

Comment articuler les méthodes qualitatives et quantitatives pour évaluer l'impact des politiques publiques ?

Un guide à l'usage des décideurs et des praticiens

Adam Baïz
Anne Revillard



COMMENT ARTICULER LES MÉTHODES QUALITATIVES ET QUANTITATIVES POUR ÉVALUER L'IMPACT DES POLITIQUES PUBLIQUES ?

Un guide à l'usage des décideurs et des praticiens

Adam Baïz, France Stratégie
et **Anne Revillard**, LIEPP





AVANT-PROPOS

L'évaluation des politiques publiques fait partie des missions de France Stratégie depuis sa création en 2013 et y occupe une part qui s'est beaucoup accrue. Ses comités d'évaluation ont remis et publié une vingtaine de rapports, sur des politiques aussi variées que les « ordonnances Travail », la fiscalité du capital, la stratégie de prévention et de lutte contre la pauvreté ou encore le plan France Relance.

En parallèle, France Stratégie conduit un **exercice** plus général de réflexion sur les méthodes et les usages de l'évaluation des politiques publiques, afin d'identifier les bonnes pratiques et de renforcer la qualité et l'utilité des travaux évaluatifs. Ont dans cette optique été réalisées et publiées : une **bibliométrie** des évaluations d'impact publiées dans les revues académiques, une série de **monographies** faisant état des bonnes pratiques évaluatives internationales ou encore une **méta-analyse** mesurant l'utilisation réelle des travaux évaluatifs en amont et en aval du vote des lois. Des séminaires sont régulièrement organisés afin d'illustrer les apports concrets de l'évaluation des politiques publiques dans la décision politique.

Ce guide fait écho à un premier **guide** de France Stratégie¹ qui présentait les principales méthodes d'évaluations d'impact, et en premier lieu les méthodes quantitatives développant un raisonnement contrefactuel. Il vise à prolonger la réflexion méthodologique en interrogeant et en illustrant plus spécifiquement la complémentarité des méthodes quantitatives et des méthodes qualitatives dans l'évaluation d'impact. Il répond à cet égard aux enjeux méthodologiques qui se posent aux comités d'évaluation de France Stratégie, comme à tout évaluateur, notamment face à la complexité des politiques à évaluer, aux contraintes opérationnelles de l'exercice évaluatif et aux attentes que suscitent ses productions. Il s'agit de produire des évaluations qui mesurent tout autant qu'elles expliquent les impacts des interventions publiques, et qui peuvent être utilisées pratiquement par des décideurs publics dans l'exercice de leurs fonctions.

¹ Desplatz R. et Ferracci M. (2016), *Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et des praticiens*, France Stratégie, septembre.

En définitive, ce nouveau guide vise à promouvoir une articulation à la fois rigoureuse et fructueuse des méthodes, tant auprès des institutions publiques qu'auprès des chercheurs et des praticiens du secteur privé. Il n'est donc pas étonnant qu'il ait été mûri et co-écrit avec Anne Revillard, directrice du laboratoire interdisciplinaire d'évaluation des politiques publiques (LIEPP), qui s'est illustré ces dernières années par son positionnement multidisciplinaire, son souci d'exigence scientifique et son effort de communication auprès des acteurs publics. Je la remercie pour cette collaboration fructueuse entre l'administration publique et le monde académique.

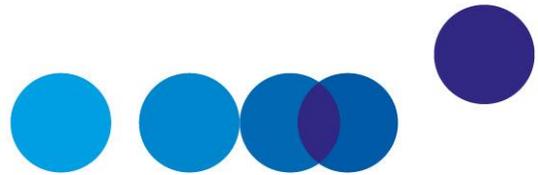
Gilles de Margerie

Commissaire général de France Stratégie



SOMMAIRE

Introduction	7
Chapitre 1 – Méthodes qualitatives et quantitatives : quelles différences ?	11
1. Les méthodes quantitatives visent à estimer l’amplitude des impacts là où les méthodes qualitatives cherchent avant tout à les expliquer.....	13
2. Les méthodes quantitatives proposent de traiter un nombre limité de variables sur un grand nombre de cas là où les méthodes qualitatives s’attachent à produire une connaissance riche sur un nombre limité de cas.....	19
3. Les méthodes quantitatives reposent surtout sur des raisonnements hypothético-déductifs là où les méthodes qualitatives mobilisent davantage le raisonnement inductif.....	23
4. Le développement de méthodes mixtes.....	26
Chapitre 2 – Quelques exemples emblématiques de méthodes qualitatives et mixtes	29
1. Outil transversal n° 1 : l’étude de cas.....	30
2. Outil transversal n° 2 : le diagramme logique des impacts.....	31
3. Méthode qualitative n° 1 d’évaluation d’impact : la cartographie conceptuelle d’impacts.....	33
4. Méthode qualitative n° 2 d’évaluation d’impact : l’analyse de contribution.....	36
5. Méthode qualitative n° 3 d’évaluation d’impact : la reconstitution de processus ou <i>process tracing</i>	37
6. Méthode mixte n° 1 d’évaluation d’impact : l’analyse qualitative comparée.....	40
7. Méthode mixte n° 2 d’évaluation d’impact : la méthode d’évaluation <i>faisabilité-impact</i>	42
8. Enrichir une expérimentation par des entretiens semi directifs.....	44
9. Une illustration des combinaisons possibles de méthodes pour approcher tout critère évaluatif.....	46
Conclusion	49
Bibliographie	51



INTRODUCTION

Dès les années 1980, et surtout depuis les années 2000, se développe le mouvement dit des « politiques publiques fondées sur des preuves empiriques » (*evidence-based policy*)¹. L'objectif est que les décisions politiques s'appuient sur des données objectives établies par des méthodes scientifiques. Il s'agit en particulier d'estimer rigoureusement l'impact des interventions publiques envisagées ou mises en œuvre, à travers des méthodes quasi expérimentales et expérimentales qui consistent à évaluer l'impact d'une intervention en comparant un groupe de traitement et un groupe de contrôle. Ces méthodes, et plus largement les méthodes quantitatives² qui reposent sur un raisonnement de type contrefactuel, se sont rapidement vu attribuer la plus grande rigueur scientifique³. Elles se trouvent portées par plusieurs réseaux, à des échelles nationales et internationales, comme la Campbell Collaboration ou l'International Initiative for Impact Evaluation. Nous pouvons citer la méthode d'expérimentation aléatoire, mais aussi la méthode des doubles différences, la méthode d'appariement, la méthode des variables instrumentales, ainsi que la simulation micro-économique ou encore la modélisation macroéconomique (voir le [Guide](#) de France Stratégie de 2016 pour une présentation détaillée de ces différentes méthodes).

Le développement des méthodes quantitatives pour l'évaluation d'impact a toutefois suscité un vif débat, encore d'actualité. Quelques critiques ont été formulées, en particulier par les chercheurs issus des sciences sociales et les évaluateurs dits *qualitativistes*. Ces critiques ne valent pas pour toutes les évaluations quantitatives, encore moins pour les plus récentes d'entre elles : de fait, les méthodes se sont rapprochées et les évaluations

¹ Bono P.-H., Debu S., Desplatz R., Hayet M., Lacouette-Fougère C. et Trannoy A. (2018), « [Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique](#) », *Document de travail*, n° 2018-06, France Stratégie, décembre.

² Schématiquement, peuvent être mentionnées l'expérimentation aléatoire, le *matching*, la régression sur discontinuité, les différences de différences, la méthode des variables instrumentales. Et, dans une moindre mesure, les études longitudinales (qui consistent à comparer un avant/après l'intervention) et les méthodes de simulation. Voir Desplatz R. et Ferracci M. (2016), [Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et des praticiens](#), France Stratégie, septembre.

³ La même rigueur et les mêmes protocoles qui caractérisaient déjà les sciences médicales dès le début du xx^e siècle.

quantitatives empruntent de plus en plus aux raisonnements culturellement associés aux approches qualitatives. Ceci étant dit, les trois principales critiques historiques ont été les suivantes :

- d'abord, il a été reproché de faire du critère d'impact le seul critère évaluatif. D'autres critères évaluatifs existent, comme la mise en œuvre ou la pertinence d'une politique publique. Ces critères valent en tant que tels pour l'évaluation des politiques publiques. Ils conditionnent et expliquent largement l'impact d'une politique publique. Ainsi par exemple, une politique pourrait ne pas produire l'impact escompté en raison d'obstacles physiques ou comportementaux corrompant sa mise en œuvre. Or les sciences sociales offrent des méthodes qualitatives rigoureuses (les entretiens, les études de cas, l'observation du terrain, etc.) qui s'attachent plus spécifiquement à évaluer ces critères qualitatifs qui peuvent aider à interpréter les résultats quantitatifs. En outre, il serait déraisonnable d'attendre d'une unique méthode qu'elle puisse répondre à toutes les questions évaluatives qui pourraient se poser ;
- en outre, il a été objecté que les méthodes quantitatives – et les méthodes expérimentales en particulier – ne sont éventuellement adaptées que pour des interventions simples (une nouvelle taxe par exemple), dans des environnements relativement stables (où il est possible de représenter et de prédire les autres variables). Elles sont, au contraire, peu adaptées : (a) aux interventions articulant de nombreux dispositifs, acteurs et environnements (pour lesquelles la mise en place d'un contrefactuel est difficile, voire impossible pour des raisons pratiques, juridiques ou encore éthiques) ; (b) aux interventions associant des mécanismes causaux intriqués, rétroactifs ou non linéaires¹ ; (c) ou encore aux interventions se réduisant difficilement à des indicateurs quantitatifs (par exemple, la gouvernance d'une entreprise, la qualité de vie, etc.). Par contraste, et parce qu'elles s'attachent généralement à ouvrir la « boîte noire » des politiques publiques, les méthodes qualitatives peuvent s'avérer plus pertinentes pour tenir compte de la complexité d'une politique publique ;
- enfin, il a été rappelé qu'il ne suffit pas que des résultats évaluatifs soient conçus comme robustes scientifiquement – ou présentés comme tels – pour qu'ils soient effectivement utilisés par les décideurs². En effet, les résultats issus des méthodes quantitatives peuvent manquer d'intelligibilité pour un lecteur profane, ou sembler abstrait et peu opérationnel³.

¹ Rogers P. J. (2008), « [Using programme theory to evaluate complicated and complex aspects of intervention](#) », *Evaluation*, vol. 14(1), janvier, p. 29-48.

² Voir notamment Cairney P. (2016), *The Politics of Evidence-Based Policy Making*, Basingstoke, Palgrave Macmillan ; Head B. W. (2010), « Reconsidering evidence-based policy: Key issues and challenges », *Policy and Society*, vol. 29(2), mai, p. 77-94 ; Head B. W. (2016), « [Toward more "evidence-informed" policy making?](#) », *Public Administration Review*, vol. 76(3), mai-juin, p. 472-484.

³ Quantifier l'amplitude d'un impact ne suffit généralement pas à en comprendre les ressorts, ni à identifier des leviers d'action pour améliorer l'intervention publique.

Par contraste, les méthodes qualitatives présentent plus souvent l'avantage de formuler des conclusions plus intelligibles et plus opérationnelles : leur appropriation par les parties prenantes est d'autant plus facilitée que ces dernières sont régulièrement associées dans l'exercice évaluatif (dans le cadre d'entretiens, par exemple).

En définitive, les méthodes quantitatives ne peuvent répondre à toutes les questions évaluatives et leurs résultats – aussi rigoureux soient-ils – ne sauraient suffire à éclairer le débat public. Prenons un exemple : l'impact du nombre d'élèves sur la performance scolaire. Certaines évaluations quantitatives ont visé à apprécier cet impact à partir de méthodes contrefactuelles¹. Il en ressort généralement que la baisse de la taille des classes impacte positivement, en moyenne, la performance individuelle des élèves et que cet impact dépend de plusieurs facteurs (comme le capital socioculturel de la famille de l'élève). Si ce résultat fait donc relativement consensus, et s'il a le mérite de la clarté, est-il pour autant suffisant pour guider l'action publique ? À partir d'une évaluation strictement quantitative de l'impact, il est permis d'en douter. En effet, un grand nombre de questions restent sans réponse. Pourquoi l'effectif de la classe a-t-il un impact sur la performance individuelle de ses élèves ? Par quels mécanismes l'impact, positif en moyenne, est parfois nul, voire négatif dans certains contextes ou pour certains types d'élèves ? Qu'en pensent les acteurs concernés (élèves, enseignants, parents d'élèves, psychologues, chefs d'établissement, etc.) ? Comment mettre en œuvre une telle réduction des effectifs en prenant en compte les potentiels biais et les obstacles opérationnels ? Et ainsi de suite. Si les évaluations quantitatives ont vraisemblablement inspiré la décision du gouvernement, en 2017, de dédoubler certaines classes primaires relevant de l'éducation prioritaire, il est fort peu probable qu'elles aient suffi à éclairer l'ensemble des questions soulevées par une telle mesure.

Que peuvent apporter les méthodes qualitatives aux évaluations d'impact ? Comment peuvent-elles être articulées de façon fructueuse avec les méthodes quantitatives ? C'est dans le cadre de ces mouvements historiques et de ces débats méthodologiques que s'inscrit le présent guide qui se veut opérationnel et pédagogique. Il s'agit ici d'étudier la complémentarité des méthodes quantitatives et qualitatives pour l'évaluation d'impact des politiques publiques. Pour permettre un recul véritable sur la méthodologie, nous commençons par rappeler les principales différences entre ces deux approches d'un point de vue épistémologique. Puis nous présentons l'émergence et l'opportunité de mobiliser des

¹ Par exemple, l'évaluation réalisée par Piketty et Valdenaire mobilise une régression sur discontinuité : Piketty T. et Valdenaire M. (2006), « [L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français. Estimations à partir du panel primaire 1997 et du panel secondaire 1995](#) », *Les Dossiers de l'Éducation nationale*, n° 173, Direction de l'évaluation et de la prospective, mars. Pour une analyse multifactorielle, voir par exemple : Bandiera O., Larcinese V. et Rasul I. (2010), « [Heterogeneous class size effects: New evidence from a panel of university students](#) », *The Economic Journal*, vol. 120(549), décembre, p. 1365-1398.

méthodes qualitatives ou mixtes en évaluation. Enfin nous développons quelques exemples emblématiques de méthodes qualitatives et de méthodes mixtes ainsi que des exemples concrets d'évaluation.

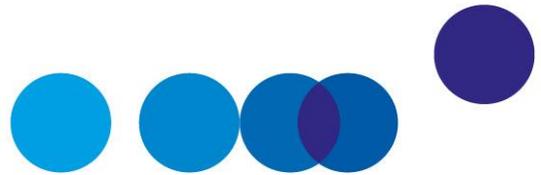
Encadré – Pourquoi une collaboration entre France Stratégie et le LIEPP ?

Le présent guide s'inscrit dans la pratique et la réflexion menée tant par France Stratégie que par le LIEPP en matière d'évaluation des politiques publiques. Il a été corédigé avec l'objectif commun de prendre du recul sur les méthodes, et de promouvoir une articulation à la fois rigoureuse et fructueuse des approches quantitatives et qualitatives, tant auprès des institutions publiques qu'auprès des chercheurs et des praticiens du secteur privé.

D'un côté, il complète le guide publié par France Stratégie en 2016¹ qui, autour de l'évaluation d'impact, porte avant tout sur les méthodes quantitatives, et évoque en complément quelques méthodes qualitatives (les entretiens, l'analyse de matériaux, l'observation du terrain et les études de cas ou monographies). Le guide s'inscrit dans un **exercice réflexif** plus général, conduit par l'institution, sur les méthodes et les usages de l'évaluation des politiques publiques. Il répond aussi à la pratique des comités d'évaluation qui sont pilotés à France Stratégie, et qui porte sur des dispositifs et politiques diverses : la stratégie nationale de prévention et de lutte contre la pauvreté, les ordonnances Travail, les politiques d'innovation, les réformes de la fiscalité du capital, etc. Face à la complexité des interventions publiques évaluées, ces comités doivent en effet mobiliser des méthodes variées, allant des entretiens aux estimations économétriques, en passant par les consultations citoyennes, la microsimulation ou encore les sondages.

D'un autre côté, le Laboratoire interdisciplinaire d'évaluation des politiques publiques (**LIEPP**) est une plateforme de recherche de Sciences Po visant précisément à promouvoir une approche multidisciplinaire de l'évaluation des politiques publiques. Ses recherches portent également sur des politiques variées – éducatives, environnementales, migratoires, fiscales, etc. – et ses méthodes reposent sur des combinaisons innovantes de méthodes quantitatives et qualitatives, visant un haut niveau d'exigence scientifique et d'intelligibilité auprès des acteurs publics. Par ses activités, le LIEPP œuvre à mieux ancrer la démarche d'évaluation dans la recherche académique en sciences sociales. Il contribue aussi au débat public et aux réflexions sur l'évaluation impliquant d'autres acteurs (administratifs, parlementaires, associatifs, etc.), dans une optique de promotion de la démarche d'évaluation et d'amélioration de sa qualité et de son utilité pour l'action publique.

¹ Desplat R. et Ferracci M. (2016), *Comment évaluer l'impact des politiques publiques ?*, op. cit.



CHAPITRE 1

MÉTHODES QUALITATIVES ET QUANTITATIVES : QUELLES DIFFÉRENCES ?

Définir les méthodes quantitatives et qualitatives n'est pas un exercice simple. De prime abord, il est tentant de les distinguer sur un plan formel : les méthodes quantitatives reposeraient sur des chiffres là où les qualitatives reposeraient sur des mots. Mais cette distinction présente immédiatement des limites. En effet, les méthodes quantitatives mobilisent toujours des mots aussi, ne serait-ce que pour interpréter les chiffres qu'elles produisent. De même, les méthodes qualitatives mobilisent généralement des chiffres, à l'instar d'un entretien sociologique qui serait accompagné de statistiques descriptives (sur la distribution d'âge des personnes interrogées, par exemple) ou d'une analyse lexicométrique.

Même si elles ne sont pas parfaites¹, les distinctions entre les méthodes quantitatives et les méthodes qualitatives sont plutôt à rechercher sur le plan épistémologique, c'est-à-dire du côté des spécificités qui entourent la production des connaissances dans chacune de ces deux approches. De façon idéal-typique, il existerait ainsi une *culture* quantitative et une *culture* qualitative²

Pour l'évaluation d'impact des politiques publiques, ces cultures se distinguent diversement. D'un côté, les méthodes quantitatives ont tendance à s'inspirer des

¹ Ces distinctions peuvent être mises à mal par des contre-exemples, lorsqu'une méthode qualitative emprunte une caractéristique présumée spécifique aux méthodes quantitatives, ou l'inverse.

² Mahoney et Goertz (2006) proposent de distinguer dix formes d'associations tendancielles pour distinguer les méthodes qualitatives des méthodes quantitatives : « statistiques *versus* logique », « estimation des effets *versus* explication d'un résultat », « approche orientée vers une population ou orientée vers les cas », etc. Voir Mahoney J. et Goertz G. (2006), « [A tale of two cultures: Contrasting quantitative and qualitative research](#) », *Political Analysis*, vol. 14(3), p. 227-249.

méthodes consacrées en épidémiologie¹. À l'instar d'un essai clinique portant sur un médicament, ces méthodes appréhendent généralement la politique publique comme un traitement (*stimulus*) touchant une population spécifique en vue de provoquer une réponse spécifique. L'amplitude de l'impact est appréciée selon un raisonnement contrefactuel, consistant à comparer un groupe de traitement à un groupe de contrôle. La causalité est le plus souvent présumée linéaire. L'environnement est supposé stable, ou du moins paramétrable (afin de pouvoir apprécier l'impact indépendamment des contextes particuliers²). Et la politique publique est généralement présumée appliquée, sauf preuve manifeste du contraire.

Par contraste, les méthodes qualitatives, plus spécifiques aux sciences sociales, reposent généralement sur des raisonnements non contrefactuels, où la causalité est par exemple appréciée au regard de l'enchaînement des phénomènes observés plutôt que par la comparaison d'un groupe de traitement avec un groupe de contrôle. L'impact peut opérer via des causalités complexes qu'il s'agit de révéler et d'expliquer. Le contexte et les représentations des acteurs sont pris en compte car ils peuvent conditionner l'impact, de même que la mise en œuvre de l'intervention est étudiée en tant que telle, car son déroulement peut également conditionner l'impact. Ainsi, la question posée n'est pas tant de savoir *ce qui fonctionne bien partout en moyenne*, mais *ce qui fonctionne mieux, où, dans quelles circonstances et pourquoi*.

À une reformulation près, trois principales distinctions épistémologiques ressortent, et nous les développons par la suite :

- les méthodes quantitatives visent à estimer l'amplitude des impacts là où les méthodes qualitatives cherchent avant tout à les expliquer ;
- les méthodes quantitatives proposent de traiter un nombre limité de variables sur un grand nombre de cas là où les méthodes qualitatives s'attachent à produire une connaissance riche sur un nombre limité de cas ;
- les méthodes quantitatives reposent surtout sur un raisonnement hypothético-déductif quand les méthodes qualitatives mobilisent davantage le raisonnement inductif, une démarche scientifique qui laisse place à la subjectivité du chercheur et des personnes interrogées.

¹ Stame N. (2009), « [Avantages et inconvénients des différentes méthodes d'évaluation : comment choisir ?](#) », in Trosa S. (dir.), *Évaluer les politiques publiques pour améliorer l'action publique. Une perspective internationale*, Institut de la gestion publique et du développement économique, p. 35-55.

² Par l'emploi notamment de variables dites *de contrôle*.

Nota Bene : La distinction opérée dans ce guide entre les méthodes quantitatives et les méthodes qualitatives doit être appréhendée de façon circonstanciée. En effet, de plus en plus rares sont les travaux qui reposeraient uniquement sur une méthode quantitative ou sur une méthode qualitative. Surtout, toute méthode quantitative peut avoir des traits communs avec certaines méthodes qualitatives, et inversement. Par exemple, les évaluations d'impact qui reposent sur des modélisations économiques empruntent *de facto* à l'approche qualitative (en cherchant à expliciter les mécanismes causaux qui sont à l'œuvre) et à l'approche quantitative (en cherchant ensuite à estimer l'amplitude des effets). Il ne s'agit donc pas d'établir une séparation nette entre ces types de méthodes, ni de prétendre représenter la diversité des travaux évaluatifs tels qu'ils existent, mais plutôt de caractériser ce qui distingue ces deux approches polaires – deux « cultures » pour citer Mahoney et Goertz (2006) – et les spécificités idéal-typiques qu'il est raisonnable de leur associer. Ainsi, il devient possible de souligner les avantages qu'il y a à croiser ces deux cultures dans le cadre des méthodes mixtes, et de comprendre les ressorts et les enjeux méthodologiques que ce croisement suscite dans l'exercice d'évaluation d'impact des politiques publiques.

1. Les méthodes quantitatives visent à estimer l'amplitude des impacts là où les méthodes qualitatives cherchent avant tout à les expliquer

L'**impact** d'une politique publique renvoie à ses effets propres. Il ne s'agit pas seulement de suivre l'évolution d'une situation, mais de parvenir à établir un lien de causalité entre l'intervention publique et l'évolution de cette situation. La notion de *causalité* est donc consubstantielle à celle d'impact¹. Sur une variable d'intérêt donnée – le taux de chômage par exemple –, il s'agit de distinguer l'effet propre d'une intervention publique des effets produits par d'autres facteurs d'influence. Par exemple, si le taux de chômage baisse de 3 % à la suite d'une réforme des cotisations sociales, rien ne garantit *a priori* que cette baisse soit imputable à la réforme, et encore moins qu'elle soit *entièrement* imputable à la réforme. D'autres facteurs (et potentiellement une infinité d'autres facteurs), comme la démographie, le PIB par habitant, le prix de l'énergie au niveau mondial ou encore les politiques éducatives, pourraient aussi exercer une influence sur le taux de chômage. Le raisonnement en causalité vise précisément à contrôler ces autres facteurs d'influence, et à séparer leurs effets des effets de l'intervention publique qu'il est question d'évaluer.

¹ À la différence du critère d'*efficacité* qui s'attache seulement à décrire l'évolution de la situation au regard des objectifs fixés.

À cet égard, les méthodes aussi bien quantitatives que qualitatives partagent ce même souci d'éliminer les explications concurrentes¹. Des différences de conception de la causalité existent toutefois.

Dans le champ de l'évaluation d'impact, les méthodes quantitatives (économétrie, modélisation mathématique) visent à évaluer l'existence d'un impact, et le cas échéant à en mesurer l'amplitude, de manière chiffrée donc. Ces méthodes, et en particulier celles qui prévalent dans les évaluations expérimentales et quasi expérimentales, partagent généralement une conception de la causalité qui peut être qualifiée de **contrefactuelle**. Un raisonnement contrefactuel consiste à poser la question suivante : *que se serait-il passé en l'absence d'intervention ?* Concrètement, il s'agit de considérer deux groupes : le groupe de traitement, où l'intervention publique a lieu, et le groupe de contrôle, où *toutes choses sont égales par ailleurs*, elle n'a pas lieu. La situation observée dans le groupe de traitement est alors comparée à celle du groupe de contrôle : les deux groupes étant initialement *identiques par ailleurs*, et en admettant qu'ils le seraient restés en l'absence de l'intervention, toute différence est alors imputée à la mise en œuvre de l'intervention dans le premier groupe relativement à son absence dans le second. Par exemple, si le taux de chômage baisse de 3 % dans une région où la réforme a lieu, et de 2 % dans une autre région identique à la première, mais où la réforme n'a pas lieu, l'impact sera estimé à -1 % sur le taux de chômage.

Toute la difficulté du raisonnement contrefactuel réside dans le point de départ, à savoir partir de deux groupes identiques en tous points, et qui le resteraient dans la durée (en l'absence de l'intervention). En pratique, il est en effet impossible de répliquer à l'identique un premier groupe en un second. Deux individus, deux groupes d'individus et *a fortiori* deux pays, aussi ressemblants soient-ils, présentent toujours des spécificités et donc des différences. Le raisonnement contrefactuel cherche donc, en pratique, à partir de deux groupes *les plus similaires possibles*, en particulier au regard des caractéristiques jugées les plus directement liées à la variable d'intérêt. Ainsi par exemple, l'impact d'une réforme des cotisations sociales sur l'emploi pourra *grosso modo* être apprécié en comparant le pays concerné avec un autre pays où le niveau des cotisations sociales est initialement proche, et où certaines caractéristiques clés, comme la démographie et le PIB par habitant, sont également proches (en utilisant par exemple la méthode dite des doubles différences). Cette conception de la causalité découle de l'argument avancé par le philosophe David Hume (1711-1776) selon lequel la causalité ne peut être directement observée, mais doit être inférée à partir de l'observation de relations entre variables et de la comparaison de situations *comparables*². Du fait des différences persistantes – qu'elles

¹ Mohr L. B. (1999), « The qualitative method of impact analysis », *American Journal of Evaluation*, vol. 20(1), mars, p. 69-84.

² Maxwell J. A. (2012), *A Realist Approach for Qualitative Research*, Londres, Sage.

soient avérées ou potentielles – entre le groupe de traitement et le groupe de contrôle, les estimations d'impact ne sauraient être établies de façon certaine. Les résultats des estimations, dans les méthodes quantitatives, s'expriment donc en principe sous la forme d'intervalles de confiance. Par exemple, l'effet d'une politique sur la création d'emplois pourra s'exprimer de la façon suivante : « [100k ; 150k] avec un degré de confiance à 95 % ». Ce résultat signifie qu'il y a statistiquement 95 % de chances, au regard des observations réalisées et des incertitudes d'estimation, que le véritable impact sur l'emploi se situe entre 100 000 et 150 000 emplois créés¹.

Du côté des méthodes qualitatives, la conception contrefactuelle est le plus souvent délaissée au profit d'une conception **processuelle** de la causalité². Dans cette conception, parfois qualifiée aussi de *physique* ou de *mécanique*³, la causalité est décrite comme un enchaînement de maillons articulés par des relations de causes à effets. Là où les méthodes quantitatives ont tendance à appréhender l'intervention publique d'un seul tenant, les méthodes qualitatives chercheront à ouvrir la « boîte noire » de l'intervention publique et à décrire comme autant de causalités intermédiaires les étapes de son processus⁴. Là où les méthodes quantitatives se contentent généralement d'inférer la causalité d'ensemble d'une intervention, les méthodes qualitatives proposeront d'observer directement chacune des causalités intermédiaires qui la composent. Reprenons l'exemple de la baisse des cotisations sociales et de son impact sur l'emploi. Dans une méthode quantitative, on cherchera à inférer ce lien de causalité, par exemple en comparant l'emploi dans le pays où la réforme a lieu avec l'emploi dans un pays similaire où la réforme n'a pas lieu. Dans les méthodes qualitatives, selon la conception processuelle de la causalité, il s'agira de décrire les principaux maillons de causalité *qui partent de la réforme jusqu'à l'emploi*, et d'observer aussi directement que possible ces causalités intermédiaires. Il sera question, par exemple, de vérifier que la réforme a bien eu lieu et qu'elle est bien connue des entreprises. Il pourra être question, ensuite,

¹ Cela implique aussi qu'il reste 5 % de chances que le véritable impact sur l'emploi soit en deçà de 100 000 emplois créés ou au-delà de 150 000 emplois créés.

² Certains auteurs vont jusqu'à contester la pertinence de l'aspiration à l'établissement de relations causales dans le contexte des méthodes qualitatives, voire des sciences sociales en général. Voir Guba E. G. et Lincoln Y. S. (1994), « Competing paradigms in qualitative research », in Denzin N. K. et Lincoln Y. S. (dir.), *Handbook of Qualitative Research*, Los Angeles, Sage, p. 105-117 ; Maxwell J. A. (2012), *A Realist Approach for Qualitative Research*, *op. cit.*

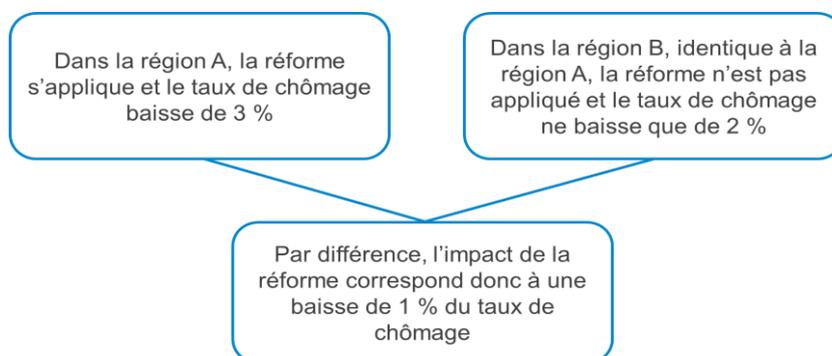
³ Cette approche de causalité vaut également dans de nombreux cadres expérimentaux en sciences dures. Par exemple, l'observation au microscope de l'action d'une molécule sur une cellule est une façon d'observer l'impact de la molécule selon une approche processuelle de la causalité.

⁴ Voir Becker H. S. (2002), *Les Ficelles du métier*, Paris, La Découverte ; Collier D. (2011), « [Understanding process tracing](#) », *Political Science & Politics*, vol. 44(4), octobre, p. 823-830 ; Trampusch C. et Palier B. (2016), « Between X and Y: how process tracing contributes to opening the black box of causality », *New Political Economy*, vol. 21(5), p. 437-454.

d'interroger les chefs d'entreprise sur leurs perceptions de la réforme, puis d'observer sur le terrain les arbitrages réalisés en termes d'embauche. La méthode qualitative pourra ainsi aboutir à des conclusions du type : « La réforme est bien connue des chefs d'entreprise et reçoit un accueil positif ; de plus, une majorité des chefs d'entreprise déclarent avoir créé plus d'emplois qu'ils ne l'auraient fait en l'absence de la réforme ». L'enchaînement de ces étapes décrit directement comment le chemin de causalité opère, et vise également à établir l'impact spécifiquement imputable à la réforme.

Figure 1 – Une représentation schématique de la causalité contrefactuelle et de la causalité processuelle

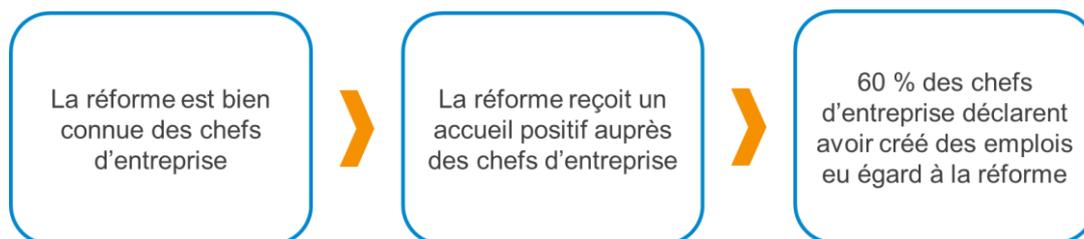
a/ La causalité contrefactuelle



Lecture : dans l'approche contrefactuelle, la comparaison du groupe de traitement et du groupe de contrôle vise à inférer l'impact de l'intervention publique.

Source : France Stratégie

b/ La causalité processuelle



Lecture : dans l'approche processuelle, il est question d'observer directement les différents maillons causaux conduisant à générer un impact.

Source : France Stratégie

Il est à noter que certaines méthodes qualitatives reposent également sur un raisonnement contrefactuel, qui diffère toutefois de celui qui opère dans le cadre des méthodes quantitatives¹. Pour rappel, le raisonnement contrefactuel dans les méthodes quantitatives

¹ Mahoney J. et Goertz G. (2006), « A tale of two cultures... », *op. cit.*

consiste à observer deux situations similaires (à défaut d'être identiques), avec ou sans l'intervention publique, et à inférer l'impact en comparant les deux situations finales. Il s'agit d'un contrefactuel en termes de *condition nécessaire*¹ : toutes choses étant égales par ailleurs, si Y a lieu en présence de X, et s'il n'a pas lieu en l'absence de X, alors X cause Y. Une variante de cette approche, que l'on retrouve dans certaines méthodes qualitatives, repose sur la notion de cause « INUS »².

Dans cette approche, il s'agit d'identifier toutes les configurations où l'impact recherché a lieu, et de les distinguer des configurations où il n'a pas lieu. Concrètement, on commence généralement par décomposer l'intervention en ses différentes composantes³, afin d'identifier celles qui peuvent – *seules ou regroupées* – générer un effet. Ainsi par exemple, en admettant qu'une politique de lutte contre le tabagisme ait été mise en place en 1980 pour la première fois, les méthodes quantitatives auront tendance à appréhender cette politique d'un seul tenant (un unique *stimulus*) et à en inférer les effets d'ensemble. Les méthodes qualitatives chercheront plutôt à décomposer – *conceptuellement* du moins – la politique : il s'agira par exemple ici d'isoler le paquet neutre, les messages d'avertissement, la taxation du paquet de cigarettes ou encore les campagnes de sensibilisation. Puis, en comparant des situations variées de mise en œuvre de la politique (avec ou sans paquet neutre, avec ou sans campagne de sensibilisation, etc.), on pourra observer les effets générés par telle ou telle composante. Le contexte même de l'intervention pourra également être décomposé en ses différentes composantes. Il s'agira, sur le même exemple, d'observer les effets de l'intervention selon le profil des consommateurs ciblés, l'activisme des lobbies ou encore les représentations culturelles associées au tabac⁴. Selon les composantes contextuelles en présence, l'impact spécifique de l'intervention pourra être ainsi relevé, alors que les méthodes quantitatives auront tendance à observer un impact *en moyenne*. Au croisement des composantes de l'intervention et des composantes contextuelles, il apparaîtra par exemple que le paquet neutre seul ou la taxe seule n'a aucun effet sur les femmes aux faibles revenus en présence d'un lobbysme fort de l'industrie du tabac ; mais que conjugués l'un à l'autre, l'impact devient important.

¹ Voir Bono P.-H. et Trannoy A. (2012), « [Causalité et contrefactualité dans l'évaluation économique des politiques publiques](#) », *Labyrinthe*, vol. 39, p. 35-53.

² *Insufficient but Non-redundant parts of an Unnecessary but Sufficient condition*.

³ Là où les méthodes quantitatives, dans une approche *behavioriste*, résument le plus souvent l'intervention à un unique stimulus dont il s'agirait d'inférer les effets.

⁴ Cette approche de la causalité est parfois qualifiée aussi de « factorielle », au sens où il s'agit d'identifier les facteurs qui génèrent ou conditionnent l'impact. Les méthodes quantitatives empruntent également à cette approche dès lors qu'elles explicitent les facteurs (dans une régression économétrique, par exemple) pouvant conditionner un impact.

Ainsi les composantes de l'intervention et de son contexte d'application sont appréhendées comme autant d'ingrédients potentiels de son succès ou de son échec¹. Par un jeu de comparaison qui est propre au raisonnement contrefactuel, il s'agit ainsi d'identifier les combinaisons de facteurs qui semblent corrélées, nécessaires ou suffisantes au succès ou à l'échec de l'intervention (nous y reviendrons avec la méthode d'analyse comparée).

En définitive, là où les méthodes quantitatives visent en principe à établir l'existence d'un impact et à quantifier l'amplitude, les méthodes qualitatives peuvent – en complément – avoir deux plus-values principales :

- en cherchant à observer le plus directement possible le processus de l'intervention, et les causalités intermédiaires qui la composent, elles peuvent aider à comprendre pourquoi une intervention produit tel ou tel impact ;
- en cherchant à décomposer l'intervention et son contexte d'application en leurs composantes clés, elles peuvent aider à révéler les facteurs (ou les combinaisons de facteurs) qui sont associés à un impact positif ou négatif, important ou faible. De façon générale, ces méthodes, selon la formule réaliste classique, permettent de passer de la question « est-ce que ça marche ? » à la question « pour quoi est-ce que ça marche et dans quelles circonstances² ? ».

Ces deux éclairages peuvent en outre avoir une portée politique et opérationnelle cruciale, en aidant à mieux cibler les interventions publiques ou à les ajuster. En outre, aucune méthode ne peut prétendre pouvoir appréhender la causalité de façon univoque et exacte. L'approche contrefactuelle peut toujours être critiquée dès lors qu'il est impossible de former deux groupes (un de contrôle et un de traitement) parfaitement identiques. La *validité interne* sera ainsi toujours limitée, et il s'agira de la renforcer en cherchant à former deux groupes les plus similaires possibles, selon des critères de similarité toujours discutables et donc à discuter. D'un autre côté, l'approche processuelle peut toujours être critiquée aussi, car il ne suffit pas d'observer un enchaînement d'étapes et d'actions pour voir des liens de causalité. La *validité interne* dépendra donc largement de la qualité des

¹ Cette conception prévaut par exemple dans l'évaluation réaliste, où la causalité est traduite par la formule « mécanisme + contexte = effet » (*mechanism + context = outcome*). Pour comprendre comment une intervention produit des effets, il faut « comprendre ce que le programme fait effectivement pour modifier les comportements, et pourquoi ce processus spécifique ne s'enclenche pas automatiquement dans toutes les situations » : Pawson R. et Tilley N. (1997), « An introduction to scientific realist evaluation », in Chelimsky E. et Shadish W. R. (dir.), *Evaluation for the 21st Century: A Handbook*, Los Angeles, Sage, p. 405-418. Pour y parvenir, l'intervention et son contexte doivent être décomposés en leurs composantes clés, et chacune de ces composantes doit faire l'objet d'une investigation à part entière : voir Robert É. et Ridde V. (2013), « L'approche réaliste pour l'évaluation de programmes et la revue systématique : de la théorie à la pratique », *Mesure et évaluation en éducation*, vol. 36(3), p. 79-108.

² Pawson R. et Tilley N. (1997), « An introduction to scientific realist evaluation », *op. cit.*

fondements théoriques qui étayent ces liens de causalité. Le croisement de ces deux approches, qui reposent sur des comparaisons ou des séquençements, doit précisément aspirer à renforcer la validité interne de l'évaluation d'impact (par *triangulation*, nous y reviendrons plus tard).

2. Les méthodes quantitatives proposent de traiter un nombre limité de variables sur un grand nombre de cas là où les méthodes qualitatives s'attachent à produire une connaissance riche sur un nombre limité de cas

Les méthodes quantitatives et les méthodes qualitatives sont usuellement distinguées aussi par la densité et l'étendue de l'information qu'elles permettent de recueillir. Alors que les méthodes qualitatives fournissent généralement une information riche et dense sur un petit nombre de cas, les méthodes quantitatives donnent accès à une information plus limitée sur un grand nombre de cas. Cette distinction vaut en particulier pour les méthodes mobilisées dans le cadre de l'évaluation d'impact des politiques publiques.

D'un côté, les méthodes quantitatives reposent, par définition, sur des indicateurs chiffrés. Les individus, par exemple, pourront être caractérisés par leur âge (0 an, 1 an, 2 ans, etc.), l'évolution de leur pouvoir d'achat d'une année à l'autre (+/-x %), leur niveau de satisfaction (-2 = très insatisfait, -1 = plutôt insatisfait, etc.) ou encore leur catégorie socioprofessionnelle (0 = ouvriers, 1 = cadres, etc.)¹. Tous ces chiffres constituent une forme de codification de l'information, et se prêtent à des analyses statistiques mêlant agrégations et comparaisons. Toutefois, aussi nombreuses soient-ils, ces indicateurs (ou *variables* ou *dimensions*) constituent une réduction de la réalité. Un individu est toujours plus que la compilation d'informations chiffrées sur son âge, son pouvoir d'achat, sa catégorie socioprofessionnelle, etc. De même, une économie est toujours plus que la compilation d'informations chiffrées sur son PIB, sa balance commerciale, son taux de prélèvements sociaux, etc. Une telle réduction présente l'inconvénient d'écarter un grand nombre (pour ne pas dire une infinité) d'informations sur les individus, les situations et les phénomènes tels qu'ils existent et se déploient. Mais elle présente l'avantage de pouvoir s'exercer sur un grand nombre de cas. Par exemple, il est techniquement possible de relever l'âge de millions d'individus. Les méthodes quantitatives chercheront à tirer profit de l'importance des échantillons observés et caractérisés pour tirer des analyses statistiquement robustes sur ces

¹ Ces exemples illustrent les quatre logiques d'interprétation des variables quantitatives : l'âge est une variable dite *de ratio* (chaque âge a un sens en soi) ; l'évolution du pouvoir d'achat est une variable *d'intervalle* (chaque évolution a un sens *relativement à une référence*) ; le niveau de satisfaction est une variable *ordinaire* (seule l'ordre des chiffres a un sens, le niveau 2 étant par exemple supérieur au niveau 1) ; enfin la catégorie socioprofessionnelle est une variable nominale (le chiffre est une simple étiquette rattachée à une catégorie).

caractéristiques. Nous y reviendrons plus tard, à propos de la généralisation des résultats. En outre, les méthodes quantitatives modélisent le plus souvent les relations entre ces indicateurs de façon très stylisée, par des fonctions mathématiques prédéfinies (une relation linéaire, une relation exponentielle, etc.). Ces modélisations (*spécifications*) n'étant jamais l'exact reflet des relations qui opèrent dans la réalité, les écarts entre l'observation et la modalisation sont mesurés par ce que l'on appelle des *résidus*.

D'un autre côté, les méthodes qualitatives se proposent d'aborder les individus, les situations et les phénomènes, non pas en les réduisant sur des indicateurs chiffrés, mais en mobilisant des mots (concepts) et des structures de mots formant des descriptions et des raisonnements. Quelques chiffres peuvent être convoqués, mais seulement en supplément. La liberté offerte par le langage¹ et la richesse de ses combinaisons permettent d'aborder la réalité de façon plus fine, plus riche, plus complexe. En particulier, les méthodes qualitatives peuvent mieux aborder des réalités se prêtant difficilement à la quantification (et à la notion d'ordre qu'elle implique), par exemple pour qualifier un ressenti subjectif, une trajectoire de vie ou encore des valeurs morales. Cet avantage a aussi son inconvénient, en miroir des méthodes quantitatives : si l'approche qualitative peut extraire une plus grande richesse de données d'un objet d'étude donné, elle ne peut raisonnablement répéter cet exercice sur un grand nombre de cas. Ainsi, et pour des raisons de temps et de ressources, les échantillons étudiés par les méthodes qualitatives sont généralement de petite taille. En outre, les données recueillies sont généralement disparates, et font difficilement l'objet d'agrégation et de comparaison. Il est, par exemple, moins aisé de comparer deux résumés de livres (une information littéraire) que de comparer leurs prix (une information chiffrée). Cette limitation justifie d'autant que les échantillons soient de petite taille. Ainsi, l'étude monographique, qui est une méthode qualitative par excellence, se définit comme une *étude approfondie, exhaustive, portant sur un sujet précis et limité*.

Les deux méthodes présentent des différences opposées. Pour les méthodes qualitatives, la densité de l'information recueillie est généralement importante et l'étendue de l'échantillon étudié généralement petit. Pour les méthodes quantitatives, c'est l'inverse. De cette différence de fond découle une seconde, relative à la *significativité* des résultats d'une évaluation d'impact. Cet enjeu suscite une réflexion transversale tant du côté des méthodes quantitatives que du côté des méthodes qualitatives².

La significativité d'un résultat, d'abord, renvoie à sa justesse pour l'ensemble de la population étudiée. Ainsi par exemple, s'il s'agit d'estimer le salaire moyen des 20 millions de salariés d'un pays, le résultat de cette estimation sera parfaitement significatif s'il tient

¹ Qui est beaucoup plus large que l'ensemble des fonctions mathématiques que mobilisent généralement les modèles quantitatifs. Le langage mathématique peut être vu comme un sous-ensemble du langage plus général.

² Mahoney J. et Goertz G. (2006), « *A tale of two cultures...* », *op. cit.*

compte du salaire de chacun des salariés, sans exception. L'estimation sera un calcul juste, par construction. À l'inverse, si l'on se contente de relever le salaire de seulement dix personnes tirées au hasard, il est fort à parier que le salaire moyen estimé ne sera pas significatif. Par un *accident de mesure*, les dix personnes interrogées pourraient en effet avoir un salaire particulièrement faible, ou particulièrement fort, et ainsi biaiser l'estimation du salaire moyen. En outre, en interrogeant dix autres personnes tirées au sort, il y a de fortes chances que l'estimation donne à chaque fois un résultat très variable. De même, si l'on se contente de relever le salaire des cadres, il est fort à parier que le salaire moyen estimé ne sera pas significatif non plus, du fait d'un *biais de sélection*. Fussent-ils des millions, le salaire des cadres n'est en effet pas représentatif de l'ensemble des salaires des salariés, ce qui biaiserait donc à nouveau le résultat. En définitive, il ressort qu'un résultat est significatif s'il vaut pour l'ensemble de la population étudiée, et pour ce faire : ou bien on étudie l'ensemble de la population (à la façon d'un recensement), ou bien on étudie un échantillon de taille suffisante pour lui être raisonnablement *représentatif*. Dans le second cas, un mode de tirage adapté doit être appliqué, par exemple le tirage aléatoire ou la méthode des quotas, afin d'éviter les biais de sélection.

Que ce soit donc avec une méthode quantitative ou avec une méthode qualitative, les résultats d'une évaluation sont parfaitement significatifs s'ils reposent sur l'observation de l'ensemble de la population d'intérêt. Par exemple, s'il s'agit d'évaluer l'impact d'une réforme sur une population comprenant au total 40 entreprises, une étude économétrique ou des entretiens sociologiques produiront des résultats parfaitement significatifs s'ils reposent sur l'observation et l'étude des 40 entreprises, sans exception.

Les difficultés émergent lorsque la population d'intérêt, c'est-à-dire la population sur laquelle il est question d'établir des résultats d'évaluation, présente une taille importante. Par exemple, s'il s'agit d'évaluer l'impact d'une réforme sur une population totale de 2 millions d'entreprises, il est non seulement impossible de mobiliser une méthode qualitative (un entretien par exemple) pour chacune d'elles, mais il est également fort probable qu'il faille se contenter d'un échantillon de petite taille (quelques dizaines d'entreprises). L'échantillon interrogé se retrouve *de facto* avec une faible représentativité, et les résultats des méthodes qualitatives auront une faible significativité. Ils seront donc difficilement extrapolables à l'ensemble des 2 millions d'entreprises : on dira que leur validité externe est faible. Cela se traduira par de fortes marges d'erreur dans les intervalles de confiance, ou par des intervalles de faible confiance (voir *supra*).

Ceci étant dit, il est à noter que les méthodes qualitatives, dès lors qu'elles appréhendent chaque situation de façon riche et précise, peuvent aussi offrir des possibilités d'extrapolation des résultats qu'elles produisent. Là où les méthodes quantitatives aspirent à des extrapolations sur des populations d'ensemble et sont impuissantes pour des extrapolations sur des cas précis, les méthodes qualitatives peuvent se distinguer en permettant exactement le contraire. Ainsi, pour apprécier l'impact d'une réforme sur une population de

2 millions d'entreprises, une méthode quantitative cherchera à observer un échantillon représentatif de la population d'ensemble, et à extrapoler l'impact observé sur l'échantillon à l'ensemble de la population. Même s'il est significatif, cet impact reste un impact *moyen* : il vaut en moyenne pour l'ensemble de la population – ou pour un sous-groupe de la population – mais ne peut être associé à une entreprise en particulier, du fait que chaque entreprise a des caractéristiques spécifiques, et en particulier des caractéristiques inobservables (par exemple, la nature des relations interpersonnelles entre les salariés). À l'inverse, une méthode qualitative caractérisera précisément une poignée d'entreprises (le chiffre d'affaires, la taille, la nature des relations interpersonnelles, etc.). Les résultats observés sur ces entreprises, s'ils sont difficilement extrapolables à l'ensemble de la population du fait de la petitesse de l'échantillon observé, pourraient ici à l'inverse être raisonnablement extrapolés à toute autre entreprise présentant le plus précisément possible les mêmes caractéristiques que l'une des entreprises observées. En d'autres termes, si deux entreprises partagent un grand nombre de caractéristiques (ce qu'une méthode qualitative vise précisément à observer), alors il est fort à parier que l'impact observé chez l'une soit extrapolable chez l'autre. Cette extrapolation *ciblée*, et non *en moyenne*, peut être très utile lorsqu'il s'agit d'appréhender des cas particuliers.

En définitive, là où les méthodes quantitatives relèvent généralement une poignée d'indicateurs sur un grand nombre d'observations, et peuvent ainsi formuler des résultats significatifs, les méthodes qualitatives peuvent – en complément – avoir deux plus-values principales :

- elles peuvent apporter un éclairage approfondi sur des cas particuliers, qui le cas échéant peut illustrer l'hétérogénéité des situations individuelles¹ ;
- elles peuvent permettre de collecter rapidement quelques premières informations, quitte à ne porter que sur des cas particuliers² (par exemple, un entretien peut être à brève échéance plus facile à réaliser qu'une grande enquête statistique).

¹ En effet, les méthodes quantitatives ont tendance à produire des résultats *en moyenne*. Ainsi par exemple, une évaluation pourrait conclure à un impact de +100 euros par ménage, *en moyenne*. Et ce résultat pourrait être fortement significatif, c'est-à-dire qu'il pourrait *en moyenne* être extrapolé à l'ensemble des ménages, au-delà de l'échantillon considéré par l'évaluateur. Mais il se pourrait que l'impact soit en réalité très variable selon les catégories de ménages (+500 euros, +20 euros, -300 euros, etc.), ce que des observations de terrain – et donc de méthodes qualitatives – pourraient aider à percevoir et à comprendre au cas par cas. Bien entendu, les méthodes quantitatives s'intéressent également à l'hétérogénéité des situations ; cette hétérogénéité est généralement appréhendée à travers des statistiques de dispersion comme la variance, l'écart-type ou encore de coefficient de variation, qui reviennent à calculer un écart *moyen* à la moyenne. Elles peuvent aussi distinguer les impacts au croisement de certaines caractéristiques (sexe, âge, région, etc.). Il reste que les impacts ainsi calculés restent des impacts moyens et non des impacts individuels : l'impact moyen auprès des femmes, l'impact moyen dans la région des Hauts-de-France, etc.

² En outre, les méta-analyses de travaux qualitatifs peuvent aider à augmenter la significativité de leurs résultats, en concaténant et en comparant leurs résultats respectifs. Voir par exemple Pawson R. (2004), « Would Campbell be a member of the Campbell collaboration? », *The Evaluator*, vol. 13, hiver.

3. Les méthodes quantitatives reposent surtout sur des raisonnements hypothético-déductifs là où les méthodes qualitatives mobilisent davantage le raisonnement inductif

Sur le plan épistémologique, les méthodes quantitatives et les méthodes qualitatives sont usuellement distinguées aussi au regard des modalités mêmes de l'élaboration théorique¹ :

- d'un côté, les approches qualitatives reposent plus facilement, bien que non exclusivement, sur des formes de raisonnements inductifs : à partir de données empiriques, et sans hypothèse préalable, il s'agit en principe d'élaborer une théorie ;
- d'un autre côté, les méthodes quantitatives sont plus fréquemment associées à des raisonnements hypothético-déductifs : à partir d'analyses préexistantes, il s'agit de formuler une hypothèse puis d'en tester la vraisemblance (confirmation/infirmer) à partir d'un protocole d'enquête empirique (une expérience, des observations, etc.).

Pour être plus exact, des hypothèses sont bel et bien formulées dans la plupart des enquêtes par méthodes qualitatives. Elles peuvent même être implicites et relever de présupposés théoriques qu'il sera essentiel d'explicitier. Mais là où le raisonnement hypothético-déductif isole l'élaboration des hypothèses et la réalisation de l'enquête comme deux phases distinctes et successives, les méthodes qualitatives favorisent généralement le va-et-vient entre ces deux étapes, entre la réflexion théorique et l'enquête de terrain. Cette démarche dite de *théorie ancrée*² consiste à rechercher sur le terrain des *cas négatifs* (ou *cas déviants*) pouvant contredire ces premières hypothèses, afin de les réviser de façon itérative. Au moment où le chercheur ne trouve plus de cas négatifs, la dernière hypothèse formulée est dite *saturée* : pour l'ensemble des cas étudiés auxquels elle s'applique, l'hypothèse pourra être considérée comme une *théorie*³. Par exemple, un évaluateur pourrait à partir d'entretiens avec dix entreprises faire l'hypothèse qu'un certain crédit d'impôt a contribué à y créer des emplois. À la recherche de cas négatifs, l'évaluateur peut chercher à réaliser des entretiens avec des entreprises issues de secteurs plus divers ou de taille différente, etc. Et à mesure des observations, il est amené à réviser son hypothèse initiale jusqu'à aboutir à une hypothèse qu'aucune observation ne contredise (jusqu'à preuve du contraire). Par exemple, l'hypothèse saturée pourrait en

¹ White H. (2013), « The use of mixed methods in randomized control trials », *New Directions for Evaluation*, vol. 2013(138), juin, p. 61-73.

² Ou *grounded theory* en anglais. Voir Glaser B. G. et Strauss A. (2017), *La Découverte de la théorie ancrée. Stratégies pour la recherche qualitative*, Paris, Armand Colin ; Strauss A. et Corbin J. (2003), « L'analyse de données selon la *grounded theory*. Procédures de codage et critères d'évaluation », in Cefaï D. (dir.), *L'Enquête de terrain*, Paris, La Découverte, p. 363-379.

³ Et ce tant qu'aucun autre cas négatif n'est à observer.

définitive être : « le crédit d'impôt aide à créer des emplois dans les entreprises de petite et moyenne tailles, et uniquement dans les secteurs manufacturiers hors Île-de-France¹ ».

À nouveau, il s'agit ici d'une distinction courante entre deux cultures évaluatives. Des contre-exemples existent. À l'aune des données massives, certaines démarches quantitatives reposent également sur un va-et-vient plus ou moins systématique entre le terrain et la théorie, selon une démarche inductive ou inférentielle. Dans le cadre du *machine learning* – ou du *deep learning*² – par exemple, les algorithmes vont en effet s'exercer et évoluer au contact des données, de façon plus ou moins itérative, et à partir d'hypothèses initiales plus ou moins explicitées.

En définitive, là où les méthodes quantitatives ont tendance à formuler une hypothèse et à porter ensuite l'évaluation sur le test de cette hypothèse, les méthodes qualitatives peuvent – en complément – avoir quatre plus-values principales.

- Elles peuvent aider à formuler des hypothèses plus pertinentes, que des méthodes quantitatives peuvent ensuite tester. Sous forme interrogative, ces hypothèses peuvent constituer des questions évaluatives également plus pertinentes, et qui n'ont pas nécessairement à se confondre avec l'intention politique de l'intervention ou le souhait du commanditaire de l'évaluation. Par exemple, dans le cadre d'une évaluation d'une politique de lutte contre le tabagisme, des entretiens avec des buralistes pourraient faire apparaître l'opportunité d'évaluer l'impact de la politique sur le chiffre d'affaires des buralistes, et leurs conséquences négatives sur la fermeture de leurs commerces et l'emploi³.
- Attentives à la diversité des terrains et à la complexité des phénomènes, elles peuvent aider les méthodes quantitatives à identifier les variables de contrôle les plus pertinentes. Sur l'exemple du crédit d'impôt, elles pourraient suggérer de tenir compte dans l'estimation économétrique de la taille et du secteur des entreprises mais aussi de facteurs moins intuitifs comme la gouvernance de l'entreprise ou sa culture organisationnelle (à condition de pouvoir quantifier ces facteurs ensuite). De façon

¹ Les méthodes quantitatives peuvent aussi aboutir à ce genre de résultats, mais sans procéder aussi systématiquement à ce va-et-vient entre le terrain et l'analyse : dans l'approche hypothético-déductive, il s'agirait plutôt de faire l'hypothèse que certains facteurs conditionnent l'impact du crédit d'impôt puis de tester cette hypothèse à travers une estimation économétrique ou des statistiques descriptives.

² Il s'agit d'algorithmes capables de mimer les actions du cerveau humain grâce à des réseaux de neurones. Les algorithmes de *machine learning* traitent des données quantitatives et structurées (des valeurs numériques), lorsque ceux de *deep learning* traitent des données non structurées comme le son, le texte ou l'image.

³ Les évaluations qui se donnent la possibilité de faire évoluer leurs questions évaluatives sont parfois qualifiées de *non finalisées* (*goal-free evaluation* en anglais). Voir par exemple : Patton M. Q. (2004), *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*, Londres, Sage ; Chanut V. (2007), *Misère de l'évaluation en ressources humaines*, Paris, Eska ; Divay S. (2012), « [Petit précis méthodologique d'évaluation qualitative](#) », *Éducation et formations*, mars, vol. 81, p. 49-54 ; Scriven M. (1972), « *Prose and cons about goal-free evaluation* », *Evaluation Comment*, vol. 4(3), p. 1-7.

corollaire, elles peuvent aider à repérer des raisonnements quantitatifs incongrus. Par exemple, il est possible grâce à une démarche statistique techniquement juste qui s'appelle le *p-hacking*¹ de produire des résultats statistiques fortement significatifs mais sans aucun fondement théorique valable, ce qu'une méthode qualitative peut aider à repérer. De même, il peut arriver, grâce à un découpage astucieux de la population en sous-catégories, de transformer des corrélations positives en des corrélations négatives ou inversement (il s'agit du paradoxe de Simpson²).

- Dans une démarche compréhensive, les méthodes qualitatives peuvent éclairer les méthodes quantitatives sur la « réception » de l'intervention publique par les acteurs ciblés³, qu'il s'agisse de leurs représentations, de leurs motivations à se conformer ou non à l'intervention, de leurs perceptions de l'intervention et de ses impacts, etc. Ainsi par exemple, une politique pourrait ne pas avoir l'impact escompté en raison de sa faible acceptabilité ou de sa faible compréhension par les publics ciblés. La prise en compte de la subjectivité des personnes concernées par une politique publique peut aider à renforcer la pertinence des questions évaluatives et à augmenter l'appropriation des résultats évaluatifs par les parties prenantes.
- De même, l'examen par l'évaluateur de sa propre subjectivité lui permet d'engager un exercice réflexif sur la conduite même de l'évaluation, qu'il s'agisse de la formulation de la question de recherche, du choix des indicateurs⁴ et des outils d'analyse, de la caractérisation des objets étudiés ou encore de l'interprétation des résultats. Par exemple, un évaluateur qui s'intéresse à la pauvreté gagnerait à interroger le concept même de pauvreté, et les biais que les définitions usuelles de la pauvreté, ou que sa propre définition, peut générer dans les résultats de l'analyse. Dans le même exercice réflexif, l'évaluateur peut aussi être attentif aux biais cognitifs qui peuvent altérer son jugement, à l'instar du biais de confirmation. La prise en considération de la subjectivité – chez les sujets évalués comme chez l'évaluateur – ne signifie donc pas un abandon de l'aspiration à l'objectivité, dont elle est au contraire une des conditions de possibilité.

¹ La technique du *p-hacking* consiste à considérer une population pour laquelle de nombreuses caractéristiques sont disponibles, et à trouver à quelques individus deux caractéristiques partagées qu'aucun des autres individus du groupe ne possède. La corrélation entre les deux caractéristiques s'en trouvera mécaniquement très significative, sans qu'il y ait forcément un fondement théorique valable.

² Voir Berger Q. et Caravenna F. (2021), « [Le paradoxe de Simpson illustré par des données de vaccination contre le Covid-19](#) », *The Conversation*, 3 novembre.

³ Revillard A. (2018), « [Saisir les conséquences d'une politique à partir de ses ressortissants : la réception de l'action publique](#) », *Revue française de science politique*, vol. 68(3), p. 469-492.

⁴ Le courant constructiviste affirme que nos concepts ne sont pas le simple reflet de la réalité mais une construction de l'esprit humain. Ainsi il n'existe pas un indicateur de la pauvreté qui représente *la pauvreté* mais seulement plusieurs façons de définir la pauvreté, et des conventions techniques et sociales qui établissent des indicateurs possibles de pauvreté.

4. Le développement de méthodes mixtes

Les méthodes mixtes¹ se développent rapidement depuis les années 2000 en sciences humaines et sociales. Elles sont employées dans des disciplines aussi variées que les sciences d'éducation, la psychologie, la santé publique, les sciences de gestion ou encore la sociologie des organisations. Des revues académiques spécialisées émergent, à l'instar du *Journal of Mixed Methods Research* et de l'*International Journal of Multiple Research Approaches*. Plus récemment, les méthodes mixtes se sont étendues à l'évaluation des politiques publiques, au terme d'un rapprochement progressif entre les approches quantitatives et les approches qualitatives².

Dans un premier temps, ces approches se sont opposées sur les plans méthodologique et épistémologique. Le clivage a été renforcé par la prolifération et la complexification des méthodes qui ont mené à une forte spécialisation des chercheurs. Selon les uns, seules les méthodes quantitatives sont aptes à produire des résultats rigoureux sur l'impact des politiques publiques, en termes de significativité notamment. Pour les autres, cette prétention hégémonique est infondée, du fait des nombreuses difficultés et défaillances que présentent en pratique les méthodes quantitatives, notamment pour approcher des causalités complexes ou pour tenir compte du contexte de mise en œuvre d'une politique publique. De vifs débats ont lieu, comme à l'occasion de la publication de l'ouvrage de King *et al.* en 1994³, où les auteurs enjoignent les *qualitativistes* à se conformer aux méthodes quantitatives, présentées comme supérieures.

Si les méthodes quantitatives continuent de jouir d'une forme d'aura scientifique en évaluation des politiques publiques⁴, un rapprochement avec les méthodes qualitatives s'est opéré au tournant des années 2000. Portées par une émulation académique, les méthodes qualitatives gagnent en légitimité. Dans un ouvrage de 2006⁵ qui fait encore référence, Mahoney et Goertz font état de la pluralité des méthodes : du fait même de ses spécificités, chaque méthode doit pouvoir être privilégiée selon la nature de l'*evaluandum* (l'objet évalué), et selon le contexte et les modalités de l'évaluation. Ainsi par exemple, une méthode qualitative peut être privilégiée, et une méthode quantitative écartée, si les données statistiques manquent, ou si le commanditaire de l'évaluation attend des

¹ *Mixed methods* ou *multimethods* en anglais.

² Aguilera T. et Chevalier T. (2021), « [Les méthodes mixtes pour la science politique. Apports, limites et propositions de stratégies de recherche](#) », *Revue française de science politique*, vol. 71, p. 365-389.

³ King G., Keohane R. O et Verba S. (1994), *Designing Social Inquiry. Scientific Inference in Qualitative Research*, Princeton, Princeton University Press.

⁴ Il suffit de penser à la notoriété de l'évaluation par expérimentation aléatoire contrôlée (*randomized controlled trial*) et à sa reconnaissance notamment via les travaux d'Esther Duflo, prix Nobel d'économie en 2019.

⁵ Mahoney J. et Goertz G. (2006), « [A tale of two cultures...](#) », *op. cit.*, p. 229.

explications sur l'échec de l'intervention dans un contexte particulier. Avant de se lancer dans une évaluation et de prétendre pouvoir appliquer une recette méthodologique, il est essentiel de se prêter à un exercice de *research design*, qui consiste à identifier ou à établir le cadre méthodologique le plus adapté¹.

C'est dans ce contexte que se développent les premières méthodes mixtes. Il ne s'agit plus seulement de reconnaître la légitimité des différentes méthodes et leurs avantages respectifs, mais de véritablement les combiner pour répondre à une même question évaluative, et plus généralement pour alimenter un même travail de recherche². Au sein d'une méthode mixte *coordonnée*, les méthodes quantitatives et qualitatives évoluent en parallèle, et leurs résultats sont confrontés à la fin de l'étude. Ce serait par exemple le cas d'une évaluation qui associe une régression économétrique et des entretiens pour apprécier et comparer le pouvoir d'achat et la perception du pouvoir d'achat. Dans une méthode mixte *intégrée*, elles s'alternent tout au long de l'étude, que ce soit au moment de la conception du protocole, de la collecte des données, de leur traitement ou encore de l'interprétation des résultats et de leur diffusion. Quelques-unes des méthodes mixtes intégrées poussent cette *intégration* au point d'aboutir à une méthodologie à part entière et d'avoir leur propre dénomination, à l'instar de la *méthode d'analyse qualitative comparée* que nous présenterons plus tard.

Les *qualitativistes* sont à cet égard les premiers à souligner la plus-value des approches mixtes auprès des quantitativistes. À partir d'une revue de littérature, Aguilera et Chevalier (2021) recensent huit principales stratégies de recherche en méthodes mixtes. D'un côté, les méthodes qualitatives peuvent fournir aux méthodes quantitatives des indices suggérant une théorie plus générale (stratégie de *confirmation*), des exemples approfondis pour illustrer des théories déjà établies (stratégie d'*enrichissement*) ou des contre-exemples³ pour suggérer une révision de la théorie (stratégie de *complexification*). De l'autre côté, les méthodes quantitatives peuvent fournir aux méthodes qualitatives les moyens d'augmenter la significativité des résultats (stratégie de *systématisation*) ou de les extrapoler à de nouvelles situations (stratégie de *généralisation*). Ensemble, les deux types de méthodes peuvent également s'alimenter sur le plan méthodologique⁴ (stratégie de *développement*) et la confrontation de leurs résultats peut servir à corroborer – ou non –

¹ Parmi les tentatives de rapprochement des deux approches, qualitative et quantitative, voir notamment : Donaldson S. I., Christie C. A. et Mark M. M. (dir.) (2009), *What Counts as Credible Evidence in Applied Research and Evaluation Practice?*, Los Angeles, Sage.

² Le terme anglais *mixed methods* souligne littéralement l'idée d'un mélange (*mixing*) des méthodes.

³ L'identification d'un cas déviant peut indiquer que la théorie s'écarte trop de certains types de situations. Elle peut aussi indiquer une erreur de mesure ou une mauvaise interprétation des chiffres.

⁴ Par exemple, une observation du terrain peut aider à identifier les variables pertinentes pour une estimation par doubles différences.

un résultat (stratégie de *triangulation*) ou à éclairer des questions évaluatives connexes (stratégie de *complémentarité*).

De fait, les méthodes mixtes sont de plus en plus courantes dans l'évaluation des politiques publiques, auprès des chercheurs mais aussi des acteurs institutionnels et des cabinets privés. En France, et de façon variable selon les producteurs des évaluations de politiques publiques, les méthodes mixtes sont d'ores et déjà présentes dans près de 40 % à 60 % des travaux d'après les chiffres de l'Observatoire de la Société française de l'évaluation¹.

Aussi, et face à ce développement, plusieurs chercheurs soulignent-ils des points de vigilance. D'abord, les méthodes mixtes exigent d'identifier les points de jonction (par exemple une variable d'analyse en commun) entre les différentes approches et d'articuler rigoureusement leurs méthodes et résultats². Il faut en outre rester attentif à leurs différences épistémologiques, par exemple lorsqu'une méthode qualitative et une méthode quantitative relèvent de deux conceptions différentes de la causalité. Ensuite, les méthodes mixtes exigent des équipes transdisciplinaires ou des experts à la fois en méthodes quantitatives et qualitatives³. Autrement, les travaux pourraient perdre en robustesse scientifique tout en laissant l'illusion de leur richesse⁴. À cet égard, certains auteurs regrettent l'utilisation cosmétique des méthodes qualitatives aux côtés des méthodes quantitatives⁵, ou la complexification excessive des résultats de l'évaluation⁶. En outre, la combinaison des méthodes peut représenter un défi organisationnel, de par les coûts supérieurs qu'elle pourrait engager en temps, en moyens humains ou encore en ressources financières. En tout état de cause, ces coûts doivent être appréciés au regard des bénéfices de l'évaluation, notamment en termes de qualité scientifique et d'utilité opérationnelle.

¹ Société française de l'évaluation (2018), *Baromètre 2018 de l'évaluation des politiques publiques en France*. Il est à noter que les méthodes quantitatives recensées par la SFE ne reposent pas nécessairement sur un raisonnement contrefactuel et peuvent se limiter à des statistiques descriptives et à des calculs de corrélation.

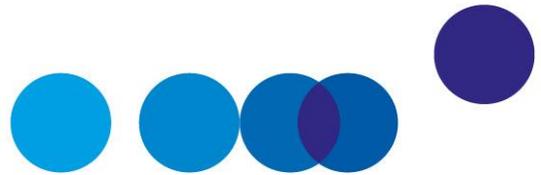
² Burch P. et Heinrich C. J. (2016), *Mixed Methods for Policy Research and Program Evaluation*, Los Angeles, Sage.

³ Voir Small M. L. (2011), « How to conduct a mixed methods study: Recent trends in a rapidly growing literature », *Annual Review of Sociology*, vol. 37, août, p. 57-86 ; Tashakkori A. et Teddlie C. (2010), *SAGE Handbook of Mixed Methods in Social and Behavioral Research*, Los Angeles, Sage.

⁴ Ahmed A. et Sil R. (2012), « When multi-method research subverts methodological pluralism. Or, why we still need single-method research », *Perspectives on Politics*, vol. 10(4), décembre, p. 935-953.

⁵ White H. (2013), « The use of mixed methods in randomized control trials », *op. cit.*

⁶ Greene J. C., Benjamin L. et Goodyear L. (2001), « [The merits of mixing methods in evaluation](#) », *Evaluation*, vol. 7(1), janvier, p. 25-44.



CHAPITRE 2

QUELQUES EXEMPLES EMBLÉMATIQUES DE MÉTHODES QUALITATIVES ET MIXTES

On trouvera ici quelques-unes des méthodes qualitatives et mixtes les plus emblématiques dans le champ de l'évaluation de l'impact des politiques publiques, au regard de leur utilisation par les évaluateurs ou de la spécificité de leurs apports théoriques. Nous présentons trois méthodes qualitatives d'évaluation d'impact – la *cartographie conceptuelle d'impacts*, l'*analyse de contribution* et la *reconstitution de processus* – puis deux méthodes mixtes, l'*analyse qualitative comparée* et la *méthode d'évaluation faisabilité-impact*. Sans constituer une boîte à outils exhaustive, ces méthodes permettent d'illustrer la diversité des angles méthodologiques et les possibilités de combinaisons, et de couvrir les principales questions évaluatives. En outre, du fait de leur foisonnement, conjugué à une certaine inflation sémantique¹, il serait difficile de faire la liste exhaustive des méthodes mixtes mobilisables pour l'évaluation d'impact des politiques publiques.

À chaque fois, nous présentons de façon synthétique les principaux éléments de méthodes et un exemple d'application. En raison de la variété des approches, des modalités et des usages, nous choisissons de nous concentrer sur ce qui nous semble constituer le socle le plus robuste et le plus opérationnel. Dans un souci de pédagogie, nous passons sous silence les nombreuses subtilités, mais nous indiquons les références académiques pour aller plus loin, ainsi que les sources des exemples cités.

Ces cinq méthodes peuvent être mobilisées sur des questions très diverses en sciences sociales, mais nous cherchons ici à illustrer leur application et leurs apports dans le cadre de l'évaluation de l'impact des politiques publiques. Aussi donnerons-nous ensuite deux exemples supplémentaires reposant sur des méthodes mixtes : la première évaluation – sur une

¹ Greene J. C., Benjamin L. et Goodyear L. (2001), « [The merits of mixing methods in evaluation](#) », *Evaluation*, vol. 7(1), janvier, p. 25-44. Par exemple, la *most significant change technique*, l'*outcome harvesting* et l'*outcome mapping* (voir *infra*) sont des méthodes très proches. De même, la méthode de *process tracing* est renvoyée à des terminologies variées dans la littérature académique : « vérification de processus », « analyse systématique de processus », « méthode de congruence », « analyse de congruence », etc.

expérimentation en matière de compétences langagières¹ – illustre les bénéfices des méthodes mixtes à la fois pour la conception, la mise en œuvre et l'évaluation d'une intervention publique ; la seconde évaluation – sur les politiques territoriales de protection de l'eau² – illustre les possibilités de combinaisons de méthodes pour approcher tout critère évaluatif.

Mais avant de détailler les cinq méthodes, nous commençons par présenter deux outils qui interviennent de façon transversale : l'*étude de cas*³ et le *diagramme logique des impacts*, un outil mobilisé par la plupart de ces méthodes mixtes et qui permet d'établir la théorie d'action de l'intervention publique.

1. Outil transversal n° 1 : l'étude de cas

En deux mots – L'étude de cas consiste à réaliser une étude approfondie sur un sujet précis et limité. Empirique, la méthode repose principalement sur des entretiens, de l'observation du terrain et de l'analyse de matériaux (écrits, photos, etc.). Il s'agit de représenter les nuances et la complexité d'un comportement, d'un phénomène ou d'une situation. La monographie est le nom que prend la restitution – généralement encyclopédique – des données collectées et leur mise en perspective.

Quel apport pour l'évaluation d'impact ? Même si elle porte *par définition* sur une situation unique, l'étude de cas peut fournir un faisceau d'indices sur l'existence d'un impact, son amplitude et ses ressorts. Par exemple, une étude de cas pourrait faire état d'une nouvelle réglementation qui interdit certaines pratiques industrielles polluantes, observer sur le terrain une disparition desdites pratiques, et mesurer une chute de la pollution d'une rivière quelques semaines plus tard. Au regard de la nature des événements et de leur enchaînement, l'étude de cas pourra raisonnablement inférer un lien de cause à effet entre la nouvelle réglementation et la chute de la pollution. À plus forte raison, la comparaison de plusieurs études de cas peut renforcer l'inférence causale, et c'est pourquoi les études de cas sont régulièrement mobilisées par des méthodes d'évaluation d'impact, comme nous le verrons ci-après. Elle permet en effet de collecter des données qui aident à formuler une hypothèse (des indices), à conforter une théorie (des preuves) ou au contraire à la réfuter (des contre-exemples).

¹ Pin C. et Barone C. (2021), « [L'apport des méthodes mixtes à l'évaluation. Une expérimentation randomisée dans le champ des politiques éducatives](#) », *Revue française de science politique*, vol. 71, p. 391-412.

² Hourcade R. et Cortinas Muñoz J. (2021), « [La mise en œuvre des politiques de protection de l'eau. Méthode et apports d'une recherche combinant QCA et process tracing](#) », *Revue française de science politique*, vol. 71, p. 413-435.

³ Les autres méthodes qualitatives pour l'évaluation de l'impact sont déjà présentées dans le guide de 2016. Voir Desplatz R. et Ferracci M. (2016), [Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et des praticiens](#), France Stratégie, septembre.

Exemple – Dans un rapport de 2018¹, l'Inspection générale des finances et le Conseil général de l'environnement et du développement durable (CGEDD) dressent un état des lieux des politiques de l'eau et de la biodiversité. À partir de l'analyse de matériaux divers (la réglementation, des bilans d'activités, etc.) et de divers entretiens, ils mettent notamment en lumière l'inadéquation des moyens au regard des objectifs visés. Par exemple, des entretiens ont été conduits auprès de directeurs de parcs nationaux au sujet des missions de police de l'environnement. Ont ainsi été mises en lumière des difficultés à la fois précises et pratiques, notamment les difficultés perçues par les agents à exercer une mission de police, généralement sans armes de service, et dans un contexte où la présence de touristes et de la population locale exige beaucoup de « doigté ».

Pour aller plus loin – Yin R. K. (1999), « [Enhancing the quality of case studies in health services research](#) », *Health Services Research*, vol. 34(5 - Pt 2), p. 1209-1224. Cet article, dans le domaine de la santé, interroge le choix des cas et la qualité des études de cas. Voir aussi Lijphart A. (1972), « Comparative politics and the comparative method », *American Political Science Review*, vol. 64(3), p. 682-693.

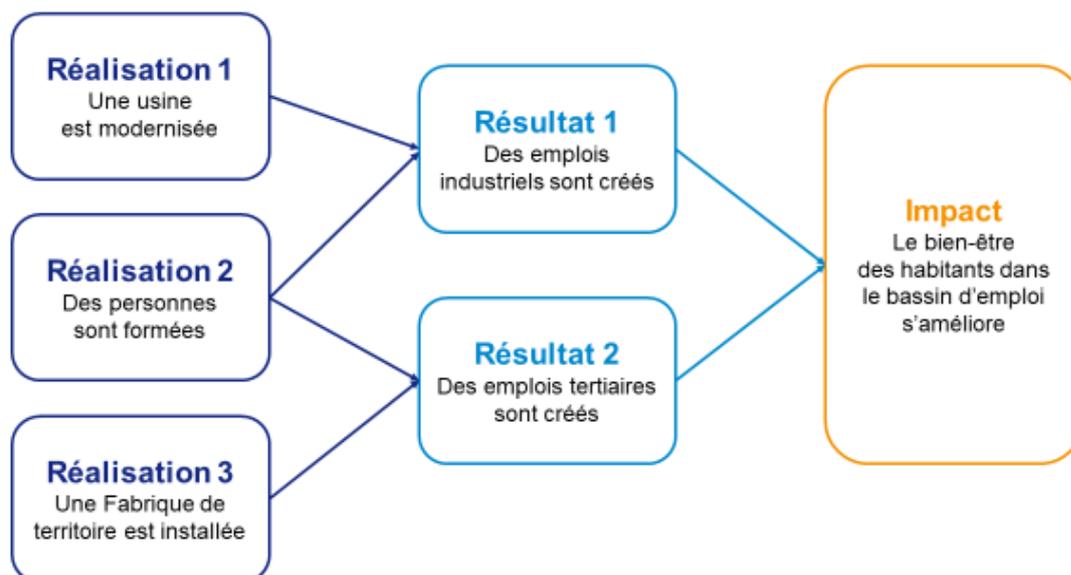
2. Outil transversal n° 2 : le diagramme logique des impacts

En deux mots – Le diagramme logique des impacts, ou logigramme, est la première étape de la plupart des méthodes qualitatives pour évaluer l'impact d'une politique publique. Cette représentation schématique de la théorie d'action, autrement appelée la logique d'intervention, décrit l'ensemble des étapes par lesquelles une intervention entend produire le changement qu'elle vise. Y sont mis en perspective, sous forme d'un enchaînement causal, les réalisations prévues, leurs résultats directs et leurs impacts attendus. Ce diagramme n'a pas vocation à décrire l'intervention telle qu'elle se déroule empiriquement, mais à modéliser la façon dont ses concepteurs imaginent qu'elle va produire un impact. Ce modèle fournit ainsi une série d'hypothèses qui pourront ensuite être confrontées à la réalité empirique. Le diagramme ci-dessous est un exemple de représentation d'un projet de revitalisation d'un bassin d'emploi.

En principe, un diagramme logique des impacts comporte trois niveaux de *maillons causaux* : celui des « réalisations », qui réunit les actions clés ; celui des résultats (*outputs* en anglais), qui réunit les conséquences directes des réalisations ; et celui des impacts (*outcomes*), intermédiaires ou finaux, qui réunit les conséquences plus larges des réalisations. La figure 2 ci-dessous donne un exemple de logigramme dans le cadre d'un projet de revitalisation d'un bassin d'emploi.

¹ Inspection générale des finances et Conseil général de l'environnement et du développement durable (2018), [L'avenir des opérateurs de l'eau et de la biodiversité](#), avril.

Figure 2 – Logigramme d'un projet de revitalisation d'un bassin d'emploi



Source : ANCT – Production de la Mission d'appui à la performance des projets (MAPP), citée dans ANCT (2020), *Conseils méthodologiques pour la conduite d'une évaluation locale d'un projet ACV*, dossier de présentation, p. 5

Une même intervention publique peut être représentée de diverses façons, en fonction des éléments que ses concepteurs jugeront significatifs ou non, et en fonction des formulations choisies. La littérature académique propose quelques conseils pour réaliser un diagramme utile à l'évaluation d'impact. En particulier :

- formuler chacun des maillons avec un verbe d'action et préciser éventuellement l'acteur à l'origine de cette action. Par exemple « Une usine est modernisée » ou « La région modernise une usine » ;
- se concentrer sur les maillons qui sont de nature à générer un effet sur les maillons suivants. *A contrario*, les simples successions temporelles sont à éviter. Par exemple, « Une usine est créée » suivi de « Des emplois industriels sont créés », plutôt que « Un entrepôt est créé », « Un fourneau est créé », etc. ;
- veiller à ce que l'enchaînement de deux maillons traduise une causalité directe. *A contrario*, éviter des *trous* dans la chaîne de causalité. Par exemple, entre le maillon « Des livres sont distribués aux élèves » et le maillon « Les élèves gagnent en compétences langagières », il peut être utile d'ajouter le maillon « Les élèves lisent davantage », car c'est par ce biais que les compétences langagières peuvent être développées, et parce qu'il n'est pas évident qu'un élève lise les livres qui lui sont distribués. L'évaluation pourra ainsi porter une attention particulière à la lecture effective des livres distribués ;

- chercher une formulation des maillons suffisamment objectivable, c'est-à-dire qui permette d'envisager un protocole de mesure capable de confirmer ou de réfuter le maillon ou son enchaînement avec d'autres maillons. Par exemple, le maillon « Des livres sont distribués aux élèves » peut être plus facilement mesurable que le maillon « Du matériel pédagogique est mobilisé » ;
- trouver un équilibre entre une formulation précise et une formulation générique de chacun des maillons. Si la précision permet de rendre chacun des maillons plus objectivable (voir le point précédent), le caractère générique d'une formulation permet de donner au diagramme logique une portée théorique plus large, avec des possibilités de généralisation et d'extrapolation. Par exemple, plutôt que de formuler le maillon « L'ONG XY utilise les résultats du laboratoire YX », il peut être utile de formuler le maillon « La société civile mobilise l'expertise de la recherche » plus générique, et donc potentiellement extrapolable à d'autres situations que celle immédiatement étudiée.

Quel apport pour l'évaluation d'impact ? Le diagramme logique des impacts est très utile pour l'évaluation d'impact, au moins à trois égards. Il permet de visualiser de façon simplifiée les composantes clés d'une intervention publique et sa portée attendue, et d'identifier les questions évaluatives les plus pertinentes. Il permet aussi d'interroger les hypothèses sous-jacentes à l'intervention, et notamment les intentions politiques du législateur. Enfin, si les parties prenantes sont impliquées pour concevoir le diagramme logique des impacts, l'exercice permet de confronter les points de vue, et de faire émerger un consensus sur la représentation de l'intervention et les modalités de son évaluation.

Pour aller plus loin – Weiss C. H. (1997), « How can theory-based evaluation make greater headway? », *Evaluation Review*, vol. 21(4), août, p. 501-524 ; et McLaughlin J. A. et Jordan G. B. (2015), « Using logic models », in Newcomer K. E., Hatry H. P. et Wholey J. S. (dir.), *Handbook of Practical Program Evaluation*, 4^e éd., San Francisco, John Wiley and Sons, p. 62-87.

3. Méthode qualitative n° 1 d'évaluation d'impact : la cartographie conceptuelle d'impacts

En deux mots – La cartographie conceptuelle d'impacts¹ consiste à demander aux parties prenantes d'une politique, ou à ses observateurs, de lister eux-mêmes les impacts qu'ils prêtent à une politique publique. À partir de leurs connaissances, de leurs expériences ou de leurs perceptions, les acteurs interrogés doivent aussi regrouper les impacts en fonction de leur proximité et les hiérarchiser en fonction de leur importance relative. Une analyse

¹ Ou *conceptual outcome mapping* en anglais.

statistique agrège les réponses obtenues et dégage des types d'impact de plus ou moins grandes importances.

Quel apport pour l'évaluation d'impact ? La cartographie conceptuelle peut être utile en amont d'une évaluation d'impact. Elle vise en effet à relever le plus exhaustivement possible les impacts potentiels d'une politique publique. En les regroupant et en les hiérarchisant, elle peut aussi aider à identifier les questions évaluatives et les indicateurs les plus pertinents. En outre, parce qu'elle cherche à tenir compte de l'ensemble des points de vue, cette méthode peut renforcer la crédibilité des résultats de l'évaluation et leur appropriation par les parties prenantes. Pour être fructueuse, cette méthode à partir d'entretiens¹ doit veiller à la diversité et la légitimité des personnes interrogées (l'expertise d'un chercheur, l'expérience d'un usager, le mandat d'un opérateur, etc.). Elle doit aussi chercher à garantir la libre expression des personnes interrogées et à favoriser la convergence – à défaut de la juste représentation – des différents points de vue.

Exemple – Dans un rapport de 2010² du cabinet Euréval, commandée par la Direction générale de la traduction de la Commission européenne, une étude a été menée sur l'impact de la traduction au niveau des institutions européennes et plus généralement dans la société européenne. À cet effet, neuf experts et chercheurs ont été interrogés, sur leurs perceptions et leurs connaissances des impacts – potentiels ou avérés – de la traduction. Pour susciter des réponses riches et exactes, des questions spécifiques ont été posées, par exemple sur les différents types d'impacts (économiques, culturels, etc.), ou encore sur les conséquences que provoquerait la disparition de la traduction.

Grâce à ces entretiens, près de 84 impacts ont été relevés³. Les participants ont ensuite été invités à regrouper les impacts en catégories. Six grandes catégories d'impacts ont ainsi été formulées⁴ : « les échanges culturels⁵ », « la globalisation de l'économie », « les

¹ Plusieurs techniques d'entretiens en sciences sociales peuvent être mobilisées ici à profit : le *focus group* qui consiste à interroger plusieurs personnes dans la même pièce afin d'observer aussi leurs interactions ; le *panel d'experts* qui correspond à une forme de *focus group* entre experts où la convergence des points de vue est recherchée ; ou encore la *méthode Delphi* qui interroge séparément des experts – et à l'écrit –, et qui leur soumet les réponses anonymisées de leurs pairs afin de recueillir de nouvelles réactions.

² Commission européenne/Euréval (2011), *Étude portant sur la contribution de la traduction à la société multilingue dans l'Union européenne*, novembre.

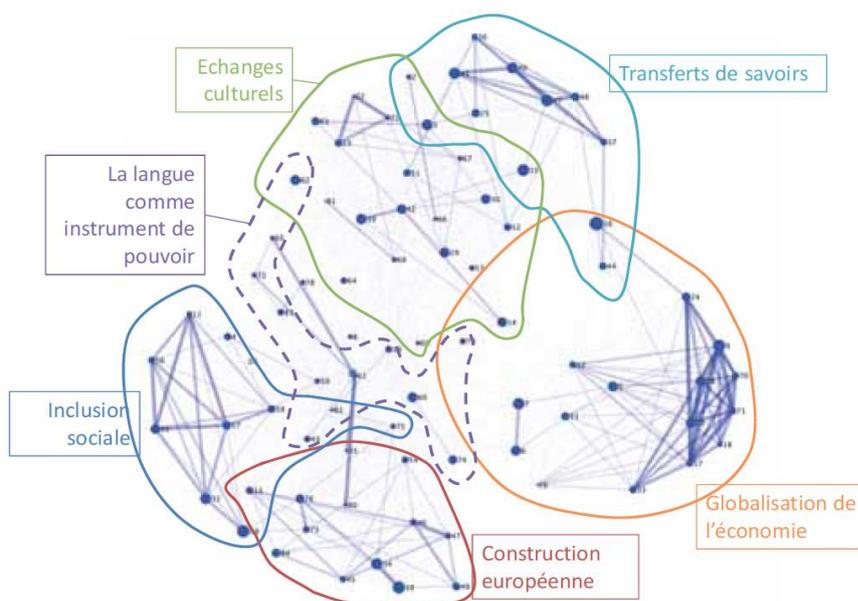
³ Par exemple : « La traduction promeut la diversité comme une caractéristique de l'identité de l'Union européenne » ; « La traduction aide les individus à mieux se comprendre dans leur environnement professionnel et personnel », « La traduction peut alimenter des tensions entre communautés culturelles (du fait d'une traduction inexacte ou inexistante) », etc.

⁴ Avec quelques recouvrements toutefois, mais relativement minimes.

⁵ Par exemple, cette catégorie a fini par regrouper les impacts suivants (issus du premier tour des entretiens) : la facilitation de la compréhension mutuelle, les rencontres et les transactions entre personnes de cultures

transferts de savoirs », « la construction européenne », « l'inclusion sociale », « la langue comme instrument de pouvoir ». Les participants ont aussi dû se prononcer sur l'importance de chacun des impacts. La synthèse de ces réponses a permis d'obtenir la cartographie suivante (voir Figure 3), où les cercles traduisent le regroupement des impacts, et où la taille des points traduit leur importance relative.

Figure 3 – Représentation des catégories d'impact à l'issue d'une cartographie conceptuelle d'impacts



Lecture : chaque point représente un des 87 impacts recensés au début de l'exercice ; plus l'impact est jugé important, plus le point est gros ; les anneaux colorés regroupent chacun un certain nombre d'impacts en fonction de leur proximité telle que perçue par les experts et les chercheurs interrogés.

Source : Commission européenne/Euréval (2011), *Étude portant sur la contribution de la traduction à la société multilingue dans l'Union européenne*, op. cit., p. 14

Cette cartographie constitue un travail préparatoire pour des évaluations d'impact, portant spécifiquement sur telle ou telle catégorie d'impact, ou sur certaines de leurs composantes.

Pour aller plus loin – Trochim W. (1989), « *An introduction to concept mapping for planning and evaluation* », *Evaluation and Program Planning*, vol. 12(1), p. 1-16. Et dans le champ de d'évaluation collaborative : Fetterman D. M., Rodríguez-Campos L., Zukoski A. P. et al. (2018), *Collaborative, Participatory, and Empowerment Evaluation: Stakeholder Involvement Approaches*, New York, Guilford Press.

différentes, l'accès pour un individu à des cultures étrangères et aux cultures anciennes, la promotion d'une langue par une autorité régionale, etc.

4. Méthode qualitative n° 2 d'évaluation d'impact : l'analyse de contribution

En deux mots – L'analyse de contribution consiste à identifier les éléments d'une politique publique qui contribuent le plus à son impact. Là où les méthodes contrefactuelles ont tendance à appréhender la politique publique d'un seul tenant et à lui attribuer un impact¹, l'analyse de contribution s'attache d'abord à décomposer l'intervention publique et à établir les chemins par lesquels ses composantes peuvent chacune générer un impact. À partir d'une collecte de données, la vraisemblance des différents chemins de causalité est testée, ainsi que l'amplitude de leurs contributions respectives à l'impact global de l'intervention publique. La collecte de données peut en outre conduire à réviser la théorie d'action, et à la tester à nouveau, de façon itérative.

Quel apport pour l'évaluation d'impact ? Cette méthode peut être utile pour identifier les composantes de la politique publique qui contribuent le plus à l'impact global. Par contraste, elle peut aider à relever les composantes qui n'ont pas d'effet, voire qui ont un effet contre-productif. Cette méthode peut donc aider à réviser les politiques publiques et à les recentrer – dans une logique d'efficacité – sur leurs composantes les plus contributives², et à écarter leurs composantes aux effets contre-productifs.

Exemple : dans un article de 2017³, Delahais et Toulemonde évaluent l'impact des actions menées par deux centres de recherche – le CIFOR et le CIRAD⁴ – sur les changements de pratiques de la filière bois dans le bassin du Congo, et en particulier sur la soutenabilité de ces pratiques. Dans un premier temps, l'action des deux centres de recherche a été décomposée en chacune de ses composantes susceptibles de générer un impact sur la filière bois. Par exemple : les centres ont développé des concepts susceptibles de modifier les représentations des acteurs ; les centres ont produit des données susceptibles de renforcer la connaissance des réalités du terrain par les acteurs ; les centres ont développé des solutions innovantes susceptibles d'inspirer de nouvelles pratiques ; les centres ont organisé des échanges entre les parties prenantes susceptibles de favoriser le partage de bonnes pratiques ; etc.

¹ À cet égard, cette approche est parfois qualifiée d'*analyse d'attribution*.

² Sous réserve qu'elles peuvent générer cet effet de façon autonome, c'est-à-dire indépendamment d'autres composantes de la politique.

³ Delahais T. et Toulemonde J. (2017), « Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management? », *Evaluation*, vol. 23(4), octobre, p. 370-388.

⁴ CIFR pour Centre for International Forestry Research et CIRAD pour Centre de coopération internationale en recherche agronomique pour le développement.

Puis une collecte de données a été réalisée afin de mesurer la vraisemblance de chacun de ces *chemins de contribution*, et son importance relative. Ces données touchent particulièrement l'enchaînement des actions et la perception des acteurs. En particulier, il est apparu que les concepts développés par les centres de recherche ont été utilisés par les cabinets de conseil au contact de la filière forestière. La causalité semble établie au regard de la temporalité : les cabinets de conseil n'utilisaient pas ces concepts avant leur développement par les centres de recherche, et ils ont commencé à les utiliser peu de temps après. Surtout, les entreprises elles-mêmes ont révélé, lors d'entretiens, qu'elles s'étaient directement inspirées des publications des deux centres de recherche. En définitive, de l'action globale des centres de recherche, le développement de nouveaux concepts est apparu comme l'une des composantes contribuant le plus aux changements de pratique de la filière bois dans le bassin du Congo. En outre, les conclusions de cette étude ont été soumises à des experts et des chercheurs, pour en évaluer la crédibilité, aider à les réviser si besoin, en répétant la démarche de façon itérative.

Pour aller plus loin – Mayne J. (2012), « [Contribution analysis: Coming of age?](#) » *Evaluation*, vol. 18(3), juillet, p. 270-280.

5. Méthode qualitative n° 3 d'évaluation d'impact : la reconstitution de processus ou *process tracing*

En deux mots – La reconstitution de processus consiste à reconstituer le chemin causal qui mène d'une situation A à une situation B. La méthode est particulièrement attentive à l'enchaînement des actions dans le temps et aux mécanismes causaux sous-jacents. La théorie d'action est formulée selon un enchaînement de mécanismes M_i menant de la situation A à la situation B : $A \rightarrow M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow (\dots) \rightarrow B$. Chacun des mécanismes causaux est ensuite confirmé ou infirmé à partir d'une collecte de données et de quatre types de tests¹ :

- Le « ***straw in the wind test*** » (*l'indice simple*) : ce test est vérifié en présence d'indices qui corroborent la vraisemblance d'un mécanisme causal, mais sans être des preuves suffisantes ou nécessaires. Par exemple, s'il est question d'étudier la causalité entre « La distribution de livres » et « Le développement des capacités langagières des élèves issus de familles défavorisées », le fait de relever une augmentation du nombre de livres dans les cartables des élèves issus de familles défavorisées semble corroborer la causalité. Mais il ne s'agit pas d'une preuve nécessaire (les élèves pourraient lire davantage sans transporter les livres dans leurs cartables) ni d'une

¹ Ces méthodes sont désormais régulièrement utilisées au sein d'autres méthodes.

preuve suffisante (il ne suffit pas qu'un élève transporte un livre dans son cartable pour qu'il lise davantage) ;

- Le « **hoop test** » (le *cerceau*) : ce test est vérifié en présence de preuves nécessaires mais non suffisantes. En reprenant le même exemple, ce test serait vérifié si on relevait une augmentation des activités de lecture chez les élèves issus de familles défavorisées. En effet, il est nécessaire que ces élèves lisent davantage pour développer leurs compétences langagières. Mais cela n'est pas suffisant car il se pourrait que les élèves consacrent davantage de temps à essayer de lire mais sans y parvenir.
- Le « **smoking gun test** » (le *pistolet fumant*) : ce test est vérifié en présence de preuves suffisantes mais non nécessaires, en d'autres termes si la preuve établit la causalité sans ambiguïté, mais ne permet pas d'exclure d'autres causes. Dans le même exemple, ce test serait vérifié si on relevait une augmentation du temps de lecture effective des livres distribués auprès des élèves issus de familles défavorisés. En effet, ce serait une preuve que la distribution de livres développe bien les capacités langagières des élèves bénéficiaires. Mais si dans le même temps, dans l'établissement, il existait un nouvel atelier de lecture du soir, il resterait possible que ce soit également cet atelier qui renforce les capacités langagières des élèves.
- le « **doubly decisive test** » (le *test doublement décisif*) : ce test est vérifié en présence de preuves à la fois nécessaires et suffisantes, en d'autres termes si la causalité est établie et si aucune autre explication causale n'est envisageable. Ce serait le cas dans l'exemple développé s'il n'existait pas d'atelier de lecture du soir ni aucun autre dispositif susceptible de renforcer spécifiquement les capacités langagières des élèves issus de familles défavorisées.

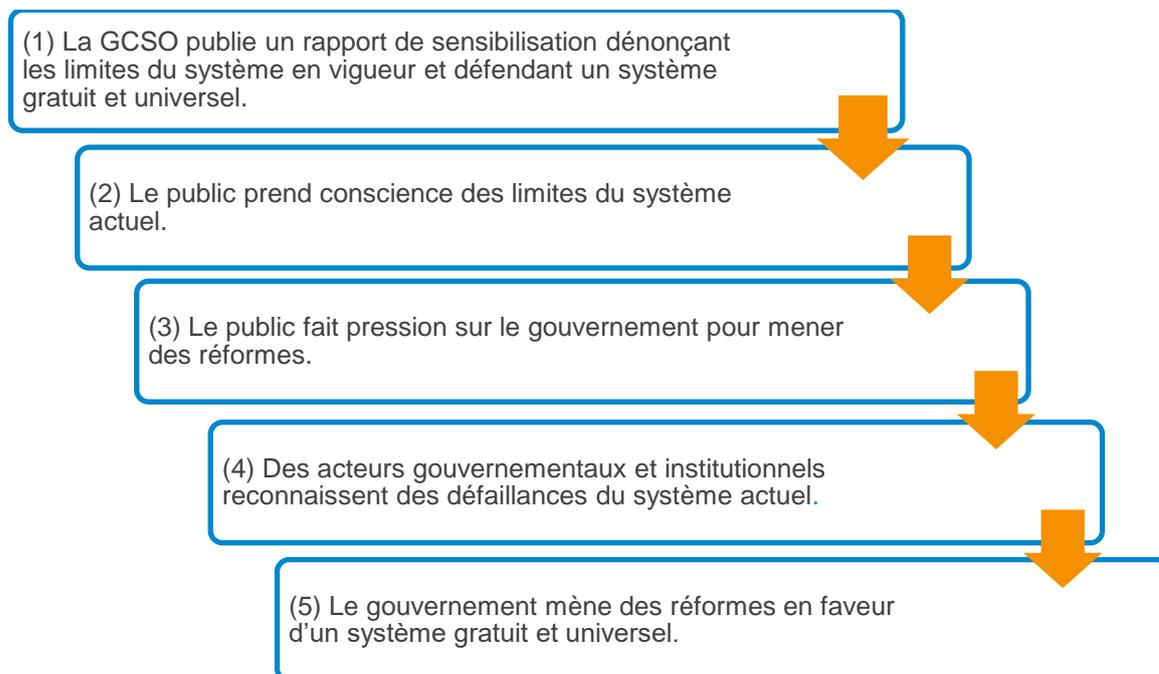
Quel apport pour l'évaluation d'impact ? La méthode est particulièrement utile lorsqu'il y a peu de cas d'observations, *a fortiori* lorsqu'il existe un unique cas d'observation (par exemple l'impact d'un événement historique sur le marché économique) et lorsqu'il n'est pas possible de comparer des situations « toutes choses égales par ailleurs ». Elle permet de comprendre la formation d'un impact, et son développement dans une temporalité et un contexte particuliers.

Exemple : dans un article de 2013¹, Stedman-Bryce s'intéresse à l'action de lobbying de la *Ghana Civil Society Organization* (GCSO) et à l'impact de cette action sur la mise en place d'un

¹ Stedman-Bryce G. (2013), *Health for All: Towards Free Universal Health Care in Ghana. End of Campaign Evaluation Report*, Oxford, Oxfam GB.

système de santé universel et gratuit. Stedman-Bryce commence par proposer une théorie de l'action autour d'un certain nombre de mécanismes causaux (voir Figure 4).

Figure 4 – Théorie de l'action proposée par Stedman-Bryce pour représenter l'impact de l'action de lobbying



Source, France Stratégie, d'après Stedman-Bryce G. (2013), *Health for All...*, op. cit.

L'articulation entre chacun des maillons est étudié. Par exemple, certaines données tendent à prouver que l'action de la GCSO a bien poussé (dans un sens causal) le gouvernement à reconnaître les défaillances du système actuel. Il apparaît notamment que le National Health Insurance Authority, un acteur institutionnel national, a reconnu avoir surestimé le nombre de bénéficiaires dans le système actuel, suite à la publication du rapport GCSO, aveu qui aurait été impensable sans la publication du rapport et la mobilisation sociale qui s'en est suivie. Le test du « *smoking gun test* » est donc vérifié. Au terme de l'analyse, l'évaluation conclut que les causalités (1), (2) et (3) sont effectivement vraisemblables, car étayées par l'observation. En revanche, la causalité (4) ne semble pas s'être produite.

Pour aller plus loin – Beach D. et Pedersen R. B. (2013), *Process-Tracing Methods: Foundations and Guidelines*, Ann Arbor, University of Michigan Press ; Surel Y. (2018), « La mécanique de l'action publique. Le *process tracing* dans l'analyse des politiques publiques », *Revue française de science politique*, vol. 68, p. 991-1014 ; Palier B. et Trampusch C. (2018), « Comment retracer les mécanismes causaux. Les différents usages du *process tracing* », *Revue française de science politique*, vol. 68, p. 967-990.

6. Méthode mixte n° 1 d'évaluation d'impact : l'analyse qualitative comparée

En deux mots – L'analyse qualitative comparée consiste à identifier les contextes où une politique publique a l'impact le plus fort, et à les distinguer des contextes où l'impact est faible, nul, voire négatif. La méthode commence par une identification des facteurs susceptibles de conditionner le succès ou l'échec d'une intervention¹. Puis, dans la mesure du possible, l'intervention publique est observée dans plusieurs contextes et sous différentes combinaisons, c'est-à-dire en présence ou non de chacun des facteurs identifiés. Une analyse statistique étudie ensuite les corrélations entre le succès ou l'échec de l'intervention, dans un contexte donné, et les caractéristiques de ce contexte. Cela permet d'identifier les facteurs, ou les combinaisons de facteurs, les plus fortement corrélés au succès ou à l'échec d'une intervention. En particulier, il est possible d'identifier les facteurs qui semblent nécessaires – ou suffisants – pour observer un succès ou un échec².

Quel apport pour l'évaluation d'impact ? Pour l'évaluation d'impact, cette méthode peut être utile pour identifier les contextes où la politique publique produit les impacts les plus importants, relativement aux contextes où elle produit des impacts plus faibles, voire négatifs. Cette approche s'insère dans l'approche générale d'une *évaluation réaliste* qui considère qu'une intervention n'a pas d'effet en soi, mais un effet qui dépend du contexte où elle s'applique. En outre, l'analyse qualitative comparée permet d'apprécier des effets cocktails, en relevant les combinaisons de facteurs qui coïncident avec le succès d'une intervention, et les combinaisons qui coïncident au contraire avec son échec. Elle permet aussi de distinguer les facteurs nécessaires ou non, et suffisants ou non à l'échec ou au succès de la politique publique, ce qui peut aider à mieux concevoir, réviser ou mettre en œuvre une intervention publique en fonction des spécificités du terrain. La méthode est particulièrement opportune pour étudier les processus de causalités multiples et imbriquées (plusieurs causes interagissent pour produire un résultat), les processus d'asymétrie (les causes d'un résultat peuvent être différentes des causes de l'absence d'un résultat), les processus d'indétermination (les mêmes causes ne produisent pas toujours les mêmes effets) et les processus d'équifinalité (les mêmes effets ne sont pas toujours issus des mêmes causes).

¹ Cette étape, qui repose généralement sur la collecte de données (revue de littérature, entretiens, observation du terrain, etc.), prend parfois le nom de *screening and scoping*.

² Dans le jargon, on dira qu'une combinaison de facteurs présente un fort indicateur de *consistency* si elle est (presque) systématiquement associée au succès (ou à l'échec) d'une intervention. Par exemple, si le dédoublement des classes semble rencontrer un succès dans 70 % des 100 classes observées où l'enseignant a plus de cinq ans d'expérience, l'indicateur de *consistency* est de 70 % (ce qui correspond à 30 % de cas déviants). De façon corollaire, une combinaison de facteurs présente un fort indicateur de *coverage* si elle recouvre l'ensemble des situations de succès (ou d'échec). En reprenant l'exemple précédent, si le dédoublement de classe rencontre également un succès dans 70 % des 100 autres classes observées où l'enseignant a moins de cinq ans d'expérience, l'indicateur de *coverage* est de 50 %.

Exemple : dans un article de 2013¹, Blackman s'intéresse aux politiques britanniques de lutte contre les grossesses adolescentes. Il propose d'apprécier, sur plusieurs territoires à risque, la réduction (ou non) du taux local de grossesses adolescentes vers la moyenne nationale. À l'issue d'un travail exploratoire (entretiens, revue de littérature), une série de facteurs susceptibles de conditionner cette réduction est identifiée : la proportion de jeunes de moins de 18 ans sur le territoire, la qualité des structures d'accompagnement, la prévalence des drogues ou encore la diversité ethnique.

Ces facteurs sont ensuite codés de façon binaire, autour de deux modalités chacun : présent (1) / absent (0) ou élevé (1) / faible (0)². Sur des territoires hétérogènes du point de vue de ces facteurs, le chercheur relève les combinaisons de facteurs qui coïncident avec une baisse des grossesses adolescentes, et celles qui coïncident avec un maintien ou une hausse des grossesses adolescentes. Par différence, il apparaît que certains facteurs semblent suffisants pour générer une réduction des grossesses adolescentes. En l'occurrence, il suffit que le territoire soit marqué par une forte diversité ethnique pour que la réduction ait lieu, quelles que soient les modalités prises par les autres facteurs. Le cas symétrique est également étudié : sont relevés les facteurs (ou les combinaisons de facteurs) qui apparaissent suffisants ou nécessaires pour que la réduction des grossesses adolescentes n'ait pas lieu. Il apparaît par exemple qu'il suffit d'avoir la combinaison « faible diversité ethnique » et « faible proportion de jeunes de moins de 18 ans » pour que la réduction n'ait pas lieu.

Une discussion est alors menée, notamment au regard de la littérature académique, pour tenter de comprendre en quoi ces associations pourraient s'expliquer. Une forte diversité ethnique, par exemple, pourrait être associée à une forte mobilisation sociale pour sensibiliser aux maladies sexuellement transmissibles, ce qui pourrait expliquer la réduction des grossesses adolescentes dans ces territoires. Ce raisonnement permet ainsi d'identifier les facteurs favorisant le succès ou l'échec d'une politique, et peut aider à adapter la politique aux spécificités des territoires, en concentrant par exemple l'effort sur les territoires regroupant les modalités les plus défavorables.

¹ Blackman T. (2013), « Exploring explanations for local reductions in teenage pregnancy rates in England: an approach using Qualitative Comparative Analysis », *Social Policy & Society*, vol. 12(1), janvier, p. 61-72.

² Dans cette méthode, il est possible de coder les facteurs selon un continuum de valeurs entre 0 et 1.

Tableau 1 – Représentation, ici réduite à deux facteurs, des résultats de la QCA

Facteurs	Forte diversité ethnique	Faible diversité ethnique
Forte proportion de jeunes de moins de 18 ans	Réduction (quasi-) systématique	(résultat non significatif)
Faible proportion de jeunes de moins de 18 ans	Réduction (quasi-) systématique	Non-réduction (quasi-) systématique

Lecture : sur les territoires présentant une forte diversité ethnique, et quelle que soit la proportion de jeunes de moins de 18 ans, il a été observé (quasi) systématiquement une réduction des grossesses adolescentes.

Source : France Stratégie, d'après Blackman T. (2013), « Exploring explanations for local reductions in teenage pregnancy rates in England », op. cit.

Pour aller plus loin – Pattyn V., Molenveld A. et Befani B. (2019), « Qualitative comparative analysis as an evaluation tool: Lessons from an application in development cooperation », *American Journal of Evaluation*, vol. 40(1), p. 55-74.

7. Méthode mixte n° 2 d'évaluation d'impact : la méthode d'évaluation faisabilité-impact

En deux mots – Dans une thèse publiée en 2018¹, Baiz développe une méthode mixte² qui permet d'évaluer conjointement la faisabilité et l'impact d'une intervention publique. La méthode consiste en deux étapes. Il s'agit d'abord de représenter la théorie d'action, en précisant notamment l'influence que les acteurs sont censés exercer les uns sur les autres : « obliger/forcer », « interdire/empêcher », « encourager/décourager », « autoriser/ permettre » et « rendre facultatif ». Par exemple, la taxe carbone pourra être représentée de la façon suivante : « Si une entreprise pollue, alors l'État l'oblige à payer une taxe proportionnelle, de façon à la décourager à polluer (davantage) ». Il s'agit ensuite – par une observation du terrain, des entretiens ou tout autre moyen empirique – de vérifier l'effectivité de chacune de ces influences, et de la traduire à l'aide d'un raisonnement probabiliste et contrefactuel. Pour vérifier par exemple si la taxe décourage la pollution, il s'agira de vérifier : (a) d'une part que la taxe est bien appliquée ; (b) d'autre part que la

¹ Baiz A. (2018), *De l'innovation des instruments de politique publique. Développement d'une méthode de conception combinatoire autour d'un langage algorithmique et application au dispositif des certificats d'économie d'énergie*, thèse de doctorat, université Paris Sciences et Lettres.

² La méthode est mixte à deux égards : elle hybride des méthodes qualitatives et quantitatives, et elle porte conjointement sur les critères de faisabilité et d'impact.

probabilité qu'une entreprise polluée baisse suite à la mise en place de la taxe, et qu'elle n'aurait pas baissé en l'absence de la taxe.

Quel apport pour l'évaluation des politiques publiques ? Cette méthode développe un langage qui permet de représenter plus explicitement les influences – et donc les types de causalité – que peuvent exercer les acteurs les uns sur les autres. Elle permet de visualiser, au niveau de granularité choisi, tous les maillons où un biais pourrait intervenir. Ainsi, les critères de faisabilité et d'impact d'une intervention publique sont appréhendés de façon conjointe : il suffit qu'une action ou une influence n'ait pas lieu comme prévu pour que l'impact de l'intervention soit altéré.

Exemple – Dans ses travaux de thèse, Baïz (2018) s'intéresse au dispositif des certificats d'économie d'énergie. De façon simplifiée, ce dispositif en France *oblige* les principaux vendeurs d'énergie à *encourager* les ménages à rénover leur chaudière, leur toiture ou encore leurs combles. À chaque fois qu'il y parvient, le vendeur d'énergie obtient un CEE, et s'il n'en cumule pas assez au bout de trois ans, il doit payer une amende à l'État. Le dispositif prévoit aussi un marché d'échanges de CEE, et des mécanismes de subvention et de délégation. La thèse propose de représenter le plus clairement possible la complexité du dispositif, puis évalue la faisabilité (l'effectivité) de chacun de ses maillons causaux. Ainsi par exemple, il apparaît que des vendeurs d'énergie se prévalent d'avoir encouragé des ménages à rénover leurs chaudières, alors que ces derniers prévoyaient de le faire de toute façon (du fait par exemple de la vétusté de leur ancienne chaudière). Cette méthode permet ainsi d'identifier le risque d'un effet d'aubaine, et toute autre sorte de biais de mise en œuvre pouvant altérer l'impact global du dispositif sur les économies d'énergie.

Pour aller plus loin – Baïz A. et Nakhla M. (2018), « [Pour une approche algorithmique de la nature protéiforme et fractale des instruments de l'action collective](#) », *Politiques et management public*, vol. 35(3-4), p. 153-172 ; Baïz A. (2018), [De l'innovation des instruments de politique publique. Développement d'une méthode de conception combinatoire autour d'un langage algorithmique et application au dispositif des certificats d'économie d'énergie](#), thèse de doctorat, université Paris Sciences et Lettres¹.

¹ Un résumé de la thèse est disponible sur [cet article](#) de *The Conversation*.

8. Enrichir une expérimentation par des entretiens semi directifs

Dans le cadre d'un projet de recherche interdisciplinaire¹ lancé en 2016, des chercheurs associés au LIEPP ont cherché à évaluer de façon expérimentale l'impact de la lecture parentale sur les compétences langagières² des enfants, et en particulier sur celles des enfants défavorisés. Dans la littérature académique, il est établi que les compétences langagières dépendent fortement des interactions entre parents et enfants. Pourtant, et de façon contre-intuitive, la trentaine d'expérimentations aléatoires portant sur la lecture parentale aboutissent à des résultats contrastés et parfois contradictoires. Deux raisons peuvent expliquer cela. D'une part, il est possible que les expérimentations aient été mal conçues ; ainsi par exemple, les parents ciblés n'ont pas été suffisamment motivés et accompagnés pour se prêter à l'expérimentation. D'autre part, il est possible que les évaluations ne soient pas parvenues à capter les effets positifs, du fait de défaillances méthodologiques ; ainsi par exemple, les échantillons étudiés ont parfois été trop petits³. L'article de Pin et Barone (2021)⁴ fait suite à cette évaluation expérimentale, et souligne la plus-value de croiser des méthodes qualitatives et des méthodes quantitatives, tant pour la qualité de l'évaluation que pour le succès de l'expérimentation.

L'expérimentation en tant que telle a consisté à tirer au sort 22 écoles maternelles en réseau d'éducation prioritaire (REP et REP+) dans plusieurs arrondissements de Paris. L'ensemble des 900 élèves a passé un premier test pour apprécier leurs compétences langagières initiales. Puis une moitié des classes dans chaque école a été tirée au sort pour bénéficier du traitement, à savoir : le prêt de deux livres par semaine à chaque enfant, et des actions d'accompagnement auprès des parents concernés. L'autre moitié des

¹ Le [projet](#) s'appelle « Favoriser les compétences langagières des enfants des milieux défavorisés : une évaluation expérimentale d'un dispositif d'accompagnement à la lecture parentale ». Il a été conçu et mis en œuvre entre septembre 2016 et juin 2018 sous la direction d'une équipe de sociologues et d'économistes de l'éducation composée de Carlo Barone, Denis Fougère et Agnès van Zanten. Cette recherche a bénéficié du soutien apporté par l'Agence nationale de la recherche (ANR) et par l'État au titre du programme d'investissements d'avenir dans le cadre du Labex LIEPP.

² Les compétences langagières renvoient en particulier à l'ampleur du vocabulaire, à la conscience phonologique ou encore à la familiarité avec les conventions de l'écrit. La lecture parentale désigne simplement l'activité où les parents lisent des livres à leurs enfants.

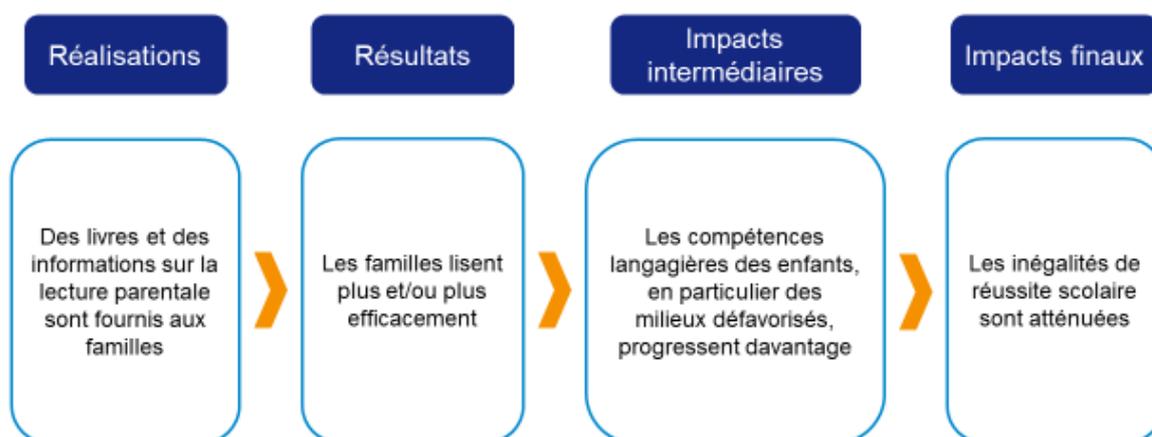
³ L'échec d'une politique peut tenir à deux raisons : l'insuccès de son exécution et/ou son inadéquation avec l'objectif visé. Voir Suchman E. A. (1967), *Evaluative Research. Principles and Practice in Public Service and Social Action Programs*, New York, Russel Sage Foundation. Il est possible aussi que la politique produise les effets escomptés mais que l'évaluation ne parvienne pas à les identifier, par manque de recul ou à cause de biais méthodologiques. Sur le sujet de l'éducation, voir aussi : Lortie-Forgues H. et Inglis M. (2019), « Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? », *Educational Researcher*, vol. 48(3), avril, p. 158-166.

⁴ Pin C. et Barone C. (2021), « [L'apport des méthodes mixtes à l'évaluation...](#) », *op. cit.*

classes n'a pas bénéficié de ce traitement, afin de constituer un groupe de contrôle. Quelques mois plus tard un nouveau test a été administré aux 900 élèves pour comparer les évolutions entre le groupe de traitement et le groupe de contrôle. Il en est ressorti des effets positifs et significatifs, mais à certains égards hétérogènes, du programme sur la fréquence de la lecture partagée et sur le développement langagier des enfants (défavorisés en particulier).

Si la méthode d'évaluation expérimentale ici employée semble classique, elle a en réalité été enrichie par la mobilisation de méthodes qualitatives. D'abord, il a été question d'expliciter la théorie de l'action (voir le diagramme logique des impacts) en représentant l'articulation théorique entre les réalisations et les impacts finaux (voir Figure 5).

Figure 5 – Théorie de l'action de l'expérimentation



Source : Pin C. et Barone C. (2021), « *L'apport des méthodes mixtes à l'évaluation...* », op. cit., p. 398

Il s'agit alors d'interroger la vraisemblance de l'enchaînement causal entre les différents maillons, notamment au regard du contexte particulier de l'expérimentation et de la diversité des profils ciblés. Ainsi par exemple, les chercheurs ont émis l'hypothèse que la distribution de livres ne pourra véritablement causer la lecture parentale par les parents défavorisés que si ces derniers sont persuadés de son bénéfice. Une attention particulière a donc été portée à la sensibilisation de ces parents pendant l'expérimentation.

Des entretiens semi-directifs ont également été réalisés avec des parents et des directeurs d'école, en amont et au cours de l'expérimentation. Ils ont eu deux principaux apports. D'une part, ils ont permis de collecter des données sur les caractéristiques des personnes concernées, et notamment sur l'environnement familial de l'enfant. Ces données ont pu ensuite être croisées avec les résultats finaux de l'expérimentation pour mieux les comprendre. Ainsi par exemple, il est apparu que le programme a eu peu d'effets sur les enfants de parents favorisés, essentiellement parce que ces derniers pratiquaient déjà la

lecture parentale. La méfiance des parents envers l'institution scolaire est aussi apparue comme déterminante sur l'amplitude des effets du programme.

Les entretiens ont par ailleurs permis d'identifier, en amont de l'expérimentation et au fil de l'eau, des difficultés pratiques pouvant entraver le déroulement de l'expérimentation ou dégrader la rigueur de l'évaluation elle-même¹. Ces difficultés concernent généralement la *non-compliance*, lorsque des bénéficiaires ne se conforment pas au programme, et le *spillover effect* lorsque, au contraire, des non-bénéficiaires parviennent à bénéficier du programme. En l'occurrence, les entretiens ont permis de souligner que certains parents bénéficiaires n'avaient pas le temps, ou que certains enfants bénéficiaires n'acceptaient pas des livres imposés. Ces retours du terrain ont permis à l'équipe de recherche d'ajuster à la marge le programme. Ainsi, les parents trop occupés (par leur travail) ont été invités à réaliser des séances de lecture courtes et à impliquer, dans la mesure du possible, les grands frères et les grandes sœurs ; et les enfants ont eu la possibilité d'emprunter chaque semaine les deux livres de leur choix. Ainsi les entretiens ont aidé à prévenir ou à désamorcer les principaux risques de neutralisation, d'altération ou de détournement de l'intervention testée. Dans cette logique d'apprentissage collectif, ils ont renforcé le sentiment d'adhésion des bénéficiaires au programme. En outre, les chercheurs ont, à l'issue de l'expérimentation, mis en ligne un *kit* pour permettre à tout établissement volontaire de réaliser sa propre expérimentation ou de généraliser ce programme.

9. Une illustration des combinaisons possibles de méthodes pour approcher tout critère évaluatif

Un grand nombre de combinaisons de méthodes sont possibles. Elles peuvent être utilisées à profit pour l'évaluation d'impact ou l'évaluation de tout autre critère évaluatif. Par exemple, dans un article de 2021², Hourcade et Cortinas Muñoz s'intéressent à la mise en œuvre de politiques territoriales de protection de l'eau face aux pollutions agricoles et industrielles en France. Pour cela, ils combinent une analyse qualitative comparée (*qualitative comparative analysis*, ou QCA), puis une reconstitution de processus (*process tracing*).

¹ La notion de *validité écologique* indique qu'une expérience en conditions réelles, contrairement à une expérimentation en laboratoire, demande à être particulièrement attentive au déroulement pratique de l'expérimentation, qui par définition est plus difficilement contrôlable. Ici l'intervention de l'évaluateur dans la mise en place du dispositif peut *a contrario* être un obstacle à la validité externe : les impacts observés en présence de l'évaluateur pourraient ne pas être observés dans un contexte plus général où l'évaluateur n'est pas présent pour apporter des correctifs et des ajustements à la mise en œuvre du dispositif.

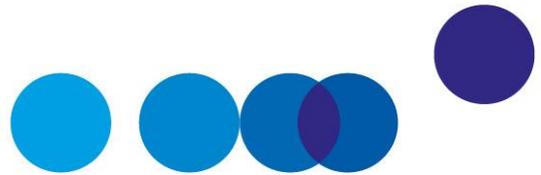
² Hourcade R. et Cortinas Muñoz J. (2021), « [La mise en œuvre des politiques de protection de l'eau...](#) », *op. cit.*

La première méthode (QCA) consiste à travailler un matériau qualitatif de façon à pouvoir coder, à l'issue de l'analyse qualitative, un certain nombre de variables permettant de tester des hypothèses causales à partir de méthodes statistiques. Dans l'étude présentée ici, la démarche démarre avec une revue de littérature, qui permet d'identifier un certain nombre de facteurs susceptibles de favoriser ou de défavoriser la mise en œuvre d'une telle politique locale : le niveau de mobilisation sociale, l'importance économique du pollueur local, le niveau d'incertitudes sur la responsabilité du pollueur, etc. Pour chacun de ces facteurs, une hypothèse de causalité est formulée. Par exemple, il apparaît vraisemblable que plus le pollueur local a un poids économique important, plus il sera à même de peser contre la mise en œuvre d'une telle politique.

Un certain nombre de territoires sont identifiés. Ils sont tous concernés par une situation de pollution chronique d'un milieu aquatique, mais ils se différencient au regard des facteurs identifiés. Ainsi dans certains territoires, le pollueur local a un poids important, dans d'autres territoires, il a un poids faible. Une collecte de données est réalisée, à partir d'une analyse documentaire, des observations du terrain et des entretiens, ce qui permet de caractériser chacun des territoires au travers de valeurs numériques. Par exemple, la valeur « 1 » est donnée à un territoire si le pollueur y a un poids économique important ; la valeur « 0 » est donnée si son poids est nul, et un nombre entre 0 et 1 pour les situations intermédiaires. Chaque territoire se voit aussi affecté une valeur au regard de la mise en œuvre (1) ou non (0) d'une politique de protection de l'eau. À partir des valeurs numériques associées à chaque territoire, une analyse statistique est conduite. Elle permet d'identifier les facteurs qui sont les plus fortement corrélés à la mise en œuvre d'une politique de protection de l'eau, et d'identifier les combinaisons de facteurs qui apparaissent nécessaires ou suffisants pour une telle mise en œuvre. Il en ressort en particulier que l'absence d'incertitudes sur la responsabilité du pollueur et l'existence de coalitions d'acteurs étendues (journalistes, chercheurs, ONG, etc.) sont fortement corrélées à la mise en œuvre de ces politiques, et ce même si le pollueur local a un poids économique important.

La méthode de *process tracing* est alors mobilisée pour enrichir l'analyse sous un angle temporel et causal. D'abord, dans une stratégie d'*enrichissement*, un cas-type est étudié de façon approfondie. Il s'agit du bassin de l'Aa, où la pollution a été longtemps tolérée, vraisemblablement pour ne pas accélérer la désindustrialisation de la région. Mais un accident industriel en 1988 suscite une forte attention médiatique et modifie l'agenda politique du territoire. Une coalition d'acteurs est constituée. La responsabilité des industries polluantes n'est plus un sujet d'incertitudes. En dépit de leur poids économique, des politiques de protection de l'eau sont donc finalement mises en place. Ensuite, et dans une stratégie de *complexification*, un contre-exemple à la théorie est recherché. En l'occurrence, dans la région de Gardanne, l'existence d'une coalition d'acteurs ne suffit pas à mettre en œuvre une politique de protection de l'eau. En regardant dans le détail, il apparaît que le ministère concerné a choisi de ne pas faire partie de cette coalition, après avoir obtenu des rapports

d'expertise contrastés. Grâce à ce contre-exemple (ou cas négatif), les chercheurs proposent d'affûter l'hypothèse causale : la coalition d'acteurs doit comprendre un acteur institutionnel pour véritablement contribuer à la mise en œuvre d'une politique de protection de l'eau.



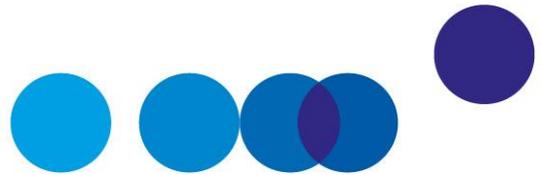
CONCLUSION

Pour évaluer l'impact d'une politique publique, aucune méthode ne peut apporter toutes les réponses ou se prétendre exempte de biais. Bien qu'elles aient historiquement bénéficié d'une aura de scientificité, les méthodes quantitatives sont également concernées par ce constat. Des méthodes qualitatives présentent des spécificités épistémologiques qui peuvent enrichir l'évaluation d'impact. En particulier, elles peuvent aider à expliquer un impact, là où les méthodes quantitatives visent avant tout à le quantifier ; elles peuvent aider à relever les particularités des situations et des phénomènes, là où les méthodes quantitatives ont tendance à calculer des moyennes et des tendances générales ; et elles peuvent contribuer, grâce à la démarche inductive, à formuler des hypothèses et des questions évaluatives plus pertinentes, que les méthodes quantitatives peuvent ensuite tester sur de plus larges échantillons d'observations.

Au sein des méthodes mixtes, les méthodes qualitatives et les méthodes quantitatives se retrouvent véritablement articulées et peuvent davantage se nourrir mutuellement. Cette articulation exige une attention particulière aux spécificités des méthodes combinées, ainsi qu'une attention à la gouvernance même de l'évaluation, puisqu'il s'agit de veiller à ce que l'évaluateur (ou l'équipe d'évaluateurs) dispose des expertises nécessaires sur les deux types de méthodes et leurs articulations. Nous avons présenté cinq exemples de méthodes qualitatives ou mixtes utilisées en évaluation, et nous les avons illustrées. Elles donnent à voir la diversité des apports des méthodes mixtes pour l'évaluation de l'impact des politiques publiques. Ainsi, et de façon schématique, une évaluation pourrait gagner à s'appuyer sur : (a) la méthode de cartographie conceptuelle, afin d'identifier les impacts potentiels et de formuler des questions évaluatives pertinentes ; (b) la méthode de l'analyse qualitative comparée afin d'identifier les facteurs potentiels de succès ou d'échec de l'intervention publique ; (c) les méthodes de l'analyse de contribution, de la reconstitution de processus ou de l'évaluation *faisabilité-impact*, afin d'explicitier et de comprendre l'enchaînement des mécanismes causaux ou leur juxtaposition. Les méthodes quantitatives strictes pourraient alors parachever l'évaluation en testant les hypothèses formulées, en quantifiant les impacts identifiés et en formulant des résultats plus généraux (*significatifs*). Les interprétations des résultats quantitatifs en seraient également facilitées, ainsi que la formulation de recommandations opérationnelles que les parties prenantes pourraient plus aisément s'approprier.

Notre introduction donnait l'exemple d'une évaluation de l'impact du nombre d'élèves sur la performance scolaire. Si une telle évaluation était strictement quantitative, elle laisserait de nombreuses questions sans réponse. En mobilisant également des méthodes qualitatives et surtout les méthodes mixtes présentées, l'évaluation pourrait être substantiellement enrichie : elle permettrait de mieux comprendre les mécanismes causaux qui lient la taille de la classe et la performance scolaire, d'apprécier la taille de classe optimale, d'identifier des situations particulières et des points de vigilance, ou encore d'apprécier les facteurs (comme l'expérience de l'enseignant) qui conditionnent cet impact. En interrogeant le terrain et les acteurs, l'évaluation s'enrichirait également de recommandations plus opérationnelles et mieux acceptées par les parties prenantes.

En définitive, les méthodes mixtes appellent à la plus grande rigueur scientifique. En combinant les méthodes quantitatives et qualitatives, elles permettent d'enrichir les évaluations de l'impact – ou de tout autre critère évaluatif – des politiques publiques, et d'envisager des ponts fructueux entre disciplines, entre chercheurs et entre acteurs de l'écosystème de l'évaluation.



BIBLIOGRAPHIE

- Aguilera T. et Chevalier T. (2021), « [Les méthodes mixtes pour la science politique. Apports, limites et propositions de stratégies de recherche](#) », *Revue française de science politique*, vol. 71, p. 365-389.
- Ahmed A. et Sil R. (2012), « When multi-method research subverts methodological pluralism. Or, why we still need single-method research », *Perspectives on Politics*, vol. 10(4), décembre, p. 935-953.
- Baïz A. (2018), [De l'innovation des instruments de politique publique. Développement d'une méthode de conception combinatoire autour d'un langage algorithmique et application au dispositif des certificats d'économie d'énergie](#), thèse de doctorat, université Paris Sciences et Lettres.
- Baïz A. et Nakhla M. (2018), « [Pour une approche algorithmique de la nature protéiforme et fractale des instruments de l'action collective](#) », *Politiques et management public*, vol. 35(3-4), p. 153-172.
- Bandiera O., Larcinese V. et Rasul I. (2010), « [Heterogeneous class size effects: New evidence from a panel of university students](#) », *The Economic Journal*, vol. 120(549), décembre, p. 1365-1398.
- Beach D. et Pedersen R. B. (2013), [Process-Tracing Methods: Foundations and Guidelines](#), Ann Arbor, University of Michigan Press.
- Becker H. S. (2002), *Les Ficelles du métier*, Paris, La Découverte.
- Blackman T. (2013), « Exploring explanations for local reductions in teenage pregnancy rates in England: an approach using Qualitative Comparative Analysis », *Social Policy & Society*, vol. 12(1), p. 61-72.
- Bono P.-H., Debu S., Desplatz R., Hayet M., Lacouette-Fougère C. et Trannoy A. (2018), « [Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique](#) », *Document de travail*, n° 2018-06, France Stratégie, décembre.
- Burch P. et Heinrich C. J. (2016), *Mixed Methods for Policy Research and Program Evaluation*, Los Angeles, Sage.

- Bono P.-H. et Trannoy A. (2012), « [Causalité et contrefactualité dans l'évaluation économique des politiques publiques](#) », *Labyrinthe*, vol. 39, p. 35-53.
- Cairney P. (2016), *The Politics of Evidence-Based Policy Making*, Basingstoke, Palgrave Macmillan.
- Chanut V. (2007), *Misère de l'évaluation en ressources humaines*, Paris, Eska.
- Collier D. (2011), « [Understanding process tracing](#) », *Political Science & Politics*, vol. 44(4), octobre, p. 823-830.
- Commission européenne/Euréval (2011), *Étude portant sur la contribution de la traduction à la société multilingue dans l'Union européenne*, novembre.
- Delahais T. et Toulemonde J. (2017), « Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management? », *Evaluation*, vol. 23(4), octobre, p. 370-388.
- Desplat R. et Ferracci M. (2016), *Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et des praticiens*, France Stratégie, septembre.
- Divay S. (2012), « [Petit précis méthodologique d'évaluation qualitative](#) », *Éducation et formations*, vol. 81, mars, p. 49-54.
- Donaldson S. I., Christie C. A. et Mark M. M. (dir.) (2009), *What Counts as Credible Evidence in Applied Research and Evaluation Practice?*, Los Angeles, Sage.
- Fetterman D. M., Rodríguez-Campos L., Zukoski A. P. et al. (2018), *Collaborative, Participatory, and Empowerment Evaluation: Stakeholder Involvement Approaches*, New York, Guilford Press.
- Glaser B. G. et Strauss A. (2017), *La Découverte de la théorie ancrée. Stratégies pour la recherche qualitative*, Paris, Armand Colin.
- Greene J. C., Benjamin L. et Goodyear L. (2001), « [The merits of mixing methods in evaluation](#) », *Evaluation*, vol. 7(1), janvier, p. 25-44.
- Guba E. G. et Lincoln Y. S. (1994), « Competing paradigms in qualitative research », in Denzin N. K. et Lincoln Y. S. (dir.), *Handbook of Qualitative Research*, Los Angeles, Sage, p. 105-117.
- Head B. W. (2016), « [Toward more "evidence-informed" policy making?](#) », *Public Administration Review*, vol. 76(3), mai-juin, p. 472-484.
- Head B. W. (2010), « Reconsidering evidence-based policy: Key issues and challenges », *Policy and Society*, vol. 29(2), mai, p. 77-94.
- Hourcade R. et Cortinas Muñoz J. (2021), « [La mise en œuvre des politiques de protection de l'eau. Méthode et apports d'une recherche combinant QCA et process tracing](#) », *Revue française de science politique*, vol. 71, p. 413-435.
- Inspection générale des finances et Conseil général de l'environnement et du développement durable (2018), *L'avenir des opérateurs de l'eau et de la biodiversité*, avril.

- King G., Keohane R. O et Verba S. (1994), *Designing Social Inquiry. Scientific Inference in Qualitative Research*, Princeton, Princeton University Press.
- Lijphart A. (1972), « Comparative politics and the comparative method », *American Political Science Review*, vol. 64(3), p. 682-693.
- Lortie-Forgues H. et Inglis M. (2019), « Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? », *Educational Researcher*, vol. 48(3), avril, p. 158-166.
- Mahoney J. et Goertz G. (2006), « [A tale of two cultures: Contrasting quantitative and qualitative research](#) », *Political Analysis*, vol. 14(3), p. 227-249.
- Maxwell J. A. (2012), *A Realist Approach for Qualitative Research*, Londres, Sage.
- Mayne J. (2012), « [Contribution analysis: Coming of age?](#) » *Evaluation*, vol. 18(3), juillet, p. 270-280.
- McLaughlin J. A. et Jordan G. B. (2015), « Using logic models », in Newcomer K. E., Hatry H. P. et Wholey J. S. (dir.), *Handbook of Practical Program Evaluation*, 4^e éd., San Francisco, John Wiley and Sons, p. 62-87.
- Mohr L. B. (1999), « The qualitative method of impact analysis », *American Journal of Evaluation*, vol. 20(1), mars, p. 69-84.
- Palier B. et Trampusch C. (2018), « [Comment retracer les mécanismes causaux. Les différents usages du process tracing](#) », *Revue française de science politique*, vol. 68, p. 967-990.
- Patton M. Q. (2004), *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*, Londres, Sage.
- Pattyn V., Molenveld A. et Befani B. (2019), « [Qualitative comparative analysis as an evaluation tool: Lessons from an application in development cooperation](#) », *American Journal of Evaluation*, vol. 40(1), p. 55-74.
- Pawson R. (2004), « Would Campbell be a member of the Campbell collaboration? », *The Evaluator*, vol. 13, hiver.
- Pawson R. et Tilley N. (1997), « An introduction to scientific realist evaluation », in Chelimsky E. et Shadish W. R. (dir.), *Evaluation for the 21st Century: A Handbook*, Los Angeles, Sage, p. 405-418.
- Piketty T. et Valdenaire M. (2006), « [L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français. Estimations à partir du panel primaire 1997 et du panel secondaire 1995](#) », *Les Dossiers de l'Éducation nationale*, n° 173, Direction de l'évaluation et de la prospective, mars.
- Pin C. et Barone C. (2021), « [L'apport des méthodes mixtes à l'évaluation. Une expérimentation randomisée dans le champ des politiques éducatives](#) », *Revue française de science politique*, vol. 71, p. 391-412.

- Robert É. et Ridde V. (2013), « L'approche réaliste pour l'évaluation de programmes et la revue systématique : de la théorie à la pratique », *Mesure et évaluation en éducation*, vol. 36(3), p. 79-108.
- Revillard A. (2018), « [Saisir les conséquences d'une politique à partir de ses ressortissants : la réception de l'action publique](#) », *Revue française de science politique*, vol. 68(3), p. 469-492.
- Rogers P. J. (2008), « [Using programme theory to evaluate complicated and complex aspects of intervention](#) », *Evaluation*, vol. 14(1), janvier, p. 29-48.
- Scriven M. (1972), « Prose and cons about goal-free evaluation », *Evaluation Comment*, vol. 4(3), p. 1-7.
- Small M. L. (2011), « How to conduct a mixed methods study: Recent trends in a rapidly growing literature », *Annual Review of Sociology*, vol. 37, août, p. 57-86.
- Stame N. (2009), « [Avantages et inconvénients des différentes méthodes d'évaluation : comment choisir ?](#) », in Trosa S. (dir.), *Évaluer les politiques publiques pour améliorer l'action publique. Une perspective internationale*, Institut de la gestion publique et du développement économique, p. 33-55.
- Stedman-Bryce G. (2013), *Health for All: Towards Free Universal Health Care in Ghana. End of Campaign Evaluation Report*, Oxford, Oxfam GB.
- Strauss A. et Corbin J. (2003), « L'analyse de données selon la *grounded theory*. Procédures de codage et critères d'évaluation », in Cefaï D. (dir.), *L'Enquête de terrain*, Paris, La Découverte, p. 363-379.
- Suchman E. A. (1967), *Evaluative Research. Principles and Practice in Public Service and Social Action Programs*, New York, Russel Sage Foundation.
- Surel Y. (2018), « [La mécanique de l'action publique. Le process tracing dans l'analyse des politiques publiques](#) », *Revue française de science politique*, vol. 68, p. 991-1014.
- Tashakkori A. et Teddlie C. (2010), *SAGE Handbook of Mixed Methods in Social and Behavioral Research*, Los Angeles, Sage.
- Trampusch C. et Palier B. (2016), « Between X and Y: how process tracing contributes to opening the black box of causality », *New Political Economy*, vol. 21(5), p. 437-454.
- Trochim W. (1989), « [An introduction to concept mapping for planning and evaluation](#) », *Evaluation and Program Planning*, vol. 12(1), p. 1-16.
- Weiss C. H. (1997), « How can theory-based evaluation make greater headway? », *Evaluation Review*, vol. 21(4), août, p. 501-524.
- White H. (2013), « The use of mixed methods in randomized control trials », *New Directions for Evaluation*, vol. 2013(138), juin, p. 61-73.
- Yin R. K. (1999), « [Enhancing the quality of case studies in health services research](#) », *Health Services Research*, vol. 34(5 - Pt 2), décembre, p. 1209-1224



Directeur de la publication

Gilles de Margerie, commissaire général

Directeur de la rédaction

Cédric Audenis, commissaire général adjoint

Secrétaires de rédaction

Olivier de Broca, Gladys Caré

Contact presse

Matthias Le Fur, directeur du service Édition/Communication/Événements

01 42 75 61 37, matthias.lefur@strategie.gouv.fr

RETROUVEZ LES DERNIÈRES ACTUALITÉS DE FRANCE STRATÉGIE SUR :



Ce guide est publié sous la responsabilité éditoriale du commissaire général de France Stratégie. Les opinions exprimées engagent leurs auteurs et n'ont pas vocation à refléter la position du gouvernement.



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*



FRANCE STRATÉGIE

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