section 2.4) et au travers d'une grille de lecture commune, à savoir : (1) le document référencé de l'étude ; (2) la politique évaluée ; (3) les commanditaires et porteurs de l'évaluation ; (4) les objectifs de l'évaluation ; (5) la méthode employée ; (6) les données analysées ; (7) les résultats ; (8) les recommandations de l'étude ; (9) une discussion de la méthode employée.

4.1 Étude qualitative : évaluation du Programme de réussite éducative

Document référencé

Commissariat général à l'égalité des territoires (CGET) (2015) *Enquête qualitative du programme de réussite éducative.*

Politique évaluée

Le Programme de réussite éducative (PRE), impulsé et cadré par l'Etat (loi de programmation du 18 janvier 2005 pour la cohésion sociale n° 2005-32), a pour objet de promouvoir l'accompagnement individualisé des enfants entre deux et seize ans en situation de fragilité éducative au sein des quartiers prioritaires. Le programme est piloté localement et implique de nombreux acteurs : Ville, Préfecture, Éducation nationale, Conseil départemental, Caisse d'Allocations Familiales, associations, parents d'élèves, etc. La prise en charge dépend des spécificités locales et se fait par l'intermédiaire d'équipes pluridisciplinaires de soutien, composées de professionnels de champs variés (scolaire, social, médical, socio-culturel, etc.). Chaque projet local est labellisé par une commission nationale CGET / DGESCO (Direction Générale de l'Enseignement Scolaire du Ministère de l'Éducation nationale). Pour les années 2014 et 2015, plus de 100 000 enfants ont bénéficié de ce programme, pour un budget d'environ 100 millions d'euros, dont 73% financé par l'Etat.

Commanditaires et porteurs de l'évaluation

L'évaluation est commanditée par le CGET (Commissariat général à l'égalité des territoires, service de l'État à vocation interministérielle), organisme qui conseille et appuie le Gouvernement dans la conception et la mise en œuvre des politiques de lutte contre les inégalités territoriales et pour le développement des capacités des territoires. L'étude est réalisée par Trajectoires, cabinet d'études membre du Groupe Reflex (groupement de cabinets spécialisées dans le conseil en politiques publiques et l'élaboration de projets territoriaux).

Objectifs de l'évaluation

L'objectif de l'étude est de mettre en lumière les forces et les faiblesses du dispositif, que ce soit au niveau de sa mise en œuvre, de sa pertinence ou de ses effets. Les questions posées sont notamment les suivantes : après dix ans de fonctionnement, comment les projets se sont-ils mis en place sur les territoires et quel est leur niveau d'adéquation avec les orientations nationales ? Sont-ils en capacité de proposer une réponse adaptée pour tous les enfants pris en charge, quels que soient le type et le niveau de difficultés ? Quels sont les effets pour les enfants qui en sont bénéficiaires ?

Méthode employée

L'évaluation appréhende les questions posées sous l'angle qualitatif. L'approche consiste en des entretiens avec les principaux acteurs et bénéficiaires du programme. Elle a pour objet de recenser les différents points de vue et peut être complétée par une recherche documentaire ou une observation sur le terrain. Les entretiens sont semi-directifs, c'est-à-dire appuyés sur des questionnaires qui orientent et structurent les réponses.

Données analysées

L'évaluation a porté sur huit projets, issus d'une sélection fondée sur les critères que sont leur taux d'individualisation (la part des enfants en parcours par rapport au total des bénéficiaires), la taille du territoire et son contexte socio-économique, et la structure juridique porteuse. Les programmes ont ainsi été choisis dans cinq régions différentes : l'Ile-de-France, le Nord-Pas-de-Calais, la Picardie, Rhône-Alpes et Languedoc-Roussillon. L'échantillon total analysé comprend 32 familles. Les enfants concernés étaient âgés de six à onze ans, et scolarisés à l'école élémentaire. Au total, 128 entretiens qualitatifs ont été menés, avec chaque enfant, si possible ses parents, mais aussi les professionnels intervenant auprès de lui parmi lesquels le référent du projet, l'enseignant, la psychologue scolaire, l'assistante sociale, l'animateur du centre de loisirs, etc. Cette phase a été menée durant une période de deux mois. Les résultats présentés dans le rapport se fondent également sur d'autres travaux menés par le cabinet Trajectoires sur un échantillon de 50 familles suivies sur trois ans.

Résultats

Les résultats sont de trois types. Tout d'abord, concernant la mise en œuvre du dispositif, l'étude a mis en évidence des organisations différentes dont l'origine est la spécificité des acteurs locaux impliqués. Le rapport a noté également que le type d'actions menées est tributaire de la zone géographique concernée (zones périurbaines et rurales, villes et zones urbaines plus importantes). Ensuite, l'analyse fait apparaître quatre groupes de bénéficiaires : le premier groupe est composé d'enfants rencontrant des difficultés scolaires ponctuelles liées à leur travail scolaire. Le second groupe est constitué d'enfants dont la principale cause de difficulté scolaire est liée à un problème précis et cerné (maîtrise de la langue française, santé, etc.). Le troisième et le quatrième groupes comprennent des enfants qui cumulent des difficultés multiples d'importances plus ou moins grandes. Enfin, l'étude met en avant des effets différenciés selon le type de bénéficiaires. Par exemple, pour le groupe 2, le PRE permet d'identifier la cause précise des difficultés et permet de repositionner l'enfant bénéficiaire sur une trajectoire positive. Pour les enfants du groupe 3, les effets du PRE sont également présents, mais sont plus longs à se faire ressentir. Pour le groupe 4, les obstacles sont nombreux et les familles tellement en difficulté qu'elles ne parviennent pas à se mobiliser dans la durée.

Recommandations de l'étude

Le rapport confirme que le PRE permet la construction de parcours personnalisés : pour une partie des enfants pris en charge, l'accompagnement produit des effets puisqu'il permet de les repositionner sur une trajectoire positive.

Pour autant, des améliorations peuvent être apportées au dispositif. Le rapport préconise notamment de : poursuivre la généralisation des parcours individuels ; renforcer le lien entre l'enseignant, l'enfant et le référent ; mieux cibler le public bénéficiaire du PRE afin de déclencher le plus d'effets positifs ; spécifier et cadrer les modalités de prise en charge des enfants les plus en difficulté ; mettre en place un dispositif d'évaluation à travers le suivi de cohortes, en lien avec l'Éducation nationale ; renforcer l'animation de certains projets par les collectivités locales pour une meilleure gestion opérationnelle des parcours personnalisés ; penser l'inscription du PRE dans le contexte intercommunal.

Discussion sur la méthode employée

L'approche qualitative a permis d'évaluer le PRE en situation de vie réelle, directement lors de sa mise en œuvre sur les enfants ciblés. Les entretiens effectués ont permis une retranscription des points de vue autour des différentes dimensions de l'intervention (mise en œuvre, pertinence, effets). La retranscription des entretiens a été utilisée pour étayer les arguments du rapport, sous forme de citations. De par la durée des entretiens, la méthode employée ne permet toutefois pas l'examen d'un

échantillon de taille suffisante. Une discussion de la représentativité de l'échantillon manque d'ailleurs à l'analyse, ainsi qu'une description plus approfondie des données.

Plus fondamentalement, on peut regretter l'absence de comparaison avec un groupe de contrôle, non bénéficiaire du programme, ce qui aurait permis de mieux cerner le contrefactuel, c'est-à-dire ce qu'aurait été la situation des enfants si le programme n'avait pas existé. On peut aussi noter la possibilité d'un biais de complaisance puisque, comme le note le rapport page 60, « il y a sans doute une valorisation des effets positifs, notamment par les référents de parcours ». Par exemple, l'approche quasi-expérimentale aurait pu être utilisée pour quantifier l'effet du programme. C'est d'ailleurs à cette fin qu'une étude de l'Institut des politiques publiques utilise la méthode d'appariement par score de propension[20].

La méthode qualitative n'en fournit pas moins des informations tangibles et utiles pour alimenter la réflexion sur la poursuite ou le redéploiement du dispositif évalué.

4.2 Évaluation participative : politiques de déplacement de nuit dans l'agglomération nantaise

Documents référencés

Ville de Nantes, Nantes métropole *(2017) Cahiers de l'évaluation des politiques publiques*, Mai.

Auran (2016) La mobilité, la nuit. *Les synthèses de l'Agence d'urbanisme de l'agglomération nantaise*, Juin.

Politiques évaluées

Il s'agit des politiques visant à faciliter les déplacements la nuit dans la ville. Pour encourager la construction d'une politique de la vie nocturne dynamique et apaisée, la Ville de Nantes a mis en place le Conseil nantais de la nuit. Associant l'ensemble des parties prenantes de la nuit, il a démarré sa réflexion en 2015 à l'occasion du diagnostic participatif et prospectif sur les nuits nantaises et a inscrit à son programme de travail 2016, l'évaluation des mobilités nocturnes.

Commanditaires et porteurs de l'évaluation

La Ville de Nantes s'est engagée à développer une pratique de l'évaluation qui s'appuie sur les principes de la Charte de la Société française de l'évaluation. La démarche associe les citoyens et la

20 Bressoux P, Gurgand M, Guyon N, Monnet M et Pernaudet J (2016) Evaluation des Programmes de Réussite Educative, Institut des politiques publiques, Etude réalisée en partenariat avec la DEPP et le CGET.

société civile. Inscrite au programme de travail 2016 du Conseil nantais de la nuit à la suite du diagnostic participatif des nuits nantaises (2015), elle s'est déroulée de mars à juin 2016, autour d'une instance d'évaluation ouverte et pluraliste.

Objectifs de l'évaluation

La démarche a visé à identifier les besoins et les usages des déplacements nocturnes (déplacements entre 19h30 et 6h et entre 6h et 19h30 hors week-end), évaluer l'efficacité et l'efficience des transports publics et étudier leur complémentarité avec les offres alternatives (privées, associatives…). Cinq questions méritaient des réponses : (1) Quels atouts et quelles faiblesses pour la desserte en transports en commun la nuit ? (2) Quels atouts et quelles faiblesses pour les autres offres de mobilité la nuit ? (3) Quels atouts et quelles faiblesses en matière de communication, de lisibilité de l'offre et d'information en temps réel ? (4) Quels atouts et quelles faiblesses en matière de sécurité et de tranquillité publique ? (5) Quelles mobilités nocturnes des travailleurs de nuit ?

Méthode employée

Un espace de travail et de débat a rassemblé des acteurs économiques des transports et de la nuit, des associations d'usagers et des citoyens issus du Conseil nantais de la nuit, du Conseil métropolitain des Usagers de l'Espace Public, du Conseil nantais de l'Égalité Femmes-Hommes et du Conseil Nantes&co 16-25. Associés à toutes les étapes de la démarche, ils ont produit avec l'appui du cabinet d'études ASDO, un avis évaluatif et des préconisations.

Données et recueil des informations

Le procédé a d'abord été l'enquête, une Enquête Déplacement Grand Territoire par l'Agence d'urbanisme de l'agglomération nantaise. Celle-ci a mis en œuvre 3 traversées nocturnes et permis de rencontrer 174 usagers. D'autres tests en situation (observations, enquêtes flash etc.) ont été réalisés débouchant sur les documents suivants :

- Cartographie collaborative : plus de 160 téléchargements effectués. Analyse des itinéraires nocturnes via une application de géolocalisation mobile *Nanteslanuit*, lors d'un week-end évènement du 19 au 22 mai 2016.
- Diagnostic de la situation actuelle : 24 personnes interviewées. Réalisation d'entretiens individuels et collectifs : élus, directions thématiques, acteurs du déplacement, de la nuit et du monde étudiant, travailleurs de nuit. Analyse de la fréquentation et des usages des modes de déplacement.
- *Benchmarking* participatif : 13 contributions collectées par l'Agence d'Urbanisme de la Région Nantaise.

Résultats, préconisations et suivi des recommandations

L'ensemble des résultats ont été partagés le 16 juin 2016. Plusieurs enseignements ont été dégagés et 33 propositions pour améliorer les déplacements la nuit à Nantes ont été formulées. La Ville et la Métropole ont ensuite étudié la faisabilité de ces préconisations et ont souhaité s'engager autour de 6 axes :

- assurer une meilleure cohérence entre l'offre de transport en commun, le rythme de la vie nocturne et les horaires de fermeture des établissements de nuit : prolongation le vendredi soir du fonctionnement de l'ensemble du réseau de nuit de 2 heures pour un alignement du réseau comme le samedi soir dès l'automne 2017 et ainsi couvrir la fermeture des établissements de nuit ;
- renforcer la couverture géographique et la fréquence des transports en commun pour le service de nuit : ajout de courses sur des lignes du tramway permettant ainsi un élargissement de l'amplitude horaire et de la fréquence de passage ; mise en place d'un bus en septembre 2018 pour renforcer la desserte de communes éloignées ; prolongation d'une ligne de bus en septembre 2018 pour desservir une zone d'activité ;
- améliorer les conditions de sécurité : renforcement des effectifs d'agents de prévention sur le réseau de nuit ;
- encourager les alternatives à l'automobile individuelle (covoiturage, auto-partage, vélo, marche à pied) : prise en compte de l'éclairage dans l'élaboration du plan piéton ; regroupement autour d'un unique opérateur de tous les services de vélos d'ici 2018 (avec un service de location de longue durée) ; dynamisation du covoiturage lors de spectacles et concerts ;
- améliorer la communication et la lisibilité des offres : expérimentation à l'entrée des établissements de nuit de supports de communication sur les solutions de « retour » grâce à un atelier de travail avec l'ensemble des acteurs. La collectivité demandera aux différents gestionnaires des parkings de développer dans le cadre de réponses aux appels d'offres des mesures et actions de communication pour valoriser les offres existantes telles que le tarif de 3 euros/nuit ;
- repenser collectivement le devenir du bus de nuit (la luciole) à la demande : l'intérêt de la démarche collaborative, constaté dans la mise en place de cette évaluation, rappelle la nécessité d'un espace de débat et de coproduction avec l'ensemble des acteurs (usagers, exploitant, acteurs économiques de la nuit, etc.) pour repenser le devenir du bus de nuit la Luciole pour 2018.

Discussion de la méthode

Comme dans toute évaluation participative, le regard se tournera vers la représentativité des usagers et des interviewés. La place respective des parties prenantes dans la démarche d'évaluation est

aussi un biais potentiel : place des élus, place des services, place des usagers, place des citoyens, place des contribuables.

4.3 Évaluation expérimentale : dispositifs d'aide à l'accès à l'emploi des jeunes diplômés

Document référencé

Crépon B, Duflo E, Gurgand M, Rathelot R, Zamora P (2014) Do labor market policies have displacement effects ? Evidence from a clustered randomized experiment. *The Quarterly Journal of Economics,* 128 : 531-580.

Politique évaluée

Les politiques d'aide à l'accès à l'emploi sont assez répandues dans les pays industrialisés et particulièrement en Europe du Nord. Le principe est qu'un intermédiaire privé (agence d'intérim ou structure associative) aide le chômeur dans sa recherche d'emploi et est rémunéré pour l'essentiel si ce dernier obtient un emploi stable. La plupart des actions entreprises à l'étranger ont un impact positif, particulièrement pour les chômeurs à faible risque de chômage de longue durée. La politique évaluée ici est un dispositif à grande échelle déployé en France à destination des jeunes diplômés de l'enseignement supérieur au chômage depuis au moins six mois.

Commanditaires et porteurs de l'évaluation

La loi de cohésion sociale de 2005 a supprimé le monopole de l'ANPE (Agence nationale pour l'emploi devenue Pôle Emploi en 2008) en matière d'assistance à la recherche d'emploi. Si chaque chômeur reste tenu de rencontrer régulièrement un conseiller de Pôle Emploi, il n'en est pas moins apparu que nombre d'entre eux avaient utilisé les conseils d'une agence d'intérim pour obtenir un emploi. S'inspirant des réformes Hartz en Allemagne, gouvernement et syndicats se sont entendus pour porter cette évaluation, conduite par les universitaires auteurs de l'étude citée et commanditée par la Direction de l'animation de la recherche, des études et des statistiques (DARES) du Ministère du travail.

Objectifs de l'évaluation

L'objectif de l'étude est de confirmer ou d'infirmer que le recours à des agences d'intérim dans l'accompagnement vers l'emploi des jeunes diplômés a ou non une influence sur l'obtention d'un travail stable (CDI ou CDD de six mois ou plus). La priorité dans la commande a été l'évaluation d'impact, même si la mise en relation du résultat au coût de la politique (recherche de l'efficience) a été ensuite un élément de la recommandation de l'étude.

Méthode employée

Il s'agit d'une évaluation expérimentale avec randomisation (répartition au hasard) des jeunes diplômés entre le groupe de traitement qui reçoit l'intervention (suivi par agence d'intérim), et le groupe de contrôle, qui n'en bénéficie pas. L'affectation aléatoire à l'un ou l'autre groupe permet *a priori* une comparaison robuste des résultats obtenus par la politique. Toutefois, la mise en place de l'intervention a des effets (éventuels) non seulement sur les jeunes qui la reçoivent, mais elle est également susceptible d'avoir un effet sur les non-bénéficiaires (jeunes diplômés du supérieur ou moins diplômés) qui, en concurrence pour des emplois en nombre limité, se verraient préférer les jeunes aidés du fait même de cette aide. L'expérimentation a dû tenir compte de ce possible effet de « chaises musicales » ou d'éviction entre demandeurs d'emploi. Cet effet peut en entraîner un autre, puisque l'éviction de non-bénéficiaires (si elle est avérée par l'expérimentation) diminue le taux d'accès à l'emploi du groupe de contrôle, et rend la comparaison d'autant plus avantageuse pour le groupe de traitement.

Données analysées

L'évaluation s'est appuyée initialement sur les résultats d'expérimentations américaines, danoises et françaises dans lesquelles l'effet de « chaises musicales » des politiques d'aide à la recherche d'emploi apparaît parfois positif, parfois mitigé. Cette difficulté à conclure étant vraisemblablement liée à l'existence de nombreux marchés locaux du travail (les bassins d'emploi locaux), l'évaluation étudiée ici a proposé une randomisation à deux niveaux, celui des bassins locaux d'emploi puis celui des individus. Dans la première étape, 235 bassins d'emploi locaux (identifiés par l'agence Pôle Emploi qui les supervise) répartis dans dix régions ont été regroupés en 47 strates de 5 bassins de taille et de population locale comparables. Au sein de chaque strate, les bassins d'emploi se sont vus assignés aléatoirement un pourcentage parmi 0 %, 25 %, 50 %, 75 % et 100 % (ce pourcentage représente la proportion de chômeurs qui rejoindront le groupe de traitement, les autres rejoignant le groupe de contrôle). Dans la deuxième étape, ce pourcentage est utilisé pour assigner aléatoirement les demandeurs d'emploi au groupe de traitement. Par exemple, dans la première étape, le pourcentage de 75 % a été assigné au quatrième bassin d'emploi et, dans la deuxième étape, c'est effectivement une proportion de 75 % des jeunes chômeurs diplômés étant éligibles à l'intervention qui est sélectionnée aléatoirement dans ce bassin pour participer au dispositif d'accompagnement.

Résultats

Sur l'ensemble des bassins d'emploi considérés, l'impact de l'intervention est à première vue positif, particulièrement pour les hommes qui ont 11 % de chances supplémentaires d'obtenir un emploi du fait de l'intervention. L'effet de « chaises musicales » est certes globalement très mitigé pour les hommes et inexistant pour les femmes. Il semble toutefois qu'il affecte les demandeurs d'emploi moins diplômés, ce qui tendrait à montrer qu'une partie des emplois obtenus par ces diplômés

du supérieur serait des emplois sous-qualifiés. En outre, l'effet de « chaises musicales » est beaucoup plus fort, tant sur les diplômés que sur les autres demandeurs, dans les zones en récession où l'emploi est plus rare. Au total, si l'on prend en compte cette éviction des chômeurs qui ne bénéficient pas de l'intervention, le nombre d'emplois net créés par la politique d'accompagnement vers l'emploi est apparu négligeable au regard des moyens budgétaires et organisationnels engagés par le dispositif.

Recommandations de l'évaluation

La recommandation est la circonspection, l'efficacité n'étant pas démontrée. L'apparent succès de l'accompagnement à l'emploi (pour les bénéficiaires, il y a un léger effet accélérateur de retour à l'emploi) ne doit pas en effet, au vu des résultats obtenus, masquer les effets négatifs induits par cette politique. Le coût de suivi étant environ le double de celui de l'accompagnement usuel, l'efficience n'est pas non plus au rendez-vous. En outre, considérant l'effet d'éviction (approximativement de 1 pour 1 dans certains bassins d'emploi défavorisés) sur les non-bénéficiaires, la recommandation de l'étude est d'avoir la plus grande circonspection quant à ces dispositifs.

Discussion sur la méthode employée

L'expérimentation a été un processus long et coûteux qui a nécessité une mobilisation importante des acteurs. Dans le cas de cette étude, la DARES, les régions, agences locales de l'emploi, agences d'intérim et demandeurs d'emploi (les taux de réponse de ces derniers ont été importants et équivalents dans les deux groupes de traitement et de contrôle) ont toutes contribué à la bonne marche de l'évaluation. Cela a été aussi un processus complexe qui a mobilisé des techniques statistiques et économétriques qui sans être insurmontables n'en sont pas moins sophistiquées.

L'importance des moyens engagés dans cette évaluation et la nature du résultat exigeaient que les recommandations soient concrètes et opérationnelles. Cela a été le cas dans cet exemple d'évaluation.

4.4 Évaluation quasi-expérimentale : les effets du recrutement en contrat aidé sur la trajectoire professionnelle

Document référencé

Benoteau I (2015) Quels effets du recrutement en contrat aidé sur la trajectoire professionnelle ? Une évaluation à partir du Panel 2008 de la DARES, *Document d'études de la DARES* : 192

Programme évalué

Les contrats examinés dans cette étude sont ceux issus de la loi de programmation pour la cohésion sociale du 18 janvier 2005, remplacés en 2010 par le contrat unique d'insertion. Il s'agit de contrats de

travail associant formation et/ou accompagnement professionnel pour son bénéficiaire et aide financière pour l'employeur. Pour les contrats d'accompagnement dans l'emploi (CAE) et les contrats initiative emploi (CIE), les personnes éligibles sont des personnes sans emploi rencontrant des difficultés sociales et professionnelles particulières d'accès à l'emploi. Pour les contrats d'avenir (CAV) et les contrats d'insertion revenu minimum d'activité (CI-RMA), les personnes éligibles sont des bénéficiaires d'allocations spécifiques (allocation aux adultes handicapés, allocation de parent isolé, etc.). Le dispositif évalué a pour objectif de stimuler à court terme le recrutement des salariés ayant une productivité plus faible et, à moyen terme, de favoriser leur insertion à la sortie du dispositif.

Commanditaires et porteurs de l'évaluation

La Direction de l'animation de la recherche, des études et des statistiques (DARES) est à l'origine de cette étude. Cette direction de l'administration publique centrale française dépend du ministère du Travail, de l'Emploi, de la Formation professionnelle et du Dialogue social. Elle a pour mission de produire des données ainsi que des analyses permettant d'éclairer le débat économique et social et d'apporter un appui à la conception et la mise en œuvre des politiques publiques sur les questions de travail, d'emploi, de formation professionnelle et du dialogue social.

Objectifs de l'évaluation

L'étude examine l'effet du programme sur le devenir professionnel des bénéficiaires et mesure la probabilité d'être en emploi à l'issue du dispositif. De par la forte hétérogénéité des publics ciblés, l'étude souligne la nécessité de distinguer les contrats aidés du secteur non-marchand (CAV, CAE) de ceux du secteur marchand (CIE et CI-RMA). Elle préconise également de bien distinguer les effets de court terme et de moyen terme du programme.

Méthode employée

La méthode d'évaluation est dite « quasi-expérimentale », c'est-à-dire qu'elle se fonde sur une observation des données en situation de « vie réelle », directement lors de la mise en œuvre du programme sur la population cible. Il s'agit ici de comparer la situation des bénéficiaires du contrat aidé avec celle des non-bénéficiaires, tout en tenant compte des différences de profil pouvant exister entre ces deux groupes.

La méthodologie statistique choisie a été celle de l'appariement par score de propension. Elle repose sur l'estimation de scores qui résument les caractéristiques initiales des individus et permettent la création d'un groupe de traitement (bénéficiaire de l'intervention) et d'un groupe de contrôle (non-bénéficiaire) comparables. Si un bénéficiaire du dispositif a les mêmes caractéristiques individuelles (âge, genre, niveau d'études, etc.) qu'un non-bénéficiaire, la méthode leur attribuera un score similaire : ils avaient les mêmes chances de participer au programme dans des conditions similaires.

Ils seront dès lors comparables et alloués respectivement aux groupes de traitement et de contrôle. Au contraire, si les différences sont trop grandes, la méthode exclura cet appariement de l'analyse et en cherchera d'autres. L'impact du programme est finalement calculé en comparant les taux d'insertion respectifs dans les deux groupes.

Données analysées

L'évaluation s'est appuyée sur le fichier administratif de l'ANPE et sur une enquête menée par la DARES fin 2008 et fin 2009 auprès d'un échantillon de demandeurs d'emploi qui se sont inscrits à l'ANPE au printemps 2005, et dont certains ont été recrutés en contrat aidé entre le deuxième trimestre 2005 et le deuxième trimestre 2007. L'enquête a été conçue dès l'origine pour permettre une comparaison du devenir des bénéficiaires avec celui de non-bénéficiaires. A cette fin, la base de données comprend de nombreuses variables portant sur les caractéristiques individuelles des enquêtés (âge, genre, santé, cursus antérieur, situation familiale, épisode de chômage, permis de conduire, etc.), leur devenir professionnel (taux d'accès à l'emploi non aidé à l'issue du dispositif), ainsi que des données de contexte macroéconomique (taux de chômage départemental).

Résultats

L'évaluation a été faite en trois temps. Premièrement, l'évaluation a examiné les caractéristiques des bénéficiaires et non-bénéficiaires avant appariements, et a mis en évidence de nombreuses différences entre les deux groupes, justifiant ainsi l'usage de la méthode d'appariement. L'étude s'est assurée ensuite de la qualité de l'analyse en vérifiant que les différences initiales sont effectivement réduites une fois les individus appariés. Enfin, dans un troisième temps, l'étude s'est focalisée sur la mesure de résultat, en comparant la probabilité d'être en emploi pour les deux groupes créés. Ces résultats sont deux types. Pour le secteur marchand, l'effet moyen du dispositif est positif : deux ans et demi après le début de leur contrat aidé, les salariés passés par un CIE ou un CI-RMA ont une probabilité d'être en emploi non aidé (CDD, CDI, intérim, création d'entreprise) respectivement de 23 et 17 points plus élevée qu'en l'absence du dispositif. Au contraire, pour le secteur non-marchand, l'effet moyen estimé est négatif. Ces résultats sont complétés par divers tests de robustesse et un examen de l'évolution des taux d'insertion dans le temps.

Recommandations de l'évaluation

L'évaluation montre une efficacité variable des dispositifs selon les modalités de leur mise en œuvre, les effets sur l'accès à l'emploi variant notamment selon la durée et le contenu du contrat, mais aussi selon le type d'employeur. Elle dévoile également la possibilité d'effets d'aubaine, par exemple si l'employeur recrute un salarié en contrat aidé au lieu de recruter ce même salarié sur un contrat de droit commun. Enfin, l'étude souligne « l'importance d'un ciblage approprié des contrats aidés – notamment selon la durée de l'épisode de chômage des demandeurs d'emploi – de manière à limiter

au maximum les effets d'aubaine, et à réserver l'aide publique aux personnes sur lesquelles ces dispositifs sont les plus efficaces. »

Discussion sur la méthodologie employée

L'évaluation fait plusieurs hypothèses. Par exemple, elle suppose que les personnes non-bénéficiaires de la mesure ne sont pas affectées par celle-ci. Elle ne prend pas en compte d'autres effets possibles comme l'effet du programme sur la croissance économique. En outre, comme le note l'étude, la méthode d'appariement permet de corriger les biais en tenant compte de caractéristiques dites « observables », c'est-à-dire pour lesquelles les données sont disponibles. L'approche suppose ainsi que les facteurs « inobservables » (comme par exemple la motivation des bénéficiaires, la connaissance de l'employeur) joueraient un rôle minime.

Malgré ces réserves, la méthodologie employée fournit assurément une mesure concrète de l'effet du dispositif. L'approche est contrefactuelle : elle repose sur la comparaison d'individus dont les trajectoires auraient été similaires en l'absence de la politique des contrats aidés, permettant ainsi d'isoler au mieux l'effet du programme.

4.5 Analyse d'impact budgétaire : genèse et construction du guide méthodologique de la Haute Autorité de santé sur l'impact budgétaire des médicaments innovants et onéreux

Document référencé

Ghabri S, Poullié A-I, Autin E, Josselin J-M (2017) Le guide d'analyse d'impact budgétaire de la HAS : un nouvel outil d'aide à la décision. *Santé Publique* 29 : 585-588

Politique évaluée

Le décret n° 2012-1116 du 2 octobre 2012 relatif aux missions médico-économiques de la Haute Autorité de santé (HAS) indique qu'une évaluation médico-économique (en l'occurrence une analyse coût-efficacité, dite « avis d'efficience ») est requise lorsqu'un produit de santé présente une amélioration importante du service médical rendu ainsi qu'un impact significatif sur les dépenses de l'assurance maladie. Cette évaluation intervient au moment du dépôt de la demande d'inscription au remboursement ou lors de son renouvellement. La Commission d'évaluation économique et de santé publique (CEESP) de la HAS transmet son évaluation au Comité économique des produits de santé (CEPS), en charge de la négociation de prix avec le laboratoire pharmaceutique. Après une phase pilote, les avis d'efficience sont devenus opérationnels depuis octobre 2013. Il est toutefois apparu qu'ils apportent au décideur public une information certes utile et nécessaire, mais qu'il convenait de compléter, d'où la réflexion sur la mise en place de l'Analyse d'impact budgétaire (AIB).

Commanditaires et porteurs de l'évaluation

Depuis l'accord-cadre (2016-2018) signé en janvier 2016 entre le CEPS et Les entreprises du médicament (LEEM), les laboratoires pharmaceutiques déposant un dossier d'efficience à la HAS doivent fournir une AIB qui sera évaluée par la CEESP et transmise au CEPS au même titre que l'analyse coût-efficacité. Cette obligation concerne les médicaments éligibles à l'évaluation médico-économique dont le chiffre d'affaires prévisionnel en deuxième année de commercialisation est supérieur ou égal à 50 millions d'euros.

Objectifs de l'évaluation

Afin de renforcer l'aide à la décision publique dans le champ de l'évaluation médico-économique, l'HAS est en charge de proposer un guide méthodologique sur l'AIB complémentaire à celui de l'analyse de l'efficience des interventions de santé. Il s'agit donc pour la CEESP de construire un instrument d'évaluation opérationnel et informatif.

Méthode employée

L'AIB est une approche financière qui estime, sur un horizon temporel à court ou à moyen terme, les conséquences budgétaires annuelles de l'adoption du remboursement d'un produit de santé. Cette estimation résulte du calcul de la différence entre les prévisions de dépenses sans et avec ce produit dans le « panier » à la disposition des patients.

Il faut expliquer ici pourquoi ne pas se contenter d'une analyse coût-efficacité. Celle-ci permettrait par exemple, pour une pathologie donnée, de comparer un médicament existant et un nouveau plus innovant. Elle mesurerait ainsi la variation de coût induite par une augmentation du bénéfice de santé pour un patient si le nouveau médicament lui est prescrit. Le problème est que, informative et utile, cette mesure « à l'échelle du patient » ne renseigne en rien sur les conséquences financières globales de l'adoption du nouveau médicament. Par exemple, une intervention « coût-efficace » permettant de réaliser des bénéfices de santé peut induire un impact budgétaire important, à l'instar des antiviraux d'action directe contre le virus de l'hépatite C. L'AIB va fournir la dimension financière complémentaire à l'analyse coût-efficacité. Elle évalue l'impact financier de la diffusion d'un nouveau médicament au regard de la population susceptible d'en bénéficier, pour le prix revendiqué par le laboratoire qui le fournit.

Données analysées

Les rédacteurs du guide (chefs de projet de la CEESP, universitaire membre de la CEESP) ont tout d'abord réalisé une recherche documentaire systématique sur la période comprise entre janvier 2000 et janvier 2016, centrée sur les publications scientifiques, les travaux des sociétés savantes et les

expériences nationales étrangères. La rédaction s'est également appuyée sur une étude pilote rétrospective portant sur les AIB soumises à la HAS de novembre 2011 à juin 2015 (optionnelles durant cette période) ainsi que sur une consultation publique du 9 mai au 21 juin 2016. Cette consultation s'adressait aux entreprises du médicament, aux sociétés savantes, aux sociétés de conseils, aux représentants d'associations de patients et aux experts nationaux et internationaux. Les recommandations émises dans le guide ont été présentées à la CEESP et approuvées par le Collège de la HAS.

Résultats

La mise en place du guide AIB de la HAS a privilégié une approche à la fois scientifique dans la construction des instruments d'analyse et participative, citoyenne, dans sa discussion. En particulier, la phase de consultation publique a donné lieu à de nombreuses contributions de qualité qui ont substantiellement amendé la version initiale. Le guide a été publié en janvier 2017 et un premier bilan sera tiré fin 2018. D'ores et déjà, il apparaît que la faible qualité des AIB fournies lors des premiers avis d'efficience (pour lesquels l'AIB était optionnelle) relevée lors de l'étude pilote appartient désormais au passé. Les laboratoires pharmaceutiques se sont emparés des recommandations du guide et l'information fournie au CEPS via la CEESP est d'autant plus fiable.

Recommandations de l'étude

Une des leçons de la mise en place du guide AIB à l'HAS est que l'évaluation des politiques publiques doit assurer l'équilibre entre exigence scientifique (le montage d'une AIB est un exercice complexe) et opérationnalité. Parmi les recommandations du guide, on notera par exemple : (1) la perspective financière doit être celle de l'assurance maladie obligatoire ; (2) l'horizon temporel est fonction de la planification budgétaire du financeur et du rythme de la diffusion du médicament, soit généralement entre 3 et 5 ans ; (3) il convient de s'assurer de la fiabilité des données cliniques nécessaires au calcul des coûts liés à la maladie ; (4) l'AIB présente des flux financiers annuels sous forme de coûts non actualisés ; (5) les sources d'incertitude sur les résultats doivent être explorées par des analyses en scénario réalisées sur les paramètres les plus importants susceptibles de faire varier les résultats, par exemple la taille de la population rejointe ou l'évolution des parts de marché des différents médicaments susceptibles de traiter la pathologie en question.

Discussion sur la méthode employée

Les enjeux du bon usage des ressources collectives sont particulièrement marquants dans le domaine de la santé. Comme le souligne le décret du 2 octobre 2012, l'évaluation des politiques publiques y revient à fournir des « recommandations et avis médico-économiques sur les stratégies de prévention, de soins, de prescription ou de prise en charge les plus efficientes et contribue à leur comparaison ou leur hiérarchisation dans un objectif de santé publique et d'optimisation des dépenses

d'assurance maladie ». A cet égard, l'AIB est un outil non suffisant mais nécessaire d'analyse de l'affectation des ressources. Face à des innovations thérapeutiques de plus en onéreuses, il permet de soulever la question de la soutenabilité de l'admission au remboursement de médicaments à la fois innovants, prometteurs et extrêmement coûteux, sachant que tout effort financier supplémentaire consenti, se traduit, à budget global constant, par des économies ailleurs et maintenant, ou par un endettement qui pèsera sur les générations futures.

4.6 Analyse coût-bénéfice : évaluation socio-économique des investissements publics du Commissariat général à l'investissement

Documents référencés

Charpin J-M, Ruat L, Freppel C (2016) *Évaluation des procédures d'évaluation socio-économique des projets d'investissements publics*. Inspection générale des finances.

Rapports de contre-expertise des évaluations socio-économiques au titre de la Mission relative aux investissements publics confiée par lettre en date du 20 août 2012 par le Premier ministre au Commissariat général à l'investissement (CGI), organisme créé par décret n°2010-80.

Politique évaluée

Il s'agit des projets d'investissements publics dits « d'avenir » en France. Ces projets sont inclus dans des programmes eux-mêmes inclus dans des politiques (ou missions), par exemple, la mission Recherche et enseignement supérieur. Selon le jaune budgétaire, l'inventaire des projets est réalisé annuellement : les ministères et établissements publics doivent fournir leurs fiches au plus tard fin juin de chaque année au CGI, alors que les fiches de l'année précédente sont tenues à sa disposition à tout moment. Les projets concernés sont tous ceux dont le financement cumulé de l'État, de ses établissements publics, des établissements publics de santé et des structures de coopération sanitaire excède 20 M€ hors taxes. Les financements européens, privés et des collectivités locales n'interviennent pas dans le calcul du seuil, laissant ainsi de côté tous les projets dont le co-financement par l'État ou ses opérateurs ne dépasse pas ce montant.

En 2017, bien qu'encore incomplet puisqu'il ne fait que très peu état de dépenses d'équipement ou de régénération des réseaux, l'inventaire a permis de recenser 472 projets. 522 fiches ont été reçues et été exploitées mais 50 d'entre elles ont été éliminées, le plus souvent car elles concernaient des projets dont les travaux avaient commencé en 2016, mais aussi parce que le financement apporté par l'État et ses établissements publics se révélait sous le seuil d'inventaire de 20 M€.

Le secteur des transports présente la particularité qu'il peut s'écouler de nombreuses années entre le débat public et l'enquête publique, puis entre l'enquête publique et le début de réalisation.

L'inventaire fait ainsi apparaître 12 projets (plus de 350 M€) dont la déclaration d'utilité publique est antérieure à 2010 et qui ne connaissent toujours aucun début de réalisation. Les hôpitaux, l'enseignement supérieur, tous ministères confondus, et la recherche présentent un délai global de 8 à 9 ans entre lancement des études et mise en service, mais si on compare les domaines de façon plus précise, ce sont les hôpitaux qui ont les délais de réalisation les plus longs, qu'ils compensent par des durées d'études plus courtes, peut-être du fait de nombreux guides et ratios disponibles pour guider les porteurs de projet.

Commanditaires et porteurs de l'évaluation et de la contre-expertise indépendante

Le CGI est le commanditaire. Les porteurs de projets doivent évaluer ou faire évaluer leurs projets. Une contre-expertise indépendante est demandée dans certains cas.

Plus précisément, en élargissant ses outils, sur la base de l'article 17 de la loi n° 2012-1558 du 31 décembre 2012 de programmation des finances publiques pour les années 2012 à 2017 (qui dispose que les projets font l'objet d'une évaluation socio-économique préalable et, qu'au-delà d'un seuil, cette évaluation est soumise à une contre-expertise indépendante préalable) et du décret n° 2013-1211 du 23 décembre 2013 relatif à la procédure d'évaluation des investissements publics, le CGI a élargi le contenu de l'évaluation socio-économique et les soumet à une contre-expertise indépendante (qu'il organise lui-même) lorsqu'ils dépassent 100 M€. Le CGI décide de l'équipe d'experts choisie en liaison avec le responsable du dossier et validée par le CGI sur la base des critères suivants : curriculum vitae, déclaration d'intérêts relative au projet, engagement de confidentialité et d'impartialité.

Selon le CGI, les experts apprécient si le porteur de projet a pu étayer le dimensionnement, les coûts, le calendrier et surtout la pertinence du projet. Le rapport peut être très technique et très détaillé sur certains points et doit répondre à des questions sur les méthodes, le respect des référentiels, la pertinence du choix des méthodes, la crédibilité des résultats des évaluations présentées.

Le CGI organise des réunions pour guider et animer le travail des experts, ainsi que le lancement, l'écriture des étapes, la date de remise du rapport et de retour d'expérience. A réception du rapport de contre-expertise, le CGI le complète d'un avis, et transmet les deux documents au porteur de projet, aux ministères concernés, au Premier ministre et au Parlement. Le CGI peut aussi procéder à l'organisation d'atelier d'experts.

Les contre-expertises sont rémunérées. Comme les avis, elles sont publiées sous la responsabilité de leurs auteurs. Le porteur du projet est libre d'apporter une réponse publique à ces documents.

Objectifs de l'évaluation

L'objectif de l'étude est l'évaluation socio-économique et pas seulement l'évaluation financière axée sur le temps de retour des investissements financés. L'objectif est la meilleure allocation des fonds publics consacrés à la préparation de l'avenir.

Méthode employée

Ce sont des évaluations de projets, les projets étant insérés dans des programmes (investissements d'avenir). L'évaluation au CGI et dans les contre-expertises est de type coût-bénéfice. La valeur actualisée nette du projet à évaluer est mesurée par la différence entre la somme actualisée des bénéfices retirés du projet et la somme actualisée des coûts qui lui sont attachés. L'efficience se rapporte à plusieurs dimensions de résultats (impact environnemental, impact sociétal, etc.). La valeur ajoutée socio-économique se définit du point de vue de la collectivité et, en conséquence, elle prend en compte, en plus des flux financiers, des coûts et des avantages non marchands qui auront pu être monétisés, le tout actualisé au taux d'actualisation fixé par la puissance publique. La mission présidée par Emile Quinet a produit un rapport en deux tomes (en septembre et en octobre 2013) principalement centré sur l'enrichissement et l'élargissement des calculs traditionnels (avec valorisation du temps, de l'environnement, et des réflexions sur la valeur de la vie humaine et la valeur du carbone).

D'autres critères sont également considérés comme par exemple la faisabilité administrative et technique, et la pertinence. Ne sont guère abordés certains autres critères, comme la cohérence ou l'acceptabilité sociale.

Données analysées

Chaque projet contient des données descriptives de l'investissement, des données comptables et financières et des données de contexte et d'environnement.

Résultats

On peut consulter le projet de loi de finances pour 2018 et en particulier le jaune budgétaire correspondant (annexe au projet de loi de finances pour l'évaluation des grands projets d'investissement public). En se limitant aux ministères pour lesquels les projets sont les plus nombreux, et donc pour lesquels la « normalisation » du dossier d'évaluation socio-économique est plus urgente, deux profils se distinguent nettement :

- S'agissant des transports, l'évaluation socio-économique et l'évaluation environnementale, généralement obligatoires, sont réalisées. En revanche, les évaluations financières ne sont

généralement pas disponibles au moment de l'enquête publique, et le choix du mode de réalisation et de la programmation financière se fait par la suite.

- S'agissant de l'enseignement supérieur comme des hôpitaux, aucune évaluation socio-économique complète n'a jamais été réalisée, mais le dossier d'évaluation socio-économique comporte cependant des éléments plus ou moins normés depuis 2013 pour les hôpitaux et depuis 2015 pour les établissements d'enseignement supérieur dépendant du ministère du même nom. L'évaluation existe cependant en ce qui concerne les aspects financiers et les coûts induits de fonctionnement. L'évaluation environnementale n'est pas systématique. Pour sa part, le secteur des hôpitaux se distingue nettement du fait que les projets doivent se conformer à de très nombreux référentiels établis à l'aide de nombreux benchmarks. Rien de semblable n'existe à ce jour pour l'enseignement supérieur.

Selon le jaune budgétaire, les avis de contre-expertises, les réserves et recommandations pour les projets portent fréquemment sur : l'optimisation du dimensionnement du projet ; la ré-estimation des coûts et la consolidation du plan de financement ; la préparation et l'accompagnement des ressources humaines aux évolutions organisationnelles ; la gestion des risques du projet et la professionnalisation de la fonction de pilotage et de suivi ; le manque de justification du mode de réalisation du projet retenu ; l'insertion d'indications des impacts indirects du projet sur l'environnement ; la mise en cohérence du calendrier de mise en service du projet avec ceux des autres projets adjacents ou plus généralement le renforcement de la coordination entre projets interdépendants mais portés par des maîtres d'ouvrage distincts.

Toujours selon le jaune budgétaire, les contre-expertises recommandent souvent : de mieux appliquer le cadrage méthodologique du rapport Quinet ou des guides sectoriels de référence ; de compléter les recommandations du rapport Quinet ; de mieux expliciter les hypothèses retenues dans les calculs ; de développer l'analyse des options de projet alternatives ; d'améliorer la fiabilité des prévisions d'activité et de recettes ; de développer les analyses de risques de manière à mieux cerner la sensibilité de la valeur actualisée nette socio-économique ; d'éviter les doubles comptes ; d'élargir le périmètre des impacts étudiés ; de progresser sur la prise en compte des évolutions de tarifs ; de développer des bases de données et des éléments comparatifs pertinents.

Recommandations de l'étude

Pour l'année 2017, les recommandations ou réserves ont aussi pointé l'absence d'analyses sérieuses des options de références et l'absence d'analyse des coûts et des impacts associés aux projets. Il faut aussi remarquer l'absence de participation de nombreuses parties prenantes ou porteurs d'intérêt (par exemple, les étudiants dans les investissements Recherche et Enseignement supérieur).

Discussion sur la méthode employée

La méthode d'évaluation coût-bénéfice incluant les effets socio-environnementaux est évidemment appropriée. Elle aurait pu être élargie en considérant l'option du développement de l'existant. Elle pourrait être approfondie avec une évaluation multicritère dans une optique de comparaisons de profils des projets concurrents au regard des nombreuses dimensions de résultats et d'impact.

Annexe 1. Aperçu des sites de ressources et extraction de leurs contenus

FRANCE

- Le site Eval fournit un glossaire **:**

  http://www.eval.fr/Pages/glossaire.aspx

- Société française d'évaluation, les cahiers de l'évaluation

  http://www.sfe-asso.fr/contributeur/sfe

- La charte de l'évaluation

  http://www.sfe-asso.fr/evaluation/charte-evaluation

- La vigie de l'évaluation

  https://www.evaluationvigie.fr/

- EVAL (réseau de praticiens de l'évaluation)

  http://www.eval.fr/Pages/AboutEVAL.aspx

- Le réseau francophone de l'évaluation

  http://www.portail-rfe.org/

COMMISSION EUROPEENNE

- **Emploi et affaires sociales.** La DG Emploi et affaires sociales et inclusion fournit de la guidance pour l'évaluation du Fonds social européen. Par exemple: Designing and commissioning counterfactual impact assessment (2012) *A practical guidance for European Social Fund managing Authorities*.

  « Les évaluations des programmes et des interventions financées par le biais du Fonds social européen (FSE) se sont avérées difficiles et n'ont souvent pas permis aux décideurs de tirer des conclusions factuelles concernant leur efficacité et leur efficience. Afin de renforcer les futures évaluations, la Commission européenne encourage les États membres à redoubler d'efforts pour développer des preuves crédibles des effets du FSE au-delà de ce qui aurait été réalisé en l'absence de soutien du FSE. Une telle preuve nécessite des évaluations d'impact contrefactuelles (EIC) c'est-à-dire une comparaison des résultats avec des estimations de ce qui se serait produit autrement. Cette orientation fournit des conseils pratiques sur certaines des questions clés qui doivent être prises en compte lors de la conception, de la commande et de la

réalisation des EIC. Il est destiné aux autorités de gestion du FSE et aux autres organes chargés de la mise en œuvre des programmes et des interventions financés par le FSE ».

- **Politique régionale de l'Union européenne.**

  « Le réseau d'évaluation comprend des représentants des États membres responsables de l'évaluation de la politique de cohésion. Il se réunit deux ou trois fois par an et est présidé par l'unité d'évaluation de la DG Politique régionale. Le réseau d'évaluation est un important forum de discussion entre la Commission européenne et les États membres sur les questions d'évaluation de la politique de cohésion, notamment les orientations méthodologiques, les indicateurs, les plans d'évaluation, les progrès réalisés par la Commission européenne et les États membres. La mission du réseau est également de faciliter l'échange d'expériences et de bonnes pratiques entre les États membres afin de renforcer les capacités d'évaluation dans l'ensemble de l'Union européenne ».

  http://ec.europa.eu/regional_policy/en/policy/evaluations/network/

BANQUE MONDIALE

- Independent Evaluation Group

  http://ieg.worldbankgroup.org/

Site Betterevaluation : http://www.betterevaluation.org/

- Une collaboration internationale pour améliorer la pratique et la théorie de l'évaluation en partageant et en générant des informations sur les méthodes et les approches. La série d'ateliers d'écriture de BetterEvaluation est une collection d'études de cas rédigées par des praticiens de l'évaluation qui discutent de leurs expériences.

- Sur ce site est proposée une longue liste des approches évaluatives comprenant les items suivants : Enquête de satisfaction, Étude de cas, Rapport de résultats collaboratifs, Analyse de contribution, Évaluation démocratique, Évaluation développementale, Évaluation de l'autonomisation, Évaluation horizontale, évaluation du changement le plus important, Cartographie des incidences, Évaluation participative, Évaluation rurale participative, Déviance positive, Expérimentation, Évaluation réaliste, Retour social sur l'investissement, Évaluation axée sur l'utilisation et la description des méthodes mixtes d'évaluation d'impact.

- http://www.betterevaluation.org/en/resources/guides/intro_mixed-methods_impact-evaluation

OCDE

- http://www.oecd.org/dac/evaluation/evaluationcommunitynews.htm

IOCE International organization of the community of evaluators

- https://www.ioce.net/members/regional-evaluation-organizations

EVALPARTNERS Plate-forme Web interactive par l'UNICEF et l'IOCE

- https://www.evalpartners.org/about/international-mapping-of-evaluation-associations

BELGIQUE

- Société Wallonne de l'Évaluation et de la Prospective

  https://www.facebook.com/laSWEP/

CANADA

- Revue canadienne d'évaluation de programme

  https://evaluationcanada.ca/fr/revue-canadienne-devaluation-de-programme

ETATS-UNIS

- American Evaluation Association

  http://www.eval.org

- Washington Evaluators

  http://washingtonevaluators.org/

SUISSE

- Société suisse d'évaluation

  http://www.seval.ch/fr/

Annexe 2. Discussion sur les critères de comparaison des méthodes

Rappelons et précisons les quatre premiers critères retenus pour apprécier les résultats d'une EPPP et formuler un jugement sur le degré de confiance que le lecteur du rapport peut attribuer aux résultats de l'évaluation :

- complétude : la méthode permet une vision à la fois précise (capacité à mesurer) et exhaustive (capacité à rendre compte des diverses dimensions) de ce que l'on cherche à estimer (point de vue des parties prenantes, ensemble des moyens engagés ou des effets, etc.) ;
- simplicité : la méthode est aisée à mettre en œuvre (des points de vue techniques, éthiques, logistiques, budgétaires, etc.) ;
- transparence : les informations sont accessibles dans les différentes étapes du processus d'évaluation (hypothèses émises, protocole choisi, données de suivi et statistiques employées), les résultats sont donc reproductibles et vérifiables, difficilement manipulables. C'est la démarche à suivre pour un chercheur qui veut publier dans une revue scientifique à comité de lecture.
- transférabilité : les résultats sont généralisables à la population cible ou à d'autres populations susceptibles de recevoir l'intervention.

La question est souvent posée de la contribution des EPPP à un meilleur fonctionnement de la démocratie libérale et même à son renforcement[21]. Pour une partie de la littérature, la question posée est celle de *l'Empowerment Evaluation*[22] (que l'on peut traduire par « Evaluation responsabilisante » ou « Evaluation émancipatrice des acteurs »). L'évaluation institutionnalisée et responsabilisante pour les acteurs et parties prenantes est un des compléments de la démocratie représentative de deux manières : d'une part, les connaissances portées à la connaissance des citoyens permettent à ceux-ci de mieux exprimer la demande de services publics. D'autre part, le recours à l'évaluation participative, au jury citoyen, enrichit le débat public et peut augmenter la qualité du résultat des évaluations. Il est de consensus que la diffusion des connaissances évaluatives, des leçons des évaluations est une chance pour la qualité du débat public. La Commission nationale du débat public en France peut à cet égard avoir une contribution importante.

[21] Barber, B R 1997. La démocratie forte. Editions Desclée de Brouwer

[22] Miller RL et Campbell R (2006) Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, 27 : 296-319. L'Empowerment Evaluation est corrélée aux nominatifs suivants : communauté apprenante, apprentissage organisationnel, responsabilité, renforcement des capacités, amélioration, appropriation, inclusion, justice sociale, démocratie, stratégies fondées sur des données probantes : community knowledge, organizational learning, accountability, capacity building, improvement, ownership, inclusion, social justice, democracy, evidence-based strategies.

Le critère de la participation citoyenne pourrait donc être ajouté en une cinquième colonne dans les tableaux de comparaison des méthodes afin d'approcher plus complètement les objectifs recherchés par le commanditaire.

Cette adjonction est cependant quelque peu prématurée aujourd'hui : elle supposerait qu'un sérieux problème de traduction en langage courant et médiatique soit résolu. Le monde de l'évaluation a engendré un vocabulaire professionnel et scientifique aussi précis que nécessaire, mais qui n'en reste pas moins difficilement accessible sans un investissement initial conséquent : **la question de la traduction des résultats, de leur médiatisation sur différents supports (journaux, magazines, audio-visuel, réseaux sociaux, wikis[23], etc.) est plus que jamais d'actualité. Des rapports d'évaluation peuvent être maltraités sur des supports d'information, en méconnaissance complète du domaine de l'EPPP**.

Pour que la démocratie fonctionne, l'exercice de présentation orale de la question évaluative, des forces et faiblesses des méthodes évaluatives utilisées, du degré de confiance dans les résultats et des possibilités de suivre ou non des recommandations doit être amélioré et le champ des EPPP pourra alors devenir un nouveau théâtre de la démocratie forte[24]. Ceci donne toute son importance à la mobilisation, dans le sens d'une meilleure information du public, de toutes les instances d'information et de communication des évaluations. Ceci devrait également susciter de nouveaux efforts de professionnalisation de la communication à propos des politiques publiques évaluées[25].

Les codes de communication des évaluations de politiques publiques, comme les codes de communication financière des entreprises ou des banques, doivent aujourd'hui croître en qualité. Si de nouvelles autorités indépendantes des EPPP ou les commanditaires actuels montant en capacité devaient aujourd'hui innover, ils devraient résoudre en pratique ces questions de langage et de communication.

[23] Selon Wikipedia, un wiki est « une application web qui permet la création, la modification et l'illustration collaboratives de pages à l'intérieur d'un site web. Il utilise un langage de balisage et son contenu est modifiable au moyen d'un navigateur web. C'est un outil de gestion de contenu, dont la structure implicite est minimale, tandis que la structure explicite émerge en fonction des besoins des usagers ».

[24] Barber B (1997) *Démocratie forte*, Desclée de Brouwer, (traduction française de l'ouvrage de 1984 : Strong Democracy: Participatory Politics for a New Age, University of California press).

[25] Bessières D (2010) *L'évaluation de la communication publique, entre norme gestionnaire et légitimités, des enjeux difficilement conciliables ?* Communication & Organisation (GREC/O), Presses Universitaires de Bordeaux : 65-76.

Annexe 3. Ressources bibliographiques

Pour aller plus loin, le lecteur peut se référer aux ouvrages et articles suivants :

- Alkin MC, Vo AT, Hansen M (2013) Using Logic Models to Facilitate Comparisons of Evaluation Theory. *Evaluation and Program Planning*, 38 : 1-88.
- Barbier JC (2014) Les sources intellectuelles de l'évaluation des politiques publiques. *Politiques et management public*, 31 : 13-30.
- Baslé M (2008) *Économie, conseil et gestion publique : suivi et évaluation des politiques publiques et des programmes*. Paris. Economica.
- Conseil d'analyse économique (2013) Évaluation des politiques publiques. *Les notes du conseil d'analyse économique*, 1 : 1-12.
- Desplatz R, Ferracci M (2016) *Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens*. France Stratégie.
- Direction générale du Trésor. *Les cahiers de l'Évaluation*. https://www.tresor.economie.gouv.fr/Ressources/Cahiers-de-lÉvaluation
- Josselin J-M, Le Maux B (2017) *Statistical tools for program evaluation. Methods and applications to economic policy, public health, and education*. Heidelberg: Springer
- Leeuw F (2003) Reconstructing program theories : methods available and problems to be solved. *American Journal of Evaluation*, 24 : 5-20.
- Lucertini G (2012) Evaluation public policies. Normative models beyond cost benefit analysis. *Thèse de Doctorat*, Université de Paris Dauphine.
- Patton MQ (2018). *Facilitating evaluation. Principles in practice*. New York: Guilford press.
- Perret B (2008) L'évaluation des politiques publiques: Entre culture du résultat et apprentissage collectif. *Esprit*, 12 : 142-159.
- Perret B (2014) *L'évaluation des politiques publiques*. Paris La Découverte.
- Rossi PH, Lipsey MW, Freeman HE (2003) *Evaluation, a systematic approach*. 7th edition. Sage editions.
- Scriven M (2008) The concept of a transdiscipline and Evaluation as a transdiscipline. *Journal of multidisciplinary Evaluation*, 5 : 65-66.
- Société française de l'évaluation. *Publications associées aux journées annuelles de l'évaluation*. Depuis 2000, voir site SFE : http://www.sfe-asso.fr/
- Vedung E (1997) *Public policy and program Evaluation*. Transaction publishers.

Liste des figures

FIGURE 1. CHAINE « INPUTS-OUTPUTS-RESULTATS-IMPACT » 9

FIGURE 2. ARBRE DE DECISION DES METHODES D'EVALUATION 22

FIGURE 3. SOCIOGRAMME SIMPLIFIE DES ACTEURS DE L'EVALUATION EN FRANCE 29

Liste des tableaux

TABLEAU 1. METHODES D'EVALUATION D'IMPACT : RECENSER LES EFFETS 24

TABLEAU 2. METHODES D'EVALUATION D'IMPACT : QUANTIFIER LES EFFETS 25

TABLEAU 3. METHODES DE SELECTION DES STRATEGIES : EVALUER LES MOYENS 26

TABLEAU 4. METHODES DE SELECTION DES STRATEGIES : RELIER EFFETS ET MOYENS 27

Liste des encadrés

ENCADRE 1. DEMARCHE DE RECHERCHE DOCUMENTAIRE EN EPPP 5

ENCADRE 2. CRITERES DE COMPARAISON DES METHODES 23

ENCADRE 3. LES FORMATIONS DES PRESTATAIRES EN EPPP 35