

# **ACTION** N° 21 **PUBLIQUE**

Recherche  
et pratiques

### Comité scientifique

**Natacha Gally**, maîtresse de conférences en science politique, Université Paris 2 Panthéon-Assas.

**Philippe Ledenvic**, membre permanent de l'Inspection générale de l'environnement et du développement durable.

**Adèle Lieber**, déléguée aux relations internationales, Direction générale des finances publiques.

**Frédérique Pallez**, professeur honoraire à l'École des Mines de Paris et chercheuse au Centre de gestion scientifique.

**Pierre-Louis Rolle**, directeur stratégie et innovation, Agence nationale de la cohésion des territoires.

**Emilien Ruiz**, historien, professeur assistant à Sciences Po Paris.

**Jérémie Vallet**, adjoint à la Directrice interministérielle du numérique, chef du département Appui, conseil et expertise, Direction interministérielle du numérique.

#### Directrice de la publication

Marie Niedergang

#### Rédactrice en chef

Marie Ruault

#### Co-rédacteur en chef

Edoardo Ferlazzo

#### Secrétaire de rédaction

Delphine Mantiene

#### Design graphique

Studio graphique des ministères économiques et financiers (SG-Sircom)

#### Mise en page et maquettage

Desk

Publication trimestrielle en accès libre – ISSN 2647-3135  
(en ligne)

Si vous souhaitez consulter les numéros de la revue *Action publique. Recherche et pratiques*,

rendez-vous sur le site  
<https://www.cairn.info/revue-action-publique-recherche-et-pratiques.htm>.

Pour vous inscrire à notre liste de diffusion :  
[recherche.igpde@finances.gouv.fr](mailto:recherche.igpde@finances.gouv.fr)

Réseaux sociaux :

Twitter : [@Igpderecherche](https://twitter.com/Igpderecherche)

[Youtube.com/igpde](https://www.youtube.com/channel/UC...)

S'abonner à la revue :



# Sommaire

## L'action publique à l'épreuve des preuves

- 5 **Éditorial**  
MARIE NIEDERGANG
- 9 **[Regards croisés] Expérimenter pour construire l'action publique : méthodes, apports et limites**  
ENTRETIEN ENTRE JÉRÔME GAUTIÉ ET SÉBASTIEN MALANGEAU
- 23 **[Étude] Comprendre les comportements pour élaborer des politiques publiques efficaces**  
NICOLAS JACQUEMET
- 33 **[Étude] Quand les « meilleures » preuves échouent : de l'usage politique des méta-analyses pour légitimer une politique de lutte contre le redoublement en Belgique francophone**  
CAROLINE DE PASCALE, HUGUES DRAELANTS ET JO HELGETUN
- 49 **[Note réactive] États-Unis : l'échelle de preuves, un nouvel outil pour justifier les décisions budgétaires**  
AMICIE DE TANOÛARN
- 55 **[L'Œil du chercheur] Revue d'articles et de thèses**



# Éditorial

MARIE NIEDERGANG



**Marie Niedergang**  
Directrice générale  
de l'Institut  
de la gestion publique  
et du développement  
économique

Sur quoi les décideurs publics sont-ils susceptibles de s'appuyer pour faire les meilleurs choix, prendre des décisions et les mettre en œuvre ? Différentes méthodes scientifiques destinées à guider les pratiques se sont développées depuis plusieurs années. Certaines, comme les expérimentations ou les méta-analyses, occupent une place importante et semblent s'être imposées parce que susceptibles de « faire preuve », d'établir une forme de vérité à partir de laquelle orienter et construire les politiques publiques. Ces méthodes, intéressantes et constructives, posent néanmoins de nombreuses questions : sur leurs caractéristiques propres, sur leurs finalités, sur l'usage qui en est fait...

L'objet de ce numéro est d'aborder ces questionnements et de les mettre en perspective. Dans le cadre d'un *Regards croisés*, **Jérôme Gautié**, professeur d'économie, et **Sébastien Malangeau**, directeur du programme France Expérimentation, évoquent plusieurs cas qui font ressortir à la fois les finalités des différentes modalités d'expérimentation et les enseignements qu'elles portent pour l'action. **Nicolas Jacquemet**, professeur d'économie, dresse dans un premier article un état des lieux de l'usage des sciences comportementales en économie et de leurs apports, notamment en matière de lutte contre la fraude fiscale. Dans un second article co-écrit par **Caroline De Pascale**, **Hugues Draelants** et **Jo Helgetun**, chercheurs en sociologie de l'éducation, est analysée la mobilisation des méta-analyses pour conduire la politique du redoublement en Belgique. Enfin, la *Note réactive* rédigée par **Amicie de Tanoüarn**, apprentie au sein du bureau de la Recherche de l'IGPDE, présente un outil d'échelle de preuves mis en place aux États-Unis pour appuyer et justifier des choix budgétaires.

Très bonne lecture !



# Regards croisés

## Entre recherche et pratiques

Les *Regards croisés* reposent sur un dialogue organisé entre une personne issue du monde académique et universitaire et une personne issue de l'administration publique sur un sujet d'intérêt commun.

Ce dialogue est animé dans le cadre d'une interview vidéo publiée sur la chaîne YouTube de l'IGPDE. Cette interview est également retranscrite et remaniée sous la forme d'un article publié dans cette revue.



# Expérimenter pour construire l'action publique : méthodes, apports et limites

Entretien entre Jérôme Gautié et Sébastien Malangeau<sup>1</sup>



**Jérôme Gautié** est professeur d'économie à l'université Paris 1 Panthéon-Sorbonne.

**Sébastien Malangeau** est directeur du programme France Expérimentation\*.

Retrouvez cet entretien en vidéo sur le site de la revue

[www.economie.gouv.fr/jgpde-editions-publications/action-publique-recherche-pratiques](http://www.economie.gouv.fr/jgpde-editions-publications/action-publique-recherche-pratiques)

## De quoi parle-t-on lorsqu'on parle d'expérimentation ?

**Jérôme Gautié** : Une première acception, assez large, du terme d'expérimentation renverrait à ce que j'appellerais l'« expérimentation sociale ». Rentrent dans ce cadre des initiatives d'acteurs

locaux qui ont des idées de changement de la législation ou de nouveaux dispositifs à déployer qui constituent une *innovation sociale*. Par exemple, dans le domaine de la politique de l'emploi, « Territoires zéro chômeur de longue durée » était l'initiative d'un acteur privé associatif (ATD Quart Monde), et a été expérimentée sur plusieurs territoires avant d'être déployée plus largement.

<sup>1</sup> Cet entretien a été animé par Marie Ruault, directrice de la Recherche à l'IGPDE. Il a été enregistré le 6 mars 2024.

\* France Expérimentation est un dispositif interministériel porté par la Direction interministérielle de la Transformation publique (DITP) et la Direction générale des Entreprises (DGE).

Dans un sens plus restreint, il peut s'agir de l'expérimentation « scientifique » dans le cadre de ce qu'on appelle l'évaluation d'impact. Et là, il s'agit vraiment d'établir scientifiquement si un dispositif marche ou pas, en examinant notamment ses effets. L'étalon-or de ce type d'expérimentation, c'est l'expérimentation contrôlée aléatoire où l'on détermine à la fois un groupe de traitement et un groupe témoin. Le groupe de traitement est celui qui entrera véritablement dans l'expérimentation et qui, notamment, bénéficiera de la mesure que l'on veut tester. Le groupe témoin, identique par ses caractéristiques, ne bénéficiera pas de la mesure. Ce type d'expérimentation se fait dans le domaine économique et social, un peu sur le modèle de ce qui se fait en médecine depuis très longtemps pour tester l'efficacité des médicaments, notamment.

**Sébastien Malangeau :** Avec France Expérimentation, on se situe dans le premier cas de figure, celui des expérimentations sociales. France Expérimentation est un programme interministériel qui vise à permettre à tout acteur portant un projet innovant, quel que soit son domaine – innovation technologique, organisationnelle, sociale... –, et qui estime que ce projet n'est pas faisable en l'état du droit, de solliciter les pouvoirs publics pour demander à ce que le droit soit adapté afin de permettre le développement de ce projet. Parmi les solutions mobilisables, il y a notamment l'expérimentation. C'est bien une demande du terrain qui est rendue possible par une adaptation du droit à l'initiative des pouvoirs publics. Ce dispositif a été lancé dans une première version en 2016, puis relancé en 2018 avec un nouveau positionnement sous l'autorité du Premier ministre, France Expérimentation étant placée sous l'autorité du délégué interministériel à la transformation publique et animée par une équipe DGE<sup>2</sup> – DITP. Le délégué interministériel est le représentant du Premier ministre, et les décisions sont prises pour le Premier ministre par son cabinet. Les acteurs de terrain sollicitent une adaptation du droit qui peut toucher à des normes réglementaires ou même à des normes législatives, les ministères instruisent les demandes et *in fine* formulent des recommandations. Au sommet de la pyramide, le cabinet du Premier ministre arbitre les solutions à donner à chacune de ces demandes. Ce dispositif très institutionnel de l'État a l'originalité d'être *bottom-up* : ce sont vraiment les acteurs de terrain qui font remonter leurs besoins auprès des

pouvoirs publics, et non les pouvoirs publics qui préjugent des attentes ou des besoins d'adaptation des innovateurs en France.

## Pourquoi l'action publique déploie-t-elle ce type d'approche ?

**Jérôme Gautié :** L'expérimentation que j'ai qualifiée de « scientifique » s'inscrit dans une démarche évaluative et vise à tester de nouveaux dispositifs. Dans la plupart des cas, c'est l'État qui prend l'initiative. La première expérimentation de ce type a eu lieu en France dans les années 2006-2007, lorsqu'il a fallu comparer la performance relative des opérateurs privés de placement et de l'ANPE pour les programmes d'accompagnement renforcé. À l'époque, il y avait un peu de tension entre l'ANPE et l'Unédic, avec l'accusation au moins implicite par l'Unédic de certaines inefficacités de l'ANPE. Le juge de paix, pour ainsi dire, a été une expérimentation où l'on a fait appliquer le même type de dispositif à des opérateurs privés de placement et à l'ANPE. Les résultats étaient comparés à ceux d'un troisième groupe, le groupe témoin de personnes qui bénéficiaient simplement de l'accompagnement « standard » de l'ANPE. Il s'agissait de recueillir des éléments de *hard evidence*, c'est-à-dire des preuves « solides », dans un processus de décision publique où l'on veut savoir ce qui marche et ce qui ne marche pas.

**Sébastien Malangeau :** Côté France Expérimentation, le cadre est forcément plus restreint. Il est lié au mandat du dispositif, qui est vraiment de favoriser l'innovation en France, avec aussi une dimension de stimulation de l'économie. Il s'agit de faire émerger des acteurs français innovants qui auraient vocation à devenir des leaders mondiaux. L'idée est vraiment de lever les obstacles juridiques, et l'expérimentation est l'un des outils mobilisables dans le cadre de ce dispositif. Dans le domaine de l'innovation, les pouvoirs publics sont confrontés à la contrainte de trouver le juste équilibre entre la souplesse nécessaire à l'innovation, et la mise en place d'un cadre juridique sécurisant permettant de maîtriser les risques. C'est toujours un exercice délicat quand on est dans une *terra incognita*... L'expérimentation juridique est là pour tester davantage l'adaptation et la proportionnalité du cadre juridique que l'objet

<sup>2</sup> La Direction générale des Entreprises (DGE) est rattachée au ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique.

même de l'expérimentation. On veut s'assurer que le cadre juridique mis en place par les pouvoirs publics sera le plus proportionné et le moins contraignant possible, et cependant suffisamment sécurisant. L'expérimentation a pour objet de tester le futur cadre définitif sur la base des enseignements tirés de la phase expérimentale. On inverse ainsi la logique traditionnelle, descendante, où les pouvoirs publics anticipent des évolutions et les cadrent selon ce qui leur paraît les critères les plus objectifs, par l'entremise des corps intermédiaires qui vont faire remonter ce qu'ils estiment être les besoins des acteurs et de la société. À l'inverse, on adopte ici une logique un peu différente, peut-être plus contemporaine, où ce sont les acteurs de terrain souhaitant très concrètement développer leur projet qui cartographient les obstacles et les font remonter à l'État. Ce dernier se met finalement dans une posture de récepteur, en analysant les demandes et en s'efforçant de trouver des solutions pour adapter le droit aux attentes des innovateurs.

## Pourquoi cette idée de disposer d'évaluations basées sur des preuves s'est-elle imposée ?

**Jérôme Gautié** : Il faut effectivement la resituer, parce que l'histoire est récente en France. J'évoquais les années 2006-2007 comme point de départ de cette expérience. C'est beaucoup plus ancien dans les pays anglo-saxons, notamment aux États-Unis, où la culture de l'évaluation systématique de chaque politique publique, avec des protocoles très formalisés et sur une base scientifique, est très ancrée. Aux États-Unis, il y a eu des expérimentations dès les années 1960, et certaines très importantes au début des années 1970 dans le domaine des politiques de l'emploi et des politiques sociales. Pourquoi les États-Unis ? Parce que c'est un pays où les clivages politiques sont très forts dans certains domaines, notamment pour tout ce qui touche au travail et aux politiques sociales. Le recours à l'expertise scientifique permet parfois de trancher ces débats qui ont des contenus idéologiques très forts. Un sociologue américain, Harold Wilensky, avait établi ce paradoxe : les États-Unis sont à la fois les moins généreux en termes de politique de

l'emploi et de politique sociale, mais c'est le pays où l'on évalue le plus ces dispositifs. À l'opposé, à l'époque, dans des pays comme la Suède, on dépensait beaucoup pour des politiques sociales et des politiques de l'emploi, mais on évaluait assez peu. Pourquoi ? Parce qu'en Suède, ces politiques étaient relativement consensuelles dans un spectre politique assez large, alors qu'aux États-Unis, la moindre mesure de politique sociale ou de politique de l'emploi, ou du marché du travail (comme le salaire minimum), est en soi beaucoup plus conflictuelle. En France, on a importé ces méthodes, parce qu'elles se sont beaucoup développées aux États-Unis. C'est devenu un peu aussi à la mode dans le champ académique. Esther Duflo, une économiste française qui a obtenu le prix Nobel<sup>3</sup>, a notamment porté ces méthodes au plus haut niveau scientifique avec d'autres. Dans le cadre de la modernisation de l'action publique et du fait d'une volonté de légitimer davantage l'action publique, l'idée d'importer cette culture de l'évaluation et ces techniques, qui font l'objet d'un consensus assez fort parmi les économistes, s'est peu à peu imposée.

## Quelles sont les attentes poursuivies par les décideurs publics à travers la démarche expérimentale ?

**Jérôme Gautié** : Je pense que les attentes des uns et des autres peuvent différer, mais il me semble important d'embarquer tous les acteurs susceptibles d'intervenir dans le processus dont on parle. Pour la Garantie jeunes comme pour beaucoup de mesures mises en œuvre et expérimentées dans le domaine des politiques de l'emploi, un conseil scientifique est mis en place. C'est plutôt lui qui formule auprès des décideurs publics la préconisation issue de l'expérience mise en œuvre. Ce conseil scientifique comprend évidemment des membres de la communauté « scientifique » (économistes, sociologues...), mais aussi des membres de toutes les administrations potentiellement concernées, et plus largement de l'ensemble des parties prenantes. Par exemple, dans le cas de la Garantie jeunes, étaient représentées non seulement les administrations du travail, de l'emploi, mais aussi notamment la direction de la cohésion sociale, les différentes administrations de

<sup>3</sup> Le prix de la Banque de Suède en sciences économiques en mémoire d'Alfred Nobel a été décerné en 2019 à Abhijit Banerjee, Esther Duflo et Michael Kremer.

la jeunesse, sans oublier les bénéficiaires, par la voix d'associations de jeunes. Il y avait enfin des représentants des missions locales – qui mettaient en œuvre la mesure. Quand on recourt à l'expérimentation, il me semble essentiel de dire comment on la conçoit et comment on l'utilise. Je crois qu'il ne faut pas tomber dans une vision trop scientifique de l'expérimentation, même quand on parle de l'expérimentation scientifique/évaluation d'impact. Parfois, certains collègues ont un peu cette vision : « La science a dit que... », et c'est une espèce de « vérité révélée », fondée sur des faits objectifs et des preuves solides, qui doit s'imposer sans débat à tous les acteurs et qui doit dicter la prise de décision publique. Ce n'est pas si simple que cela, parce que l'expérimentation elle-même peut être plus ou moins « solide » d'un point de vue méthodologique (ce qui renvoie à sa « validité interne »), mais surtout, elle peut ne valoir que dans le contexte – temporel et spatial – dans lequel elle a été déployée, ce qui pose la question de sa « validité externe » : on n'est pas sûr que ses résultats seront valables dans un autre contexte. Avant d'être trop assertif, il faut avoir beaucoup d'éléments de « preuves ». Par exemple, pour la Garantie jeunes, d'autres évaluations ont été menées par la suite qui ont plutôt confirmé les résultats de la première expérimentation. Dans l'histoire des politiques publiques, notamment aux États-Unis, on a des exemples célèbres d'évaluations scientifiques qui avaient donné des avis très tranchés sur certains dispositifs qui, par la suite, ont été contredits par de nouvelles études, avec de nouvelles méthodes et de nouvelles données. Cela incite à beaucoup de modestie. Un des exemples les plus célèbres concerne le cas du salaire minimum aux États-Unis. Jusqu'au début des années 1980, il y avait un très fort consensus selon lequel le salaire minimum détruisait des emplois. Les sondages montraient que plus de 90 % des économistes étaient persuadés que le salaire minimum était une mauvaise chose. Des chercheurs sont arrivés avec de nouvelles méthodes, de nouvelles données, etc., et ils ont renversé la table en montrant que non, finalement, cela n'avait pas d'effets si négatifs. David Card a été récompensé il y a quelques années par le prix Nobel pour ces travaux-là. Un autre exemple très célèbre, ce sont les travaux sur la peine de mort aux États-Unis. Quand la Cour suprême a changé sa doctrine concernant la peine de mort au début des années 1970 pour réautoriser la peine de mort au niveau des États, elle s'est appuyée notamment sur une étude économétrique qui aujourd'hui est

considérée comme très fragile d'un point de vue méthodologique. C'est un exemple extrême, mais qui souligne la responsabilité énorme qu'ont les scientifiques et qui devrait les rendre d'autant plus humbles. Il faut avoir cette modestie pour relativiser l'usage de l'expérimentation scientifique. Une évaluation donne lieu à un avis, une « évidence » parmi d'autres<sup>4</sup>, qui doit être confrontée à d'autres évidences, à d'autres éclairages. Finalement, la décision politique est un peu comme la décision du juge : c'est à partir de faisceaux d'indices que l'on décide de l'opportunité ou pas. Parfois, on peut ne pas aller dans le sens de l'évaluation scientifique pour toutes sortes de raisons : les raisons politiques, les pressions de lobbies, ou la crainte de l'impopularité... ou parce qu'on estime qu'il y a d'autres considérations à prendre en compte, ou que les éléments de preuve ne sont pas suffisamment probants pour que la décision publique aille dans le sens que préconiseraient les économistes.

**Sébastien Malangeau** : Sur les expérimentations juridiques, une des motivations que l'on peut relever, c'est peut-être une évolution due au contexte international, à ce qui a pu se passer avec la mondialisation qui s'est intensifiée dans les années 2000-2010, et puis aujourd'hui, je pense avec l'émergence d'acteurs vraiment mondiaux. La confrontation entre innovation et régulation par les pouvoirs publics existe de tout temps. Jusqu'à récemment, les acteurs économiques étaient, en termes de puissance, sérieusement inférieurs aux États. La question était alors de savoir si les pouvoirs publics pouvaient adapter le droit au rythme des innovateurs. S'ils ne le faisaient pas, cela bridait l'innovation. Désormais, certains acteurs ont la force économique pour tenter d'imposer leurs solutions partout dans le monde sans tenir compte du cadre juridique en place. Dans ce contexte, l'expérimentation est particulièrement pertinente car elle permet de tester des cadres qui ne sont pas forcément définitifs. Dès lors, la prise de décision et la conception du cadre peuvent se faire plus rapidement que si l'on attend d'avoir validé toutes les hypothèses. Un exemple me semble assez parlant dans notre domaine, c'est celui d'Uber. Quand Uber arrive en France, son activité est illégale. Pourtant, sa force de frappe fait qu'il arrive en disant : « Je vais concurrencer les taxis. Mon activité est nouvelle et je ne me préoccupe pas de savoir si elle est légale. » Uber est tellement fort que, *in fine*, les pouvoirs publics se retrouvent à négocier, probablement dans une position d'infériorité relative avec un très gros acteur, un nouveau

4 En anglais, *evidence* signifie « preuve ».

cadre. Dans ce type de scénario, la question qui peut se poser, c'est : même si l'on n'a pas de certitude sur le cadre à déployer, ne vaut-il pas mieux anticiper avec des acteurs un cadre intermédiaire coconstruit, mais sur lequel les pouvoirs publics peuvent garder la main, plutôt que de dire « Non, on ne change pas le cadre tant qu'on n'a pas une vision parfaitement claire » ? Et peut-être alors que demain, le marché aura déjà décidé pour les pouvoirs publics. C'est la même chose pour Booking et Airbnb : on le voit, ces acteurs ont émergé sur des absences de réglementation. L'expérimentation, en tout cas dans le domaine juridique, revient aussi à assumer, dans une posture d'humilité, que l'on peut tester des propositions sans les avoir validées totalement, pour s'assurer que l'on suit le rythme susceptible de nous être imposé par des acteurs externes. C'est enfin un très gros enjeu d'acceptabilité. L'expérimentation ouvre en effet la palette des solutions à la disposition des pouvoirs publics pour renforcer l'acceptabilité, puisque la décision, qui n'est pas présentée comme définitive, sera véritablement objectivée. C'est au fond toutefois un peu théorique, parce qu'aucune décision des pouvoirs publics n'est irréversible, mais cela inscrit, je pense, la mesure testée dans une perspective d'évaluation et d'ajustement en fonction des retours du terrain.

## Pourriez-vous décrire la fabrique d'une expérimentation à but évaluatif ?

**Jérôme Gautié** : Dans le domaine des politiques de l'emploi, je reprendrai ici l'exemple emblématique de la Garantie jeunes qui a été la mesure phare des dix dernières années en faveur des jeunes les plus vulnérables sur le marché du travail. Elle a de surcroît fait l'objet d'un lourd protocole d'évaluation qui s'est étalé sur cinq ans à partir de 2013. Ce qui est également intéressant avec la Garantie jeunes, c'était qu'elle avait ces deux dimensions : à la fois une expérimentation institutionnelle et une expérimentation au sens plus scientifique du terme, d'évaluation d'impact. Sur le plan institutionnel, l'objectif n'était pas seulement d'avoir un nouveau dispositif d'accompagnement des jeunes et de voir dans quelle mesure il marchait ou ne marchait pas, ou était plus efficace qu'un autre, mais c'était aussi de changer les pratiques des opérateurs publics qui mettaient en œuvre ce dispositif, à savoir les missions locales. C'est-à-dire que ce dispositif avait pour ambition de faire opérer des changements de pratiques institutionnelles

chez certains opérateurs. En quoi consistait-il ? Les grands principes de la Garantie jeunes, c'était d'abord de s'adresser à un public particulièrement vulnérable qu'on appelle les *Neet* (pour *Not in education, employment or training* : « Ni en formation, ni employé, ni stagiaire ») – c'était la première fois qu'on utilisait ce concept comme catégorie administrative de ciblage de la politique publique. Le dispositif reposait aussi sur différents principes assez novateurs en termes d'accompagnement : d'une part, articuler accompagnement collectif et accompagnement individuel, mais un accompagnement collectif *en cohorte*. C'était très original. Ce n'est pas seulement le fait que le jeune participait à des ateliers collectifs classiques dans les missions locales ou à Pôle emploi, mais qu'il y entrait par cohorte de 15 à 20 jeunes, tous suivis pendant un an avec des ateliers collectifs pour créer un esprit d'équipe. D'autre part, cet accompagnement renforcé reposait sur le principe du *work first*, qui remettait en cause le processus séquentiel classique de l'accompagnement : on commence par faire le diagnostic du jeune et de ses difficultés, en déterminant les freins d'accès à l'emploi, on essaye de traiter ces difficultés (et notamment d'élaborer un projet professionnel), et ensuite on accompagne vers l'emploi. Dans la nouvelle conception, on adopte un processus itératif : le projet et le diagnostic se construisent en itérant des phases en expérience de travail, avec le soutien de la mission locale d'accompagnement. Dès qu'on pouvait mettre le jeune en expérience de travail, on le faisait sous des statuts juridiques différents. Et le jeune, fort de cette expérience, revenait à la mission locale pour un débrief. C'est à partir de là qu'on voyait à la fois ses lacunes en termes de compétences, mais aussi son projet professionnel qui s'élaborait. Un autre pilier était l'« intermédiation active », c'est-à-dire que l'agent qui suivait ce jeune était supposé le faire non seulement dans les phases d'accompagnement à la mission locale, mais aussi en emploi pour capitaliser son expérience et faciliter l'accompagnement par l'employeur, en préparant l'employeur à l'accueil de ce type de public qui, parfois, pouvait être désocialisé ou particulièrement fragile. Enfin, le quatrième pilier était une allocation à peu près du montant du RSA, pour donner au jeune sa capacité d'agir, et conçue comme un soutien à sa recherche d'emploi plutôt que comme une aide sociale. Côté expérimentation scientifique, on a choisi un certain nombre de territoires où a été appliquée cette mesure et on a pris des territoires témoins où l'on a regardé la même population. Ont ainsi été comparés les jeunes bénéficiaires dans les territoires de traitement et les populations équivalentes dans les territoires témoins qui ne

bénéficiaient pas encore de la mesure, mais qui auraient été éligibles à cette mesure. On a regardé ce qui se passait au bout d'un certain nombre de mois, avec des interrogations à 6 mois, à 12 mois, à 18 mois. Et on s'est aperçu, en comparant le groupe de traitement et le groupe témoin, que ce type d'accompagnement avait eu un effet positif sur le taux de retour à l'emploi des jeunes bénéficiaires de la Garantie jeunes. Cette évaluation quantitative à partir de critères classiques comme le taux d'emploi, la vitesse du retour à l'emploi, etc., a été complétée par des indicateurs sur le bien-être des jeunes sur un plan plus qualitatif.

## Quels étaient les acteurs impliqués pour l'évaluation de la Garantie Jeunes ?

**Jérôme Gautié :** À la différence des expérimentations juridiques qui partent des acteurs locaux, les innovations sociales sont impulsées par l'État lui-même, même si la volonté naît des retours du terrain. En l'occurrence, l'initiative partait de la Direction générale de l'emploi et de la formation professionnelle (DGEFP). C'est donc l'État qui souhaite innover, sous forme d'expérimentation institutionnelle. Mais l'État, c'est beaucoup d'acteurs autour de la table. De là, la mesure est un peu une mesure « kit ». Elle n'est pas définie de façon parfaite au niveau central pour s'appliquer de façon uniforme sur le terrain. C'est plutôt un ensemble d'outils mis à disposition des acteurs locaux chargés de mettre en œuvre la mesure. Évidemment, il y a la mission locale qui, très concrètement, est en charge, mais il y avait aussi des commissions au niveau local animées la plupart du temps par le préfet ou, à l'époque, par la Direction régionale du travail, de la formation professionnelle et de l'emploi ou sa branche départementale. Ces commissions réunissaient tous les acteurs étatiques et associatifs qui pouvaient être en contact avec des jeunes potentiellement bénéficiaires et qui devaient à la fois les orienter vers la mesure, les sélectionner et aussi faire un peu de suivi, parce que le jeune pouvait être « renvoyé » s'il ne suivait pas un peu le protocole et, notamment, s'il ne respectait pas le contrat d'engagement qu'il signait en entrant dans la mesure. Beaucoup d'acteurs étaient mobilisés. Et évidemment, le dispositif que l'on peut qualifier de « Garantie jeunes », comme sa mise en œuvre, étaient en fait différents d'un territoire à l'autre. Tout dépendait de la façon dont les acteurs locaux s'étaient emparés d'un cadre commun, d'une boîte à outils commune, outillée par la DGEFP. On a ainsi constaté des

différences assez importantes de mise en œuvre par les acteurs locaux. Le cadrage de la DGEFP spécifiait qu'il devait y avoir une phase d'accompagnement collectif, une phase d'accompagnement individuel. Les phases d'accompagnement collectif étaient mises en œuvre de façon différente. Dans certaines missions locales, c'était six ou huit semaines d'affilée et ensuite, seulement de l'accompagnement individuel. Dans d'autres missions locales, on alternait accompagnement collectif et accompagnement individuel. Les ateliers n'étaient pas les mêmes... Il y avait une marge de manœuvre des acteurs, et là, on rejoint tout à fait l'aspect d'innovation sociale institutionnelle. C'était aussi aux acteurs locaux de construire eux-mêmes le dispositif. L'intérêt pour le décideur public, ce n'est pas seulement de savoir si le dispositif marche ou ne marche pas, mais de voir où il marche mieux ou moins bien, et d'essayer de comprendre pourquoi. Cela élargit aussi le spectre des questions.

## Comment se déploie la démarche du côté de France expérimentation ?

**Sébastien Malangeau :** Je pense que dans toute expérimentation, il y a une dimension plus ou moins prononcée de co-construction. L'idée, c'est bien de ne pas avoir une réponse binaire. Plus globalement, dans l'innovation sociale, c'est le *process* qui enrichit la version définitive si elle est appelée à s'appliquer. France Expérimentation, c'est un dispositif de l'État, mais qui n'a pas le monopole de l'expérimentation. Les ministères, les pouvoirs publics peuvent tout à fait décider de mettre en place des expérimentations sociales. Nous sommes là comme une boîte à outils, pour collecter les demandes des acteurs, avec cet état d'esprit que l'État par essence ne peut pas préjuger des innovations de demain, qui plus est de toutes natures : des évolutions sociétales peuvent entraîner des demandes d'adaptation en matière sociale, de nouveaux types de cohabitation dans le logement social peuvent nécessiter d'adapter les critères... Cela peut être une innovation sociale, une innovation technique ou une innovation organisationnelle. On traite bien de dérogation juridique, c'est notre mandat, donc il faut vraiment qu'*a priori*, le projet soit bloqué par le droit. Nous avons un réseau de correspondants dans l'ensemble des ministères. Dès lors qu'un acteur innovant, quel qu'il soit – entreprise, particulier, collectivité territoriale, association... –, nous fait valoir que son projet innovant n'est pas, selon lui, faisable en l'état du droit, nous (c'est-à-dire

l'équipe France Expérimentation) réceptionnons ce projet et nous sollicitons tous les ministères intéressés par la thématique ; pas seulement le ou les ministères qui ont la main sur le blocage juridique allégué, mais aussi tous les ministères intéressés. En guise d'illustration, on a eu récemment une expérimentation, qui est toujours en cours, à l'initiative du groupe Novo Nordisk, le leader mondial du traitement du diabète. Cette entreprise faisait valoir qu'en France, à l'heure actuelle, le recyclage de la partie plastique des stylos à insuline était impossible, parce qu'à une époque plus lointaine et après un certain nombre de scandales ou de difficultés sanitaires, tous les dispositifs médicaux devaient être incinérés. Cet acteur est venu nous voir et nous a dit : « Moi, j'ai une méthode innovante pour recycler ces produits qui sont incinérés. C'est bon économiquement mais c'est surtout bon pour la société et pour l'environnement. Voilà le protocole que je peux mettre en place. J'aimerais que vous m'autorisiez à le tester. » C'est une proposition qui est faite. Nous saisissons donc les ministères concernés : les ministères de la Santé, de la Transition écologique, de l'Économie, pour que chacun puisse donner un avis en droit et aussi en opportunité. L'analyse en droit peut avoir une série de conséquences. La première conséquence possible, qui n'est pas négligeable dans le cadre de France Expérimentation, c'est que les ministères reviennent vers les acteurs en disant : « En réalité, votre projet est faisable dans l'état du droit. Tel que vous nous le décrivez, on peut vous sécuriser en vous indiquant quelles sont les dispositions de la réglementation actuelle qui vous permettent le développement de votre projet. » Si le blocage juridique est confirmé, chaque ministère va se prononcer en opportunité. Faut-il donner suite à la demande ? Éventuellement non, pour quelle raison ? Et si cela a du sens, quels seraient les scénarios envisageables ? Répondre exactement à la demande du porteur ? Faire une contre-proposition et dire : « Cette thématique nous intéresse, mais peut-être pas exactement à l'identique de vos propositions » ? Ou bien suggérer que le droit soit directement et définitivement modifié ? Si on est déjà parfaitement certains du résultat, alors que tous les acteurs en sont convaincus, ne faisons pas une expérimentation. Les ministères formalisent leurs avis et ces avis, ainsi que celui de l'équipe France Expérimentation, sont ensuite soumis au cabinet du Premier ministre qui va prendre la décision sur la suite à donner. L'expérimentation est une des voies de sortie. Si on modifie le cadre juridique de manière définitive, cela se fait de manière tout à fait traditionnelle. Si l'on va dans la voie de l'expérimentation, il y a deux processus en parallèle. Le premier, c'est la modification du

texte bloquant, selon les procédures classiques, ce qui n'est pas sans limite en termes de temps. Par exemple, s'il faut un décret modificatif en Conseil d'État, s'il faut un article de loi qui modifie l'article de loi bloquant, cela peut être long et assez complexe. En parallèle est défini le protocole expérimental. L'idée est bien de le coconstruire, évidemment avec l'acteur à l'origine de la demande, mais pas exclusivement, puisque les pouvoirs publics se placent dans une perspective globale. Il y a une proposition de départ, mais si d'autres acteurs arrivent ou si l'État estime que le protocole devrait être adapté, à l'issue de cette discussion, un protocole expérimental est formalisé, qui va permettre d'assurer le bon suivi de l'expérimentation et d'en tirer des enseignements. Quant à la durée, c'est un élément important. Les expérimentations juridiques, en général, durent de 3 à 5 ans. Moins de 3 ans, compte tenu de la lenteur du processus, et pour faire des collectes, c'est un peu court. Plus de 5 ans, c'est plus complexe pour le juge, car le « temporaire qui dure » est un peu difficile à justifier. On estime qu'au-delà de 5 ans, si vraiment on veut tester quelque chose, on peut modifier la norme directement. Dans la sphère publique, il n'y a pas un domaine qui ne voit pas une nouvelle loi arriver tous les 2-3 ans. Donc si on allait au-delà de 5 ans d'expérimentation, autant passer par la voie de la modification directe et éventuellement de sa rectification.

## Quels sont les critères et les modalités de généralisation d'une expérimentation ?

**Jérôme Gautié** : Je reprends l'exemple de la Garantie jeunes. On voit que ça marche plus ou moins bien selon les endroits, mais on voit qu'en moyenne, ça marche. La question de la généralisation était aussi posée par les pouvoirs publics, qui ont d'ailleurs généralisé la mesure. On pourrait penser qu'une fois que les scientifiques ont parlé, la généralisation est automatique, et que si ça marche, on généralise, alors que si ça ne marche pas, on abandonne. C'est la logique du « stop ou encore », mais c'est en fait un peu plus compliqué que cela. Pourquoi ? Parce que même d'un point de vue scientifique, le problème qui se pose dans les évaluations avec expérimentations aléatoires qui sont les mieux faites, c'est le problème évoqué plus haut de ce qu'on appelle la validité externe. C'est-à-dire qu'après avoir testé la mesure dans un certain nombre de missions locales, sur un certain nombre de territoires, on a un résultat que l'on

pense solide, parce qu'on a bien respecté un protocole : c'est ce qu'on appelle la validité interne. En revanche, qu'est-ce qui se passerait si on changeait d'échelle, et notamment si on généralisait ? Deux problèmes se posent alors. Premièrement, les résultats qu'on a obtenus sont peut-être en partie à relier au contexte particulier des territoires sur lesquels on a déployé ces expérimentations, et/ou des spécificités des missions locales qui sont intervenues, etc. Et ça, c'est vraiment une différence entre les expérimentations dites scientifiques dans le domaine économique et social et dans le domaine de la médecine, par exemple. Quand un médicament est testé dans le cadre d'un protocole, avec un groupe témoin qui a un placebo, mais qui ne le sait pas, et un groupe de traitement qui a le vrai médicament, si les échantillons sont représentatifs de la population totale, si ça marche pour 5 000 personnes, il n'y a aucune raison que ça ne marche pas pour 60 millions de personnes. Dans le domaine économique et social, c'est différent. Il y a des spécificités des missions locales, des territoires, etc. Et la représentativité du contexte est beaucoup plus compliquée à établir, les expérimentateurs aléatoires en ont bien conscience. C'est quelque chose qui fait encore débat. Une des réponses consiste à dire : « Il faut multiplier les expérimentations dans le temps et dans l'espace. » Le contexte, ce n'est pas seulement le contexte géographique, cela peut être aussi un contexte conjoncturel. On voit toujours dans le domaine des politiques de l'emploi qu'une même mesure peut ne pas avoir les mêmes effets en bonne conjoncture du marché du travail et en mauvaise conjoncture du marché du travail – et pas seulement en termes d'ampleur de l'effet, mais parfois même dans le sens de l'effet. On peut très bien imaginer que ça joue positivement dans certains contextes conjoncturels et négativement dans d'autres. Ce n'est pas impossible, et c'est un premier problème. Un deuxième problème est lié plus directement au changement d'échelle lors de la généralisation. Quand vous faites une expérimentation, vous travaillez sur des groupes de bénéficiaires de taille limitée. Le principe, c'est que le traitement que vous imposez à certains bénéficiaires ne doit en aucun cas affecter la situation et la trajectoire des membres du groupe témoin. Sinon, cela fausse la comparaison, puisque le groupe témoin constitue ce qu'on appelle le contrefactuel (« Que se serait-il passé si la mesure n'avait pas été déployée ? » : le groupe témoin est là pour ça). Or, souvent, dans le cadre des politiques de l'emploi, on n'est pas trop sûrs. Le fait, sur un territoire donné, d'avoir accru l'employabilité de certains jeunes ne se fait-il pas au détriment des jeunes exactement identiques, par leurs caractéristiques, mais qui n'ont pas pu

bénéficier de la mesure et qui peut-être étaient dans votre groupe témoin ? Du coup, si l'effet est en partie dû à cet effet d'éviction, quand vous généralisez la mesure à tout le monde, ce « + » que vous aviez constaté va par définition disparaître. C'est un risque que l'on peut avoir pour certaines mesures, dont il faut bien avoir conscience : au fur et à mesure que l'échelle progresse, un rendement décroissant de votre mesure par rapport à des effets que vous aviez mesurés à échelle plus petite.

**Sébastien Malangeau** : Pour ce qui nous concerne, il est délicat d'apporter une réponse très précise. Selon la définition que le Conseil d'État avait donnée lors d'une étude sur les expérimentations mises en œuvre par les pouvoirs publics, l'expérimentation permettait de tester temporairement un dispositif et sur la base d'un cadre rigoureux, d'en faire une évaluation pour éclairer les pouvoirs publics. Dans le cas d'espèce, « éclairer les pouvoirs publics » est probablement le terme le plus important. Dans le cadre de France Expérimentation, c'est *in fine* le Premier ministre, par l'intermédiaire de son cabinet, qui prend la décision. Sa décision ne peut pas être contrainte. Donc tous les éléments de contexte peuvent participer de la décision : le contexte budgétaire, économique, les sensibilités politiques, voire corporatistes, ou les sensibilités locales... D'expérience, l'expérimentation porte finalement moins sur l'objet même (« Est-ce que c'est une bonne idée de permettre ce nouveau service, cette nouvelle activité ou ce nouveau produit en France ? »), que sur la modalité d'encadrement. Comme le *process* d'adaptation juridique est très long, si le cabinet du Premier ministre donne le feu vert pour lancer une adaptation juridique, c'est parce qu'on estime collectivement – même si parfois ce présupposé n'est pas totalement perçu par les administrations – qu'à terme cette faculté pourra être généralisée. La question devient plutôt : comment la généraliser de la manière la plus adaptée, c'est-à-dire la moins contraignante pour les acteurs, mais suffisamment efficace pour les pouvoirs publics et les superviseurs ? C'est un peu du *bêta-test* de la norme : on teste un encadrement et, sur la base des échanges de terrain, on voit si le curseur a été mis trop fortement ou insuffisamment du côté de la maîtrise des risques. Il s'agit d'objectiver que pendant 2-3 ans, il n'y a pas eu d'incidents ou que, sur la base des incidents relevés, s'ils sont mineurs, on a tiré les ajustements *ad hoc* pour avoir une version définitive qui corresponde pleinement aux attentes de l'ensemble des acteurs. Et d'éviter un des effets que l'on entend souvent, qui peut être réel ou partiellement fantasmé, à savoir que les pouvoirs publics ont encore produit une norme qui n'est pas adaptée au terrain. Il serait

complexe de donner une ligne de séparation. Une généralisation serait impossible, par exemple, si le cadre juridique temporaire s'était avéré totalement inopérant et si les pouvoirs publics avaient estimé qu'il n'y a pas d'alternative pour encadrer une activité. Ce qui, dit comme cela, paraît plutôt improbable. Il s'agit bien plus d'affiner le cadre définitif que de décider si on généralise ou si on met fin à une mesure. Cependant, sur le principe, rien n'est exclu. Avec France Expérimentation, on a eu une cinquantaine de projets en expérimentation. On n'a pas véritablement eu d'exemple d'expérimentation qui ait conduit à un arrêt du dispositif. Finalement, on a généralisé, mais pas toujours sur les conditions initiales.

## L'action publique a-t-elle raison, selon vous, de s'inscrire dans ce type de démarche ?

**Jérôme Gautié :** Oui, si elle ne s'inscrit pas dans une démarche qui reste trop technocratique. Je reprends le fil des expérimentations scientifiques. C'est vrai qu'il faut sortir d'une conception un peu trop simpliste où un État central prend des mesures qui vont s'imposer telles quelles aux acteurs qui les mettent en œuvre – Pôle emploi, les missions locales, par exemple, dans le domaine des politiques de l'emploi –, avec d'un autre côté l'expert (économique, surtout), qui fait ce type d'évaluation, qui vient mesurer ce qui se passe et qui dit « ça marche », « ça ne marche pas », ou « ça marche plus ou moins bien », etc. Et puis, on prend automatiquement la décision du « stop » ou « encore ». Je crois qu'au contraire, et dans le prolongement de ce qui a été dit, l'expérimentation scientifique entre dans un cadre plus large qui est de contribuer à établir un diagnostic partagé. C'est un terme qui était utilisé à l'époque du Commissariat général au plan, lorsqu'on mettait les acteurs autour de la table et que l'on essayait d'avoir un diagnostic partagé à travers un certain nombre de données, d'expertises, qui ne viennent pas seulement de la sphère académique, mais aussi des acteurs de terrain... Et la question était : « Que faut-il faire dans tel domaine ? » On est revenu à cette volonté d'établir un « diagnostic partagé » avec des institutions comme le Conseil d'orientation des retraites (COR), le Conseil d'orientation pour l'Emploi (COE), etc. Et, dans le cadre de l'évaluation plus spécifiquement, avec la mise en place des Conseils scientifiques que j'évoquais plus haut. Et je crois qu'effectivement, cette culture

de l'expérimentation, c'est une façon de mettre tous les acteurs autour de la table, et de le faire avec un outil légitime. On regarde ensemble ce qui se passe, on discute. On implique l'ensemble des acteurs dans toute la chaîne. Il ne s'agit pas d'être simplement « réactif », en recueillant le retour des acteurs une fois la mesure déployée. Il s'agit d'associer les acteurs dès le départ : mettons en place un protocole d'expérimentation, entendons-nous sur ce protocole et déjà, commençons à discuter à partir de ce protocole. C'est la grande vertu de ce processus d'expérimentation.

**Sébastien Malangeau :** Je partage totalement cet avis. Il est vrai que dans une période où l'on parle beaucoup de délégitimation de l'action publique, c'est un des *process* essentiels de co-construction, de concertation. Je suis très sensible à la notion de diagnostic partagé qui résume bien la situation. Déjà, partageons un constat, et construisons ensemble des pistes pour avancer sans préjuger de ce que sera le résultat dans sa totalité, en partant de l'idée que ces itérations vont vraiment, dans le cas de France Expérimentation, améliorer la norme. De manière plus générale, elles permettront de trouver les solutions les plus adaptées. C'est vraiment la manière vertueuse de mener une expérimentation. Cela correspond aux tendances actuelles et il est au fond assez logique qu'elles se développent. Ce qui serait peut-être un peu moins vertueux, c'est l'expérimentation qui s'apparenterait à de la temporisation. Si une décision est évidente, mais difficile à prendre, et si l'on se dit qu'on va faire une expérimentation pour reporter la décision... À une époque, on disait « Pour enterrer un problème, créons une commission ! » Si pour enterrer un problème, on fait une expérimentation, c'est de nature à casser l'objet et à nous renvoyer à ces problèmes de légitimité. Mais dès lors que le diagnostic est clair, que la perspective est globalement tracée et qu'ensuite, on se met collectivement d'accord sur le fait que l'expérimentation va permettre de créer la maille la plus fine et la plus satisfaisante pour tous, c'est un dispositif vertueux.

## L'expérimentation comme méthode présente-t-elle des limites ? Et comme outil d'aide à la décision ?

**Jérôme Gautié :** Si je m'en tiens à l'expérimentation scientifique, la mise en œuvre n'est pas si simple que cela. Pour l'expérimentation aléatoire, on se

dit : « C'est assez simple, on va tirer au sort et on va regarder ce qui se passe ». Il y a pourtant des obstacles de différentes natures. Cela vaut de façon plus générale pour l'expérimentation sociale (nous avons évoqué « Territoires zéro chômeur de longue durée »), mais peut-être plus encore dans un cadre d'expérimentation scientifique d'évaluation d'impact. Il y a des protocoles très contraignants pour les acteurs. D'abord, on a parfois du mal à trouver des agences locales de l'emploi qui veulent participer à l'expérimentation. Et quand elles participent à l'expérimentation, il y a des obstacles éthiques. Ce n'est pas évident, quand vous êtes un agent local de l'emploi, et que vous faites face à un jeune qui a besoin d'une mesure, de l'exclure du « traitement » parce qu'il est dans le groupe témoin. Dans certaines expérimentations (notamment aux États-Unis), on s'est aperçu que certains agents locaux de l'emploi, dans des agences qui participaient à l'expérimentation, « trichaient » un peu. Pour compenser le fait que le jeune n'avait pas été sélectionné pour faire partie du groupe de traitement, ils lui donnaient un peu plus d'appui par ailleurs ou par d'autres dispositifs, etc. Évidemment, cela faussait l'expérimentation. Il y a des obstacles très concrets comme ceux-ci. Dans un autre registre et de façon plus générale, quand on pense à l'usage de l'expérimentation scientifique, dans l'articulation entre le moment scientifique et le moment de la décision publique, il y a la question de la temporalité qui est très importante. Le temps du politique, parfois, est plus court que le temps du scientifique. Le scientifique aime avoir au moins 18 mois, voire deux ans. Ensuite, il lui faut le temps de traiter les données. Il aime avoir le recul temporel le plus important possible, donc interroger à un an, à deux ans, etc. Pour lui, ça serait l'idéal, mais le politique est parfois pressé. Un des exemples les plus célèbres, c'est le Revenu de solidarité active (RSA) : alors qu'on avait mis en place des protocoles d'expérimentation scientifique, à un moment donné, le politique dit : « Je généralise, même si les résultats officiels de l'évaluation scientifique ne sont pas encore là. » C'est un problème. Un autre point qui peut être évoqué concernant l'expérimentation scientifique, et qui peut frustrer le décideur public, c'est que si les expérimentations aléatoires, notamment quantitatives, sont très bonnes pour dire : « ça marche », « ça ne marche pas », « avec tel impact », etc., elles le sont moins pour expliquer *pourquoi* ça marche et *comment* ça marche. On a des chiffres, on a des statistiques, qui permettent de constater par exemple un différentiel positif de taux d'emploi entre les jeunes bénéficiaires et les jeunes non-bénéficiaires, donc ça marche. Mais qu'est-ce qui s'est passé sur le terrain pour

expliquer que ça marche en moyenne, et que ça marche plus ou moins bien selon les territoires où l'expérimentation a été déployée ? Ce type d'évaluation doit toujours être articulé – c'est ce qui est fait désormais assez systématiquement en France, au moins dans le domaine de la politique de l'emploi – avec des approches plus qualitatives qui ouvrent la « boîte noire » et vont voir sur le terrain comment les acteurs mettent en œuvre les mesures. Il y a vraiment une complémentarité entre le quantitatif et le qualitatif. Cela introduit une complexité supplémentaire, parce que l'on est généralement obligé, du fait de contraintes temporelles, de lancer les deux en même temps. Or, souvent, quand on a des retours du qualitatif, on aimerait modifier le protocole quantitatif de l'expérimentation, moduler ceci ou cela, mais il est trop tard pour le faire. On a ces problèmes de mise en œuvre, qui renvoient à des problèmes de temporalité, soit internes au protocole d'évaluation lui-même, soit dans l'articulation entre ce protocole et la décision publique.

**Sébastien Malangeau :** Je souscris totalement à cette idée. Je peux partager de manière beaucoup plus pratique les limites que nous avons identifiées sur France Expérimentation, où on met vraiment en œuvre, et cela du côté des pouvoirs publics. La première, c'est qu'il demeure toujours une différence importante entre le temps du droit et le temps de l'innovation, même si notre objectif est de compresser cet écart. Prenons la pire hypothèse, un blocage au niveau législatif : entre la collecte de la demande, le temps de l'instruction, l'arbitrage du Premier ministre, même si dans l'idéal on trouve un projet de loi très rapidement, un projet de loi entre son examen et la décision du Conseil constitutionnel, c'est souvent une année entre le passage des différentes assemblées. Donc même si on arrive à avoir une demande de dérogation législative opérationnelle deux ans plus tard – parce qu'il y a le texte de loi, mais aussi les textes réglementaires qui viennent ensuite –, c'est très, très court pour l'administration. Pour un acteur innovant, si vous êtes une start-up, deux ans c'est très long ! Donc nous pouvons parfois nous réjouir, mais nous sommes conscients que malgré tout, dans ce domaine-là, l'idée c'est de faire au mieux, sachant qu'il y aura toujours une forme de frustration pour beaucoup d'acteurs, surtout les plus petits et les plus jeunes. Un grand groupe international peut dérouler des projets à trois, quatre, cinq ans. Un acteur plus modeste a une temporalité plus courte. Deux autres limites qu'on découvre avec le temps, d'abord c'est la difficulté à massifier. France Expérimentation a la vertu de recevoir des demandes de tous types d'acteurs,

parfois des demandes individuelles ; les pouvoirs publics donnent suite à l'expérimentation, et après on n'atteint pas vraiment la masse critique. Donc *in fine*, quand on arrive en bout d'expérimentation, ça devient presque une auto-évaluation des pouvoirs publics de « Est-ce que le cadre qu'on a défini nous paraît satisfaisant ou non ? » Si ce cadre dérogatoire n'a pas été assez employé, la qualité objective de l'expérimentation sera délicate à établir. De manière malheureuse, on va passer de quantitatif à qualitatif sans que ça soit totalement satisfaisant. D'un autre côté, si on faisait un filtrage en demandant un nombre minimum d'acteurs, on se priverait dans un domaine comme l'innovation. Peut-être que la plus grande innovation de demain, paraît par essence totalement incongrue ou anecdotique à l'immense majorité des acteurs. C'est un équilibre assez subtil à trouver, et il faut parfois accepter que les expérimentations ne soient pas utilisées autant qu'on le souhaiterait. C'est aussi notre responsabilité collective de mieux communiquer. On peut sûrement mieux faire, mais il y a parfois une difficulté des acteurs à s'emparer des opportunités parce que l'expérimentation peut faire peur, parce qu'on est scruté par les pouvoirs publics, qu'il y a un protocole expérimental, donc il y a peut-être des obstacles aussi un peu psychologiques. Une difficulté pour nous – qu'on a su traiter, mais qu'on n'avait pas traitée au début –, dans ce temps long, c'est qu'on ne peut pas attendre le terme de l'expérimentation pour décider, parce qu'économiquement, quand vous êtes un acteur, que vous avez fait des investissements et que vous devez attendre la fin de la troisième année d'expérimentation pour que l'on vous dise « C'est une très bonne idée, mais maintenant arrêtez tout, on revient avec la loi modifiée dans deux ans, deux ans et demi », ce n'est pas satisfaisant. C'est quelque chose qu'on a appris : notre temps d'expérimentation doit intégrer un *temps masqué*. Dans le domaine réglementaire, par exemple, si la modification réglementaire prend un an, on dit au ministère : « Faites une expérimentation de quatre ans, mais la décision doit être prise à trois ans pour que l'acteur connaisse l'issue de l'expérimentation, et que si l'expérimentation est appelée à s'éteindre ou à donner lieu à des bases différentes, il puisse minimiser ses pertes et préparer sa sortie d'expérimentation. Et puis, si l'expérimentation doit être généralisée, s'assurer qu'en temps masqué, le cadre juridique sera adopté et qu'il n'y aura pas de rupture d'expérimentation. » Ça, c'est quelque chose qu'on n'avait pas intégré au début et qui n'a pas été sans nous poser des difficultés. Cela fait partie des choses qu'on apprend. Nous sommes toujours sur des temporalités assez longues.

**Jérôme Gautié** : Sébastien évoquait le fait qu'il y a parfois des acteurs frileux à expérimenter : on lance une expérimentation, mais il n'y a pas forcément les troupes derrière qui sont prêtes à expérimenter. À l'opposé, il peut y avoir un enthousiasme des acteurs de terrain qui disent : « Pourquoi même expérimenter, en fait ? » Ils prennent cette voie-là parce que c'est la voie qu'on les autorise à prendre, mais ils sont tellement persuadés que c'est la bonne mesure à prendre qu'il peut même y avoir de la frustration à se dire : « Mais pourquoi est-on encore dans une phase d'expérimentation ? Mais pourquoi nous impose-t-on des protocoles d'évaluation aussi contraignants qui ne sont que des obstacles ? » Avec « Territoires zéro chômeur », il y a des tensions entre les acteurs de terrain et les évaluateurs, qui peuvent être de différents niveaux. Il y avait notamment des experts économiques qui disaient : « Ouh là là, attention, on n'a pas de preuve vraiment très objective que ça marche au sens où le diraient les économistes, c'est-à-dire avec un groupe témoin, avec ce qu'on appelle un contrefactuel, etc. », alors que les acteurs de terrain rétorquaient : « Mais si, on le voit sur le terrain, on le sent, etc. » Ce décalage, qui peut vraiment créer des tensions importantes, peut naître aussi du fait que les différents acteurs n'ont pas forcément la même vision de l'expérimentation, notamment parce qu'ils ne privilégient pas forcément les mêmes objectifs. Je reprends le cas de la Garantie jeunes : il n'y a pas eu trop de tensions parce que ça marchait en termes de retour à l'emploi, et que tout le monde était content que ça marche. Mais les associations de jeunes insistaient sur le fait que la variable emploi n'était pas la seule importante (ni peut-être même, la plus importante), il y a le fait qu'on a re-socialisé des jeunes, le fait qu'on leur a donné des ressources pendant une certaine durée... Et que tout cela doit aussi compter dans « l'évaluation ». Et donc, que doit-on évaluer exactement ? Sur quelle(s) variable(s) de résultats doit-on porter l'attention ? Un des problèmes de l'évaluation scientifique, et notamment quand elle est quantitative, c'est ce qu'on appelle l'« effet lampadaire », c'est-à-dire que l'on va se concentrer uniquement sur ce que l'on peut mesurer, ce sur quoi on va pouvoir rassembler des données. Un taux d'emploi, c'est assez facilement mesurable, mais d'autres variables sont beaucoup plus difficiles à mesurer. Il y a un concept qui est assez important en politique de l'emploi, c'est l'autonomie. De nos jours, on évoque beaucoup la nécessité de rendre les chômeurs plus autonomes, avec toute cette problématique d'*empowerment*, pour utiliser un terme anglo-saxon... L'autonomie, l'*empowerment*, la capacitation, comment peut-on les mesurer ?

Les psychologues ont des tests psychométriques, mais qui sont encore assez peu utilisés, notamment en France. Il est plus facile de mesurer les taux d'emploi, donc les économistes vont se concentrer là-dessus. Les effets en termes d'autonomie, en termes de bien-être, en termes d'acquisition de compétences socio-comportementales, sont beaucoup plus délicats à mesurer... On a désormais construit des indicateurs et des protocoles

d'enquête, mais pendant très longtemps on était assez désarmé là-dessus. Donc tous ces aspects positifs et qualitatifs d'une mesure passaient sous les radars. Et ce sont justement ceux auxquels les acteurs de terrain sont particulièrement sensibles, parce qu'ils sont les mieux placés pour les observer, alors que le statisticien-économiste, ne les perçoit pas forcément, et du coup, ne va pas les juger pertinents. De là peuvent naître des tensions.

# Études



# Comprendre les comportements pour élaborer des politiques publiques efficaces

Nicolas Jacquemet

**L'efficacité des interventions de politique publique dans le domaine économique et social dépend fortement de la manière dont les agents qui y seront exposés réagiront à leur mise en place. Sous l'impulsion de travaux inter-disciplinaires qui font une large place aux sciences du comportement, la recherche en économie a connu une inflexion comportementale qui a considérablement enrichi la représentation classique des déterminants du comportement sur laquelle s'appuient ses recommandations de politique publique. Cet article propose un panorama des implications de ce renouvellement profond de la discipline sur l'élaboration des politiques publiques, appliqué notamment au cas de la fraude fiscale. Au-delà de sa manifestation la plus connue, les *nudges*, cette approche comportementale en économie permet à la fois de mieux comprendre les effets attendus d'outils classiques d'intervention publique, et de faire émerger de nouveaux leviers d'intervention.**

*« Il y a en France treize cent vingt mille maisons de paysans qui n'ont que trois ouvertures, dix-huit cent dix-sept mille qui ont deux ouvertures, la porte et une fenêtre, et enfin trois cent quarante mille cabanes qui n'ont qu'une ouverture, la porte. Et cela, à cause d'une chose qu'on appelle l'impôt des portes et fenêtres. Mettez-moi de pauvres familles, des vieilles femmes, des petits enfants, dans ces logis-là, et voyez les fièvres et les maladies ! Hélas ! Dieu donne l'air aux hommes, la loi le leur vend ! »*

Victor Hugo, *Les Misérables*.

Les décideurs publics ont depuis longtemps appris à leurs dépens que des politiques publiques efficaces ne sauraient se concevoir à comportement constant. Les acteurs de l'économie s'adaptent aux conséquences de leurs décisions et réagissent à l'environnement institutionnel dans lequel ils évoluent. Une intervention de politique publique revient à revisiter ces conséquences (en augmenter le coût par l'impôt, en diminuer le coût par la subvention, en limiter l'accès par des conditions légales d'éligibilité, etc.), et conduit donc inévitablement à des changements de comportement. Que ces réactions limitent l'efficacité de la politique mise en place (comme dans le fameux exemple de l'impôt sur les fenêtres, qui désespère jusqu'à Victor Hugo) ou qu'elles coïncident au contraire avec les objectifs visés, la réflexion

en matière d'élaboration des politiques publiques doit s'appuyer sur une compréhension fine des déterminants des décisions des agents de l'économie, afin de pouvoir anticiper leur réponse aux nouvelles conditions qui seront créées par le déploiement des mesures envisagées.

Pour nourrir ces réflexions sur la manière dont les acteurs de l'économie réagissent aux conséquences de leurs décisions, et donc à des changements dans ces conséquences, l'économie s'est dotée d'une représentation des décisions qui lui vaut, aujourd'hui encore, une très mauvaise réputation : le fameux *homo economicus* qui fonde la réflexion en économie est parfaitement informé des conséquences de ses décisions, parfaitement conscient de ce qu'il souhaite et ne souhaite pas, et parfaitement à même de synthétiser l'ensemble

de ces informations afin de prendre la décision la meilleure pour lui-même ; en bref : il est parfaitement rationnel. Aussi caricaturale que soit cette représentation, elle est suffisante pour prédire qu'un contribuable qui contrôle l'assiette sur laquelle porte un impôt aura tendance à réduire cette assiette pour en réduire la charge... Elle ne permet pas, en revanche, de comprendre pour quelle raison étrange des contribuables qui reçoivent une simple lettre leur rappelant l'importance des impôts pour le financement des biens publics, ou leur indiquant que la très grande majorité des contribuables déclarent fidèlement leurs revenus, puisse suffire à faire doubler le montant d'impôt collecté (Bott et al., 2019)...

À l'instar de cet exemple, de nombreux ressorts des décisions économiques échappent à la représentation traditionnelle (et délibérément simplificatrice) des déterminants du comportement dont s'est dotée la science économique, et ont des conséquences majeures sur le fonctionnement de l'économie. Si les comportements ont toujours été au cœur de sa réflexion, la recherche en économie a connu au cours des dernières décennies une révolution comportementale qui s'est nourrie de collaborations avec la psychologie et les sciences cognitives. Cette évolution doit beaucoup au développement de l'utilisation d'expériences contrôlées, qui ont le double avantage de permettre d'observer à petite échelle les effets attendus de différentes mesures de politique publique, et de permettre une compréhension fine des déterminants des comportements économiques. Au-delà de sa manifestation la plus connue (et la plus controversée) dans le domaine des politiques publiques, les *nudges*, cette inflexion comportementale infuse aujourd'hui la réflexion académique dans tous les domaines d'application de la science économique. Ces développements sont désormais suffisamment mûrs pour permettre d'éclairer la décision publique (et de réhabiliter l'économie aux yeux du grand public !), comme nous essaierons de le montrer à partir de l'exemple de la fraude fiscale.

## L'économie expérimentale, un outil précieux d'aide à la décision publique

Ce tournant comportemental a été rendu possible par l'émergence d'une nouvelle méthode, l'économie expérimentale (Jacquemet, Le Lec et L'Haridon, 2019 ; Eber et Willinger, 2012), qui permet de confronter la manière dont les décisions sont effectivement prises à la compréhension qui

en est donnée par la théorie économique. Cette méthode consiste très concrètement à créer des jeux dont les règles sont aussi proches que possible du contexte de décision étudié par la théorie économique, puis d'observer les décisions qui sont prises par des participants dont la rémunération dépend directement de leurs décisions et de celles des autres. Pour obtenir des mesures aussi précises que possible, ces expériences sont souvent réalisées dans ce que l'on appelle un « laboratoire », qui n'est rien d'autre qu'une salle informatique où les participants découvrent les règles du jeu, prennent des décisions et obtiennent un paiement qui dépend de leurs décisions en application de règles préalablement énoncées.

L'économie expérimentale constitue, en tant que telle, un outil puissant d'aide à l'élaboration des politiques publiques. La méthode expérimentale permet en effet d'observer les comportements qui résultent de différentes règles de fonctionnement de l'économie dans une version « miniature » que l'on a créée en laboratoire selon la situation économique à laquelle on s'intéresse. Ces règles sont autant de modalités de régulation différentes de l'économie qui ont pour objectif de modifier les décisions prises et d'atteindre un équilibre différent : l'assurance chômage, par exemple, vise à accompagner la recherche d'emploi et à assurer le risque de perte d'emploi. Pour penser l'assurance chômage, il faut comprendre les comportements de recherche d'emploi et se demander quel est le système qui favorise un retour à l'emploi le plus rapide possible, et vers l'emploi de la meilleure qualité possible. La méthode expérimentale peut donc être utilisée comme banc d'essai pour tester différentes décisions de politiques publiques en observant les conséquences de ces règles sur les décisions de recherche d'emploi dans l'expérience. C'est, dans un tout autre domaine, la démarche qui a été adoptée aux États-Unis pour choisir la meilleure méthode d'attribution des licences de téléphonie mobile de 3<sup>e</sup> ou 4<sup>e</sup> génération : les autorités ont déterminé les revenus générés au bénéfice du gouvernement en fonction de différentes règles d'enchères testées en laboratoire.

En permettant une compréhension plus fine des déterminants des comportements économiques, l'économie expérimentale aide également à réfléchir différemment aux conséquences de différents types de politiques publiques, et à développer de nouveaux outils d'intervention. Par exemple, l'épargne consiste à reporter une partie de son revenu vers le futur, diminuant ainsi le pouvoir d'achat immédiat au profit de la consommation à venir. Dans l'analyse de cet arbitrage, l'approche économique traditionnelle suppose que

les épargnants prennent des décisions cohérentes dans le temps : la manière dont un *homo economicus* compare ses possibilités de consommation entre aujourd'hui et demain devrait être identique à la comparaison qu'il opère entre demain et après-demain. Dans le cadre d'une expérience, cette hypothèse peut-être facilement testée en proposant aux participants de choisir différents gains monétaires dont le versement est décalé dans le temps. Les résultats généralement observés sont sans appel : l'hypothèse de cohérence temporelle décrit mal les décisions d'allocation observées sur le terrain. *Homo economicus*, le pendant charnel d'*homo economicus* que l'économie comportementale s'efforce de décrire, semble avoir une « préférence pour le présent » : beaucoup de participants préfèrent recevoir 100 euros le jour même plutôt que 101 euros le lendemain, alors même qu'ils sont disposés (le jour J) à faire le choix inverse s'il concerne deux versements futurs (entre J+1 et J+2). Pour revenir à l'assurance-chômage, des recherches récentes montrent les limites des dispositifs d'allocations dégressives ou des sanctions fondées sur des objectifs de retour à l'emploi (Paserman, 2008) : la perspective de la diminution future des allocations n'a en effet que peu d'effet sur le retour à l'emploi car, compte tenu de la préférence des individus pour le présent, elle est trop éloignée dans le temps pour être prise en compte dans leurs comportements de recherche d'emploi. Mieux vaut mettre en œuvre des objectifs de court terme, et multiplier les paliers d'indemnisation qui ont un effet de rebond important sur les comportements de recherche d'emploi.

## Les prémices des interventions comportementales : les *nudges*...

L'axiome de la rationalité parfaite implique que chacun est en pleine capacité de prendre les décisions les meilleures pour lui-même et constitue par là l'un des arguments les plus forts en faveur des politiques libérales : chaque membre de la société doit disposer de la plus grande liberté décisionnelle possible dans la mesure où il est le mieux à même de savoir ce qui est bénéfique pour lui-même.

Or, l'économie expérimentale a permis de faire émerger un ensemble convergent de faits stylisés qui distinguent les décisions effectivement prises des décisions attendues de l'*homo economicus* rationnel : outre la préférence pour le présent, les décisions ont notamment tendance à être prises en comparaison d'un point de référence, et les pertes sont ressenties plus durement que les gains ne sont sources de satisfaction (ce qui conduit par exemple à accorder plus de valeur à un bien que l'on possède qu'au même bien que l'on chercherait à acquérir) ; ces décisions sont influencées, par conformisme, par celles que prend autrui dans des circonstances similaires ; elles sont sensibles à l'image sociale qui leur est attachée (qui conduit à se comporter différemment lorsque l'image sociale est activée par une prise de décision en public), mais aussi à l'image de soi qui en découle (qui conduit à prendre des décisions différentes lorsque l'on est confronté aux décisions qui ont été prises par les autres), etc.

Nombre de ces exemples conduisent à des décisions qui peuvent être contraires aux intérêts de l'agent économique lui-même. La préférence pour le présent, par exemple, conduit à prendre des décisions de consommation à l'instant présent qui seront regrettées dans le futur. Ce constat a des conséquences importantes sur la manière de penser les politiques publiques : il pourrait notamment conduire à favoriser des politiques purement paternalistes (i.e., qui cherchent à orienter les décisions dans une direction considérée *a priori* comme souhaitable par la puissance publique). La puissance publique s'efforcera alors de contraindre les décisions (par exemple, en rendant l'épargne obligatoire) puisque la liberté décisionnelle des agents économiques conduit à des situations regrettables. Mais ce serait substituer une nouvelle imperfection – celle de la puissance publique – à celle que l'on voulait résoudre – celle de la rationalité. La réussite des politiques paternalistes est en effet elle-même déterminée par la capacité de l'autorité en charge de leur mise en œuvre à recueillir toute l'information nécessaire sur les situations individuelles pour décider des orientations souhaitables.

L'approche connue sous le nom de *nudge* (« coup de pouce » en français)<sup>1</sup> propose une voie originale, entre liberté des acteurs et interventionnisme : le paternalisme libéral. Plutôt que d'imposer et de contraindre, cette démarche propose de

1 Et fortement associée aux noms de Thaler et Sunstein (2010). Voir, notamment, Ferey, Gabuthy et Jacquemet (2013) pour une discussion détaillée.

tenir compte des fondements psychologiques du comportement qui rendent caduque l'hypothèse selon laquelle la liberté décisionnelle conduit à l'optimalité des décisions individuelles. Cette démarche a par exemple été appliquée avec succès à l'épargne : des travaux ont montré que la règle par défaut d'adhésion des employés américains au programme d'épargne retraite de leur entreprise constituait un outil efficace pour contrecarrer les effets de la préférence pour le présent (Choi, Laibson, Madrian, & Metrick, 2004). Le simple passage d'une exclusion par défaut à une adhésion par défaut lors du recrutement conduit à une augmentation considérable des montants épargnés par les salariés. Cette intervention est à la fois paternaliste, puisqu'elle infléchit les décisions individuelles dans une direction considérée comme souhaitable, et libérale, puisque la décision d'épargne relève de la volonté individuelle.

Cette approche ouvre un vaste champ de modalités nouvelles d'interventions publiques : des interventions subtiles, souvent peu coûteuses car agissant sur les ressorts psychologiques de la prise de décision, mais aux conséquences potentiellement importantes, qui vont d'une disposition habile des plats dans le linéaire d'une cafétéria afin de favoriser des compositions de plateau plus conformes aux recommandations diététiques, à la réduction du nombre de choix disponibles pour faciliter la prise de décision et limiter les conséquences de la complexité, en passant par la mise en place de rappels réguliers d'obligations à venir afin d'en augmenter la saillance et de s'assurer qu'elles seront honorées. Si l'on peut regretter que la force de l'idée et le talent de ses initiateurs aient conduit à élargir la terminologie jusqu'à lui faire perdre l'essentiel de sa signification (la notion courante de *nudge* faisant désormais référence à toute intervention qui s'appuie sur les déterminants psychologiques des comportements), l'intérêt d'une approche comportementale des interventions publiques a été bien compris dans nombre de pays occidentaux. L'administration de David Cameron s'est ainsi dotée d'une *behavioral team*, et la présidence Obama s'est appuyée sur une *nudge brigade* pour élaborer de nouvelles interventions en matière de tabagisme, de lutte contre l'obésité ou encore d'efficacité énergétique.

En France, cette approche n'a été exploitée que de manière éparse et localisée<sup>2</sup>. Le ministère de la Transition écologique et solidaire y a eu recours

lorsqu'il a décidé d'activer par défaut l'option d'impression recto verso au sein de ses services, conformément à la circulaire du 3 décembre 2008 relative à l'exemplarité de l'État au regard du développement durable. On reconnaît également cette approche dans le changement récent de réglementation s'appliquant au don d'organe, par lequel toute personne devient donneuse par défaut – sauf à exprimer son désaccord – plutôt que non-donneuse – sauf à exprimer son accord.

## ...et au-delà : l'approche comportementale des politiques publiques

Cette réticence à mettre en œuvre les recommandations issues des sciences comportementales doit beaucoup au fait que les *nudges* sont fréquemment perçus (en partie à raison) par les décideurs publics comme des outils d'intervention qui s'apparentent aux techniques du marketing, qui se déclinent en un nombre d'outils limités, permettant de résoudre des problèmes locaux, et produisant des changements de comportement certes peu coûteux, mais souvent limités à des effets de court terme et d'ampleur modeste. S'ils restent la forme la plus connue d'intervention publique fondée sur les sciences comportementales, et que l'on doit à cette approche d'avoir imposé les sciences comportementales dans le débat public, le succès des *nudges* a cependant eu pour effet pervers de finir par cacher la forêt constituée de l'ensemble du spectre des interventions de politique publique efficaces, parce que fondées sur les enseignements des sciences comportementales.

Contrairement au terrain naturel d'application des *nudges*, de très nombreux aspects de la prise de décision économique mis en évidence par les sciences comportementales ne conduisent pas à des décisions contraires à l'intérêt de l'agent économique qui les prend – ouvrant la voie à la possibilité d'une intervention fondée sur les préceptes du paternalisme libéral – mais correspond au contraire à une décision éclairée qui a pour seul inconvénient de ne pas coïncider avec le modèle de décision formulé par l'économie classique (au point qu'il est devenu courant d'accoler à ces ressorts psychologiques de la décision le

<sup>2</sup> Notons cependant que les actions de la Direction interministérielle de la Transformation publique sont de plus en plus fortement ancrées dans les sciences comportementales.

qualificatif de « biais »), et oblige à en reconsidérer les recommandations<sup>3</sup>.

La théorie économique constitue un cadre de réflexion approprié pour comprendre l'effet sur les comportements des incitations financières, que celles-ci visent à dissuader certaines décisions (par l'intermédiaire de taxes ou d'amendes) ou au contraire à les encourager (par des subventions ou des primes). Ces incitations se traduisent implicitement par des variations de prix, qui tendent naturellement à dissuader les décisions dont les prix s'accroissent au bénéfice de celles qui deviennent financièrement plus rentables. La politique fiscale constitue en ce sens un outil de pilotage particulièrement efficace permettant d'aligner les comportements économiques avec les objectifs poursuivis par la puissance publique via une modulation des taux de taxe en fonction des activités économiques que l'on souhaite encourager ou dissuader. Aussi puissant que soit ce raisonnement, et aussi fiables que soient les analyses de politiques publiques qui en découlent dans bien des domaines, la réponse comportementale aux incitations se révèle plus complexe et plus subtile qu'il y paraît. Les incitations financières peuvent être le vecteur de déterminants non-monétaires du comportement, susceptibles d'en atténuer l'efficacité, voire de les rendre contre-productives.

Une première raison est liée à l'information véhiculée par les incitations. Imaginons un employeur qui confie une tâche à un employé, et que la difficulté de la tâche ne soit connue que de l'employeur. Afin de mettre en place des incitations suffisantes, l'employeur aura tendance dans cette situation à proposer une rémunération d'autant plus élevée que la tâche est difficile. Loin de rendre la tâche attractive, ce surcroît de rémunération est cependant de nature à signaler à l'employé la très grande difficulté de la tâche. Face à cette difficulté, l'employé peut avoir intérêt à diminuer son effort. Au lieu d'encourager l'employé à consentir les efforts nécessaires pour s'acquitter de la tâche, des incitations financières particulièrement fortes peuvent paradoxalement avoir un effet décourageant. Dans le domaine des politiques publiques, l'étude de Frey et Oberholzer-Gee (1997) a montré que le simple fait de proposer un dédommagement financier aux habitants d'une zone géographique où l'implantation d'un site nucléaire était envisagée conduisait paradoxalement à augmenter les réticences de ces derniers, car le dédommagement était interprété comme le signal d'un risque d'incident élevé.

Un phénomène similaire a pu être mis en évidence s'agissant des décisions qui comportent une dimension morale. Celles-ci sont fortement influencées par l'idée qu'on se fait de soi-même, et par l'idée que les autres se font de nous, sur la base de nos actes. Or, l'introduction de conséquences financières pour ces mêmes actes ne se cumule pas avec ces jugements moraux : lorsqu'une action procure un bénéfice financier, elle n'est associée à aucune récompense sociale, ni à ses propres yeux, ni aux yeux des autres. Faisant disparaître les conflits moraux associés aux décisions ciblées, les incitations monétaires peuvent donc paradoxalement décourager les décisions qui sont récompensées, car elles sont perçues comme étant guidées par le gain matériel qu'elles procurent plutôt que comme un signe d'élévation morale.

Cette idée a fait son apparition dans la pensée économique à l'occasion d'un intense débat entre le père fondateur des recherches en politiques sociales, Richard Titmus, et les plus fameux économistes des années 1970, convaincus que les forces de marché s'appliquaient à toute forme de bien et que l'introduction de récompenses financières pour les dons de sang ne pouvait qu'en accroître l'offre. Titmus était au contraire convaincu qu'une telle mesure serait contre-productive et conduirait à une baisse des dons, car l'introduction de récompenses financières annihilerait l'une des motivations importantes de la décision de se porter volontaire : le sentiment d'avoir fait une bonne action, et la récompense que constituent l'estime de soi et l'estime des autres qui en découlent. Les travaux de recherche qui ont été menés depuis confirment sans ambiguïté cette intuition : l'introduction de récompenses financières n'augmente aucunement le volume des dons de sang, et pose des questions qui restent largement débattues sur la qualité des dons attirés par de telles récompenses (Mellström et Johannesson, 2008).

Un exemple célèbre en est fourni par une expérimentation réalisée en Israël, et qui visait à limiter les retards très fréquents des parents qui viennent chercher leurs enfants à la sortie de la crèche. Dans la mesure où ce type de comportement n'est pas sanctionné (si ce n'est par l'attitude peu amène du personnel qui se trouve obligé de rester au travail au-delà de l'horaire de fermeture), un raisonnement utilitariste conduit à penser que l'instauration d'une amende monétaire devrait créer un coût conduisant les parents à être plus vigilants sur leur heure d'arrivée. Non seulement

<sup>3</sup> Voir Gabuthy, Jacquemet et L'Haridon (2021) pour de plus amples illustrations dans le domaine des politiques de santé, des politiques environnementales ou encore des politiques touchant le marché du travail.

la mise en œuvre d'une telle amende n'a pas produit les effets attendus, mais elle a même conduit à un doublement du nombre de retards – en comparaison des crèches dans lesquelles cette amende n'avait pas été mise en place (Gneezy et Rustichini, 2000), car l'amende était interprétée par les parents comme une garde additionnelle payante, la tarification d'un service nouveau qui consistait à choisir l'heure à laquelle il était possible de venir chercher ses enfants. L'introduction d'une amende poussait ainsi les parents à raisonner en termes d'arbitrage entre le coût que représente l'amende et le bénéfice d'être en retard, plutôt que de s'efforcer d'être à l'heure pour se conformer à un principe valorisé à la fois socialement et individuellement : le respect du personnel.

## Le mystère de la fraude fiscale<sup>4</sup>

La fraude fiscale constitue une limite évidente au recours à la fiscalité comme outil de pilotage des décisions économiques. C'est d'ailleurs la principale difficulté que la fraude fiscale pose à la politique économique : aussi élevé que soit son montant (estimé en France à plus de 80 milliards d'euros, soit le budget annuel du ministère de l'Éducation nationale<sup>5</sup>), la fraude fiscale resterait un problème assez superficiel si elle se traduisait uniquement par une diminution des recettes de l'État. Il suffirait en effet de tenir compte de la fraude dans la détermination des taux de taxes pour suppléer à ses conséquences purement budgétaires. Outre qu'une telle hausse des taux de taxe peut affecter les comportements économiques (car la taxation tend à affaiblir les incitations liées à la rémunération des activités productives), la fraude fiscale suscite cependant de réelles difficultés dès lors qu'elle est pratiquée de manière inégale par différentes catégories de population. Lorsque son intensité varie au sein de la population, elle se traduit notamment par une subvention des contribuables honnêtes au profit de ceux qui se soustraient à leurs obligations fiscales. Lorsqu'elle est inégalement répartie en fonction des secteurs d'activité ou des caractéristiques individuelles des contribuables, elle contrarie en outre les objectifs qui président à l'élaboration du système de taxation. La fraude fiscale déséquilibre enfin les objectifs redistributifs de la politique économique si elle est inégalement pratiquée entre les contribuables en fonction de leur niveau de revenu, et la réallocation

des activités de production si un secteur est fortement taxé afin de contenir son expansion, mais se prête plus facilement que d'autres à la fraude.

La fraude fiscale est longtemps restée un sujet difficile à aborder pour la théorie économique, car l'application simple du calcul coût-bénéfice auquel est supposé se livrer le contribuable « rationnel » conduit à un paradoxe : contrairement à une idée largement répandue, les bénéfices de la fraude fiscale sont tellement élevés, et le risque de sanction est tellement faible, que la fraude fiscale est étonnamment limitée dans l'ensemble des économies développées et devrait atteindre des montants encore plus élevés que ceux qui sont constatés. Plutôt que la fraude qui en constitue le pendant, c'est donc la disposition à payer l'impôt qu'il convient d'expliquer pour comprendre les déterminants des décisions des contribuables.

Afin de comprendre les raisons de ce « mystère de la fraude fiscale » (ou plutôt, le mystère que constitue son absence), une première génération de travaux d'économie comportementale s'est tournée vers la psychologie de la moralité et a introduit l'idée qu'il existait une « morale fiscale », une disposition intrinsèque à s'acquitter de ses obligations fiscales. Il s'agirait alors de comprendre quelles sont les racines d'une telle morale fiscale, autrement dit les raisons pour lesquelles certaines personnes décident de se comporter de manière honnête tandis que d'autres s'engouffrent dans la possibilité de dissimuler leurs revenus. Pour répondre à cette question, des travaux récents ont combiné l'application de la méthode expérimentale à l'observation des comportements de fraude fiscale, avec une batterie de questionnaires de personnalité empruntés à la psychologie et visant à documenter les composantes de la morale fiscale qui distinguent les contribuables les uns des autres. Ces questionnaires visent à mesurer au niveau individuel les traits de personnalité qui sont liés à l'adoption de comportements moraux selon la psychologie de la moralité : la propension à se soumettre à une norme, l'empathie et enfin la propension à ressentir honte et culpabilité. Les résultats montrent que même s'il existe des relations entre certains aspects de la personnalité et les comportements de fraude, aucune de ces relations ne permet de dresser un profil type du fraudeur fiscal. Lorsque l'ensemble des caractéristiques individuelles sont prises en compte, elles n'expliquent jamais plus de 20 % des décisions de fraude fiscale observées.

4 Cette sous-partie est fortement inspirée de Jacquemet, Luchini et Malézieux (2021).

5 Cette évaluation émane du syndicat Solidaires-Finances publiques (2018).

La conclusion générale qui se dégage de ces résultats est que, pour nombre de décisions de la vie quotidienne, la moralité intrinsèque due à la personnalité n'a que peu à voir avec l'adoption de comportements moraux (Terestchenko, 2005). Appliqués à la fraude fiscale, ces résultats impliquent que faire appel à la moralité des contribuables, leur rappeler les conséquences sociétales désastreuses de la fraude fiscale, et invoquer leur sens des responsabilités peut n'être que de peu d'effet sur une décision qui, pour une large frange de la population, est principalement motivée par des raisons externes. Une explication souvent avancée d'une telle absence de déterminisme des caractéristiques individuelles sur les comportements est que la personnalité et l'identité individuelle affecteraient principalement l'intention d'adopter un comportement particulier. Mais cette intention ne se transforme pas nécessairement en action : l'un des médiateurs cruciaux de cette transformation des intentions en action est le contexte dans lequel les décisions sont prises.

## La banalité du bien : l'influence du contexte sur la « morale » fiscale

Un vaste ensemble de recherches en psychologie sociale montre en effet qu'une même personne peut, suivant le contexte, être amenée à prendre des décisions qui servent ou desservent le bien commun. Une expérience célèbre s'intéresse par exemple au comportement de l'utilisateur d'une cabine téléphonique, confronté lorsqu'il en sort à un passant (complice de l'expérience) qui trébuche et répand devant lui les documents qu'il portait. L'attitude moralement appropriée est bien évidemment d'aider cette personne à rassembler les documents qu'elle a éparpillés. Selon une approche purement comportementaliste de cette situation, la question qui se poserait est de savoir de quelles informations il faudrait disposer pour pouvoir prédire l'attitude qu'adoptera la personne qui est confrontée à cette situation en sortant de la cabine téléphonique : est-il suffisant de connaître son âge, son sexe, sa profession, de disposer de mesures objectives de sa personnalité, etc., pour se faire une idée précise de sa « moralité » et donc de sa disposition à aider un inconnu dans le besoin ?

La réponse sans ambiguïté des nombreuses expériences de psychologie sociale qui utilisent ce scénario est qu'une même personne adoptera des attitudes totalement différentes suivant le

contexte. L'une des variations de cette expérience consistait par exemple à déposer une pièce de monnaie dans la cabine avant que la personne dont on étudie le comportement y entre. Le simple fait d'être exposé à cette bonne surprise fait passer la proportion de gens qui décident d'aider le passant malchanceux de 4 % à 87,5 %. L'attitude positive qui préside à des actes généreux est ainsi sensible à des détails aussi futiles qu'un simple coup de chance.

Afin d'illustrer les différents types de ressorts psychologiques permettant d'influencer les comportements, Robert-Vincent Joule et Jean-Léon Beauvois (Joule et Beauvois, 1998) présentent le cas imaginaire d'une institutrice qui est contrainte de quitter sa salle de classe avant ses élèves, le soir à l'heure de la sortie. Soucieuse de la consommation d'énergie, elle demande à ses élèves de bien vouloir penser à éteindre la lumière en quittant la salle de classe, après leur avoir longuement expliqué les enjeux économiques autant qu'écologiques d'un geste aussi simple. À son arrivée le lendemain matin, elle constate que la lumière est restée allumée toute la nuit. Par souci pédagogique, elle décide de quitter à nouveau la classe en avance (délibérément cette fois), et met en œuvre une stratégie de motivation qu'elle espère plus efficace en insistant cette fois sur le sens des responsabilités des élèves et demande à chacun de faire appel à sa conscience en quittant la classe. La même déception l'attend cependant le lendemain matin. Elle persiste à obtenir de ses élèves le comportement qu'elle juge socialement approprié, renouvelle sa déclaration, puis ajoute un autre élément de contexte en se tournant cette fois vers la classe et en demandant quel élève est prêt à prendre la responsabilité d'éteindre la lumière. Aucun volontaire ne se manifeste et elle désigne alors un élève choisi au hasard, lui demande s'il veut bien s'en charger, s'assure de son plein accord une seconde fois suite à sa réponse affirmative, et quitte la classe. Les chances sont élevées que la lumière serait cette fois-ci bel et bien éteinte au retour de l'institutrice.

Le contraste entre les trois tentatives de l'institutrice pour influencer ce comportement auquel les élèves attachent peu d'importance (comme le montrent leurs réticences à se porter volontaires) est pourtant révélateur des différents leviers psychologiques sur lesquels il est possible d'agir. La première s'apparente à une stratégie de persuasion : elle participe de l'idée que les gens agissent uniquement en fonction de leur personnalité et de leurs convictions, de sorte qu'il suffit de les convaincre du bien-fondé d'une action pour les voir l'adopter. La seconde prend acte de ce que

les convictions ne trouvent pas toujours de traduction dans les actes en l'absence de motivation intrinsèque à le faire. Pour stimuler la motivation intrinsèque, elle complète la stratégie de persuasion d'un appel au sens des responsabilités, qui vise à mettre l'accent sur l'importance qu'il convient d'accorder aux conséquences éthiques de ses actes. La troisième et dernière tentative relève d'une approche radicalement différente des déterminants du comportement et fait appel à une motivation purement externe : la décision délibérée de la part de l'élève d'accepter une demande (savamment formulée de telle sorte qu'il soit très difficile de la refuser). Plutôt que les qualités intrinsèques de l'élève (qui avait été choisi au hasard), sa sensibilité à la cause environnementale ou encore sa volonté de préserver sa relation avec son institutrice, c'est la décision d'accepter la demande explicite de l'institutrice qui prédispose cet élève à adopter un comportement que la persuasion s'est avérée incapable d'obtenir.

## La fraude fiscale sous serment : quels contribuables tiennent leurs engagements ?

La théorie de l'engagement en psychologie sociale formalise ce levier d'action sur les comportements et montre que la décision est un élément central de contexte : le facteur déterminant pour expliquer l'adoption d'un comportement n'est pas l'ensemble de raisons (bonnes ou mauvaises) qui ont conduit à la prendre, mais l'acte même d'avoir pris une décision dans le passé qui est cohérente avec ce comportement. Tout se passe comme si le fait même d'avoir décidé produisait « un effet de gel » : les actes passés produisent un engagement, un « lien qui unit l'individu à ses actes », qui se traduit par des comportements ultérieurs cohérents avec cette première action.

Appliquée aux comportements de fraude fiscale, la théorie de l'engagement élargit considérablement la palette d'outils de politique fiscale disponibles : des modifications de l'architecture de la phase de déclaration de revenus sont ainsi susceptibles de produire un engagement en direction de l'honnêteté des contribuables. L'une des techniques d'engagement à la fois les plus simples, les plus efficaces, et les plus transposables sur le terrain consiste à faire précéder l'étape de déclaration du revenu par la signature d'un code d'honneur

ou d'une simple promesse à se comporter de manière honnête. Une étude récente (Koessler et al., 2019), réalisée en coopération avec les services fiscaux suisses, permet de mesurer l'effet d'une déclaration d'impôts modifiée, dans laquelle les contribuables peuvent promettre formellement de payer leurs impôts à temps. La formulation d'une telle promesse donne en outre accès à un tirage au sort permettant d'obtenir une récompense (financière ou en nature). En comparaison de contribuables qui bénéficient des mêmes récompenses, mais n'ont pas eu la possibilité de formuler une promesse, les résultats de l'étude montrent que la formulation d'une promesse adossée à une récompense en nature conduit à une amélioration substantielle du paiement de l'impôt. Les résultats de cette étude peuvent néanmoins difficilement être généralisés, car la formulation d'une promesse reste libre, et il est difficile de savoir si la promesse stimule effectivement le paiement de l'impôt, ou si les contribuables les plus disposés à respecter les délais impartis sont aussi ceux qui décident de formuler une promesse.

Les expériences en laboratoire permettent d'apporter une réponse plus affirmative à cette question. Un certain nombre de travaux (Zickfeld et al., 2024 ; Jacquemet et al., 2020) confirment en effet l'efficacité d'une forme particulière d'engagement, qui prend concrètement la forme d'un serment sur l'honneur à dire la vérité. En matière de fraude fiscale, un tel serment conduit à une augmentation massive, de l'ordre de 50 %, du montant d'impôt collecté. Cet effet est cependant concentré sur un sous-ensemble très particulier de contribuables : ceux dont les décisions, en l'absence d'engagement, se porteraient sur des choix intermédiaires qui ne correspondent ni à une préférence forte pour la fraude ni à une préférence forte pour l'honnêteté. Les fraudeurs convaincus, qui dissimulent l'intégralité de leur revenu, et les contribuables scrupuleux, qui en déclarent l'intégralité, sont quant à eux insensibles à l'engagement.

Les outils non monétaires de la politique fiscale fondés sur l'engagement tirent ainsi leur efficacité de leur capacité à faire basculer vers les comportements socialement appropriés les contribuables dont les écarts de comportement sont hasardeux plutôt que délibéré et consciemment choisis. Ces outils constituent un levier efficace, à la mise en place peu coûteuse, pour discipliner les fraudeurs « communes » commises par légèreté ou négligence. Ces mêmes outils semblent en revanche inefficaces sur les actes de fraude qui relèvent de décisions délibérées et mûrement choisies, contre lesquelles l'efficacité des dispositifs de sanctions

et de détection est au contraire avérée. En ce sens, les incitations non monétaires fondées sur l'engagement et les outils traditionnels de la politique fiscale sont complémentaires et ciblent des types de fraudes radicalement différents.

## Un éclairage qui doit laisser toute sa place au débat public

Bien qu'elle reste encore confidentielle pour le grand public, la révolution comportementale qu'a connue la réflexion en économie infuse désormais l'ensemble de ses champs d'application et permet de fonder des recommandations de politique publique qui prennent en compte les ressorts psychologiques des décisions économiques. Au-delà des *nudges*, qui font l'objet de réticences légitimes dans les sphères de la décision publique, les développements comportementaux de l'économie permettent d'élaborer de nouveaux outils d'intervention, mais aussi de mieux comprendre les effets d'outils traditionnels tels que les incitations financières. Ces recommandations peuvent être fondées scientifiquement grâce au recours à l'expérimentation en laboratoire, qui permet de tester

un vaste éventail d'interventions avant qu'elles ne soient mises en œuvre à grande échelle, ou validées empiriquement par des observations en conditions réelles. De telles expérimentations permettent d'éviter les errements les plus flagrants et d'ajuster le contenu des interventions envisagées afin qu'elles atteignent au mieux leurs objectifs.

L'économie comportementale comme l'économie expérimentale n'ont d'autre ambition que de décrire les conséquences sur les comportements des interventions envisagées. Ni l'une ni l'autre n'est en mesure (ni ne prétend) pouvoir trancher ce qui devrait être, quels objectifs doivent être poursuivis et quelle hiérarchie doit être établie entre les conséquences positives et négatives du déploiement d'une intervention de politique publique. Ces questions normatives, qui interrogent la légitimité des objectifs poursuivis et des moyens mobilisés pour les atteindre, relèvent du processus de délibération collective et de la prise de décision politique. L'éclairage de l'économie (comportementale) ne permet rien de plus que de s'accorder sur l'efficacité de certaines mesures par rapport aux objectifs préalablement définis. Ce faisant, elle peut néanmoins contribuer à libérer le débat public de ces questions pour qu'il puisse se porter sur la légitimité des fins et des moyens.

**Nicolas Jacquemet** est professeur d'économie à l'université Paris 1 Panthéon-Sorbonne (CES) et à PSE-École d'Économie de Paris, où il dirige le master *Economics and Psychology* ainsi que le projet de recherche « Approches croisées de la décision » de la Chaire Hermès-PSE « Ouvrir la Science économique ».

# Références

- Bott K. M., Cappelen A. W., Sørensen E. Ø. et Tungodden B. (2019)**, “You’ve Got Mail: A Randomized Field Experiment on Tax Evasion”, *Management Science*, 66, pp. 2801-2819.
- Choi J. J., Laibson D., Madrian B. C. et Metrick A. (2004)**, “For Better or for Worse: Default Effects and 401(k) Savings Behavior”, in D. A. Wise (Ed.), *Perspectives on the Economics of Aging*, University of Chicago Press.
- Eber N. et Willinger M. (2012)**, *L'économie expérimentale* (2<sup>nd</sup> éd.), La Découverte.
- Ferey S., Gabuthy Y. et Jacquemet N. (2013)**, « L'apport de l'économie expérimentale dans l'élaboration des politiques publiques », *Revue française d'économie*, vol. XXVIII, n° 2, 2013, pp. 155-194, <https://www.cairn.info/revue-francaise-d-economie-2013-2-page-155.htm>.
- Frey B.S. et Oberholzer-Gee F. (1997)**, “The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out”, *American Economic Review*, 87, pp. 746-755.
- Gabuthy Y., Jacquemet N. et L’Haridon O. (2021)**, *Économie comportementale des politiques publiques*. La Découverte.
- Gneezy U. et Rustichini A. (2000)**, “A Fine Is a Price”, *Journal of Legal Studies*, 29, pp. 1-17.
- Jacquemet N., Le Lec F. et L’Haridon O (2019)**, *Précis d'économie expérimentale*, Economica.
- Jacquemet N., Luchini S., Malézieux A. et Shogren J. F. (2020)**, “Who’ll stop lying under oath? Empirical evidence from tax evasion games”, *European Economic Review*, 124, 103369, <https://doi.org/10.1016/j.euroecorev.2020.103369>
- Jacquemet N., Luchini S. et Malézieux A. (2021)**, « Qui fraude le fisc ? », *La Vie des idées*, 22 juin, <https://laviedesidees.fr/Qui-fraude-le-fisc>.
- Joule R. V. et Beauvois J.-L. (1998)**, *La soumission librement consentie*, Presses universitaires de France.
- Koessler A.-K., Torgler B., Feld L. P. et Frey B. S. (2019)**, “Commitment to pay taxes: Results from field and laboratory experiments”, *European Economic Review*, 115, pp. 78-98.
- Mellström C. et Johannesson M. (2008)**, “Crowding Out in Blood Donation: Was Titmuss Right?”, *Journal of the European Economic Association*, 6(4), pp. 845-863.
- Solidaires-Finances publiques (2018)**, *Quand la baisse des moyens du contrôle fiscal entraîne une baisse de sa présence*, Rapport du syndicat national, <https://solidairesfinancespubliques.org/component/edocman/999-rapport-quand-la-baisse-des-moyens-du-controle-fiscal-entraîne-une-baisse-de-sa-présence-chute-inquietante-de-la-couverture-du-tissu-economique-et-fiscal/viewdocument/999.html?Itemid=>
- Paserman M. D. (2008)**, “Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation”, *The Economic Journal*, 118, pp. 1418-1452.
- Terestchenko M. (2005)**, *Un si fragile vernis d’humanité : banalité du mal, banalité du bien*, La Découverte.
- Thaler R. H. et Sunstein C. R. (2010)**, *Nudge : la méthode douce pour inspirer la bonne décision*, Éditions Vuibert.
- Zickfeld J. H. et al. (2024)**, “I Solemnly Swear I’m Up To Good: A Megastudy Investigating the Effectiveness of Honesty Oaths on Curbing Dishonesty”, Preprint, <http://dx.doi.org/10.31234/osf.io/hctxe>.

# Quand les « meilleures » preuves échouent : de l'usage politique des méta-analyses pour légitimer une politique de lutte contre le redoublement en Belgique francophone

Caroline De Pascale, Hugues Draelants et Jo Helgetun

**Cet article considère le rôle des preuves dans l'action publique à partir d'une étude de cas consacrée à l'utilisation des méta-analyses dans le domaine de l'éducation. Les méta-analyses sont considérées comme les meilleures preuves dans le paradigme de l'*evidence-based policy*, ce qui les rend particulièrement attractives pour les décideurs. En Belgique francophone, les méta-analyses ont été utilisées pour influencer la politique éducative, afin de légitimer la décision de supprimer le redoublement. Cependant, elles présentent des limites qui sont analysées dans l'article et dont les conséquences potentiellement négatives sont illustrées par l'échec de la politique de lutte contre le redoublement en Belgique francophone. Les auteurs concluent que l'usage des méta-analyses dans l'éducation confère aux décideurs un faux sentiment de certitude et qu'il nécessite une réflexion plus approfondie sur les finalités de l'éducation et les implications de ces preuves prétendument supérieures sur la politique éducative.**

Depuis une cinquantaine d'années, le fait de mobiliser des « preuves » est devenu incontournable, non seulement pour informer, mais aussi pour déterminer la politique publique, conduisant à ce que l'on appelle aujourd'hui communément l'élaboration de politiques fondées sur des preuves (*Evidence-Based Policy-Making*) (Draelants & Revaz, 2022). Bien que l'utilisation des travaux de recherche par le politique conduise parfois à une sélection partielle, voire partielle, des preuves retenues pour fonder l'action publique (*Policy-based Evidence-Making*) (Chung, 2016), cette montée en puissance des preuves semble conférer une légitimité renouvelée, garantie par la science, à la politique (Helgetun, 2021).

Dans ce nouveau paradigme, des universitaires de diverses disciplines ont noté que la méta-analyse

se situe au sommet de la hiérarchie des preuves, dans la mesure où elle est considérée comme un « étalon-or » (*gold standard*) de la recherche (Haidich, 2010 ; Knogler *et al.*, 2022). Méthode scientifique qui consiste à synthétiser statistiquement les résultats d'études indépendantes pour tirer des enseignements généraux sur un problème donné, les méta-analyses sont de manière croissante consultées, encouragées et utilisées par les gouvernements du monde entier pour justifier leurs réformes (Ahn *et al.*, 2012).

Cet article entend questionner les usages en plein essor des méta-analyses au cœur de l'action publique. En rappelant la définition et les origines de cette méthode, le texte qui suit interroge dans un premier temps l'attrait croissant que la méta-analyse a exercé sur les décideurs pour

fabriquer l'action publique, alors même qu'elle présente d'importantes limites, notamment lorsqu'elle est appliquée aux sciences sociales. Dans un deuxième temps, l'article s'appuie sur une étude de cas consacrée à la politique de redoublement en Belgique francophone. Depuis maintenant des années, le taux de redoublement y est exceptionnellement élevé par rapport à celui des autres pays de l'OCDE, ce qui a conduit à une volonté politique de le réduire. La politique initiée dans les années 1990 a donc été celle d'une lutte contre le redoublement, préalablement justifiée par des méta-analyses. En mettant en relief les effets nuancés de l'usage politique de cette méthode scientifique, l'article discute la place à accorder aux méta-analyses dans la politique éducative, et plus généralement dans l'action publique.

## Faire preuve par la synthèse : les méta-analyses, nouvel outil de légitimation de la politique publique

### Aux origines des méta-analyses

La méta-analyse est « une approche statistique qui a pour but de rassembler les résultats empiriques d'études singulières pour en faire une synthèse reproductible et quantifiée » (Laroche, 2015, p. 7). La méta-analyse se fonde donc sur deux éléments principaux : d'une part, elle synthétise les résultats d'études sur un sujet, nonobstant leurs qualités disparates. D'autre part, elle apporte un surcroît de savoir au regard des revues de littérature en faisant ressortir des résultats généraux, et augmente donc la capacité de la science à rendre compte d'une réalité sur un sujet. En se présentant comme la solution au problème posé par l'abondance d'informations dans la recherche (Glass, 1976), la méta-analyse a gagné en popularité, en particulier dans certains domaines. C'est en médecine qu'elle a acquis son statut actuel et s'y est imposée comme la norme d'excellence en matière de recherche. Depuis lors, elle est considérée comme une forme de connaissance essentielle prisée par les universitaires et les gouvernements qui cherchent à comprendre « ce

qui fonctionne le mieux » (Goldacre, 2013). Ces promesses ont contribué au succès incontestable des méta-analyses dans divers domaines, dont la psychologie où elles ont obtenu un statut privilégié de synthèse de référence, ainsi qu'en éducation, où leur usage est en pleine expansion.

Des chercheurs n'ont pas attendu qu'un « label » soit inventé pour synthétiser des études quantitatives portant sur un objet de recherche similaire. Il y a cent cinquante ans, les astronomes s'interrogeaient déjà sur la manière d'exclure (Peirce, 1852) puis de combiner (Airi, 1861) des observations de mesures physiques de faible qualité. Ceci a abouti à l'élaboration d'une théorie générale à ce sujet (Newcomb, 1886). Deux décennies plus tard, en épidémiologie, Pearson (1904) combinait différentes séries de données sur la fièvre entérique, dont il estimait la taille trop petite pour élaborer des conclusions si elles étaient prises individuellement. Dans les années 1930, c'est au travers de questions d'optimisation agricole (Dunlop, 1933 ; Lush, 1931) qu'une série d'ouvrages voit le jour pour synthétiser les résultats de divers travaux (Cochran, 1937 ; Snedecor, 1937), en même temps que les statisticiens poursuivent leurs recherches sur ce thème (Fisher, 1934 ; Pearson, 1933). Au milieu des années 1960, l'enseignement des techniques de synthèse sous-tendant la méta-analyse a déjà lieu dans les universités (Rosenthal, 1991). Le point de bascule de l'article de Glass (1976)<sup>1</sup>, n'est donc pas tant de développer une méthode d'analyse inédite, que de la nommer et d'en dessiner les contours.

### Une méthode présentant des limites scientifiques...

Depuis sa popularisation il y a près d'un demi-siècle (Glass, 1976 ; Smith & Glass, 1977), la méta-analyse a fait l'objet de réactions enthousiastes, mais aussi de vives controverses, y compris dans le champ où elle a été sans doute la plus développée, la médecine (Ioannidis, 2016).

Une des premières limites relevées par la littérature académique concerne la qualité des études incluses dans les méta-analyses. Ainsi, des données recueillies par des procédés de piètre qualité aboutissent à des méta-analyses de qualité tout aussi discutables (Ioannidis, 2016) (ce qui a été nommé le problème du « *junk in – junk out* »). Les limites de la méta-analyse proviennent aussi de caractéristiques méthodologiques qui lui sont propres, notamment

1 En 1976, Gene Glass publie un article intitulé « Primary, secondary, and Meta-analysis of Research », dans lequel il souligne que, face à l'abondance de recherches sur de nombreux sujets, la synthèse de données permet de faire émerger « des sources d'information jusque-là inexploitées » (p. 4), et déterminer, en particulier en éducation, les pratiques qui sont réellement efficaces.

lorsqu'il s'agit de saisir les causes des variations de résultats (Laroche, 2015). En outre, elle comporte parfois des défauts de conception spécifiques, liés aux critères de sélection des études intégrées dans l'analyse ainsi qu'aux problèmes de sélection des variables et d'hétérogénéité (Gurevitch et al., 2018 ; Kitsiou et al., 2017). Pour donner un exemple simple, si on synthétise des études qui évaluent l'effet de l'alimentation sur la prise de poids, mais qu'une étude inclut la présence d'écran lors des repas, tandis qu'une deuxième étude mesure par ailleurs l'activité physique, cela produira des résultats différents. Une méta-analyse a comme contrainte de tenir compte de cette hétérogénéité.

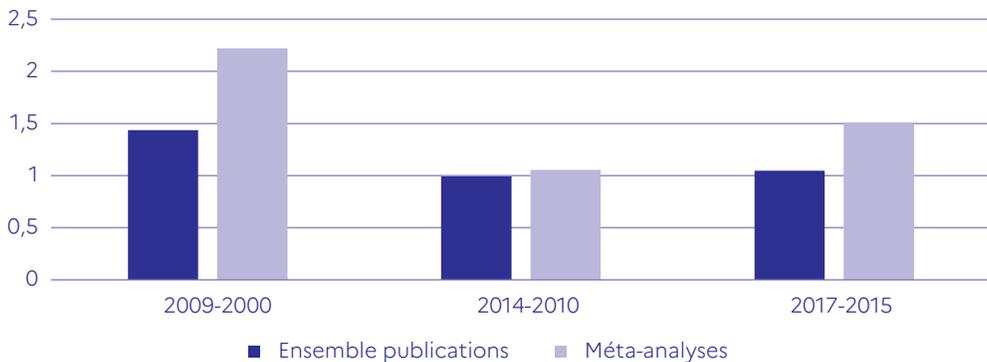
En somme, les critiques sur les méta-analyses s'articulent principalement sur deux principaux axes (Glass, 2015). La première critique consiste à se demander si on peut comparer des pommes et des poires. Après tout, sauf dans le cas de la réplification d'une autre étude, une méta-analyse va forcément intégrer des travaux qui ont porté sur des effets divers et des contextes différents. À cela, l'argument de Glass (2015) est de préciser que, dans le cas des pommes et des poires, la méta-analyse compare des fruits : c'est le critère de comparaison dont il faut évaluer la pertinence. Deuxième critique générale, celle de la terre plate (Glass, 2015) : les méta-analyses verraient le monde *a fortiori* social de manière trop lisse, ce qui favoriserait la possibilité de trouver des effets globaux, mais ferait perdre de vue les variations de contexte. Glass (2015) répond à cela que les discussions sur

la variabilité deviennent inutiles dès lors que la méta-analyse prouve qu'un effet fonctionne. Il prend l'exemple du monde éducatif : si on prouve que la réduction de la taille des classes améliore la qualité de l'enseignement, il n'est pas nécessaire de prolonger la discussion dans l'étude des contextes. Cependant, les méta-analyses sont loin de produire des résultats toujours aussi univoques, ce qui montre que tous les effets n'ont pas la possibilité d'être généralisés de la même manière.

Mais, et c'est notre argument principal, ces critiques méthodologiques proviennent surtout de disciplines scientifiques, comme la médecine, où la place et la légitimité des méta-analyses, bien qu'ayant donné lieu à des critiques partielles, ne sont pas fondamentalement discutées. *A fortiori*, dans les champs issus des sciences sociales, le recours aux méta-analyses pose de nombreuses questions, alors même que ces dernières y sont utilisées de manière croissante.

L'exemple de la recherche en éducation est à ce titre probant. L'éducation est en effet entrée dans un paradigme médical (Tröhler, 2015) fondé sur un positivisme social qui suppose qu'on peut, par exemple, décrire la meilleure manière d'enseigner. La méta-analyse a alors d'autant plus séduit qu'elle devait permettre de faire le tri au sein d'une littérature foisonnante et parfois contradictoire sur un même sujet (Glass, 1976). Dans ce contexte, le nombre de méta-analyses progresse plus rapidement que le nombre d'autres méthodes d'analyses.

**Comparaison des taux d'augmentation des publications et des méta-analyses dans le domaine de l'éducation**



Source des données : Education Resources Information Center (ERIC) ProQuest<sup>2</sup>

<sup>2</sup> Base de données relative aux publications en matière d'éducation financée par le département de sciences de l'Éducation des États-Unis.

Ainsi, les méta-analyses se présentent comme étant capables d'établir non seulement ce qui marche, mais surtout ce qui marche le mieux. Or, la question de ce qui marche le mieux n'est pas neutre : c'est une question axiologique qui ne peut se départir des finalités des systèmes éducatifs (Biesta, 2007). Bref, en éducation, ce qui marche dépend des missions de l'école qui sont aujourd'hui plurielles, à l'image de la pluralité même du monde social. L'usage de la méta-analyse en éducation pose donc un problème normatif dans la mesure où la valorisation intrinsèque du quantifiable et l'argumentaire démonstratif qui l'accompagnent (du type « *evidence shows* ») tendent à occulter la réflexion sur les finalités de l'éducation (Helgetun & Menter, 2022 ; Lynch et al., 2020).

Il peut paraître paradoxal que le nombre de méta-analyses, une méthode de synthèse, augmente plus rapidement que l'ensemble des autres travaux académiques en matière éducative. Au-delà des controverses purement scientifiques, il nous semble que l'une des explications au recours académique de ce type de travaux résulte du fait qu'ils sont jugés comme attractifs par le politique (Weingart, 1999).

#### Un effet de « science » contestable

Bruno Latour (2004) distingue deux types de faits scientifiques, ceux établis à de purs desseins de connaissance (*matters of fact*) et ceux conçus pour interpeller (*matters of concern*), pour capter l'attention et peser dans le débat public, à l'instar des faits rassemblés par le GIEC (Bensaude-Vincent et Dorthe, 2023). Les méta-analyses produisent des faits de ce second type puisqu'elles ont explicitement vocation à résumer toute l'information scientifique sur un sujet, en disant ce qui « marche » ou pas. Si elles ont été critiquées dans le champ académique, c'est justement sur des points qui en font tout l'attrait pour le politique.

En effet, en science, on reproche aux méta-analyses d'agréger des phénomènes hétérogènes, mais aussi d'aller trop loin et de manière trop caricaturale dans l'abstraction, qui pourtant est nécessaire dans les sciences (Demeulenaere, 2006). La question pertinente est donc celle de la qualité de l'abstraction, de son caractère plus ou moins valide pour rendre compte des phénomènes étudiés. Si le risque d'une modélisation excessivement simplifiée de la réalité peut donner lieu à des débats vifs dans le champ scientifique, dans le champ

politique les enjeux sont tout autres et la simplification y est généralement prisée. Comme l'indique Latour, la politique est un travail de traduction, de médiation et de représentation, les décideurs ne se soucient pas beaucoup de l'exactitude et d'un « transfert d'information exacte et sans déformation sur le monde social ou naturel » (Latour, 2002, p. 145). Il importe aussi de pouvoir trancher les débats, et la méta-analyse est un outil très puissant de ce point de vue, permettant de condenser un ensemble de recherches à la lumière d'un critère commun et de dire de quel côté penche la balance. En disant ce qui marche le mieux et donc ce qu'il faut faire, la méta-analyse intéresse *a priori* tout décideur.

De plus, s'appuyer sur ce type de connaissances permet de clore les débats et de dire ce qu'il faut faire avec autorité (compte tenu du statut éminent de la méthode). Les conclusions des méta-analyses sont perçues comme particulièrement solides et difficilement contestables. Les faits établis permettraient alors de dépolitiser les débats en dépassant les clivages partisans et *in fine* de (re)légitimer l'action publique, censée être plus rationnelle, puisque fondée sur la science. Autrement dit, l'attrait du politique pour les méta-analyses, comme pour les statistiques, tient dans ce que Pierre Bourdieu (1984) appelle l'« effet de science », ce qui est d'autant plus le cas ici puisque la méta-analyse est placée au sommet de la « pyramide des preuves ».

Enfin, la méta-analyse identifie des effets globaux. Là encore, cela peut paraître réducteur d'un point de vue scientifique, car les contextes ne sont pas pris en compte, mais c'est aussi ce qui intéresse les politiques. Les gouvernements sont confrontés à des populations, et l'action publique doit s'appliquer à grande échelle (Draelants et Revaz, 2022). Les méta-analyses facilitent la décision dans la mesure où elles permettent des mesures qui évitent de considérer les cas particuliers, de raisonner en moyenne et de faire émerger une action publique « efficace en probabilité ».

Pour toutes ces raisons, les méta-analyses sont faites pour intéresser l'action publique. Mais le prestige particulier dont jouit cette méthode est-il vraiment justifié ? L'efficacité politique est-elle au rendez-vous ? L'examen des usages politiques des méta-analyses dans la politique de redoublement en Belgique francophone fait apparaître des effets contrastés.

# Rendre l'action publique plus efficace par les méta-analyses : l'espoir déçu de la politique de redoublement en Belgique

## Les éléments-clés des réformes récentes de la politique belge sur le redoublement

La Belgique compte l'une des plus grandes proportions d'enfants ayant redoublé dans le monde occidental : en 2020 30,2 % des élèves de 15 ans ont redoublé dans l'ensemble de la Belgique. Dans le même temps, 46 % des jeunes de 15 ans de la Fédération Wallonie-Bruxelles (ou FWB, la partie francophone de la Belgique) ont redoublé. La FWB a le taux de redoublement le plus élevé de l'OCDE. À titre de comparaison, en 2020, la moyenne de l'OCDE est de 11 % (OCDE, 2020). Le redoublement a généré un coût supplémentaire d'environ 387,5 millions d'euros pour l'année 2019 (FWB,

2021). Ces taux de redoublement hors-norme sont donc devenus une préoccupation centrale des politiques éducatives en Belgique depuis le début des années 1990.

À ce jour, trois textes législatifs différents ont explicitement visé à réduire ces taux de redoublement dans l'enseignement secondaire<sup>3</sup> (voir tableau).

Depuis 30 ans, les discours politiques sur le redoublement en Belgique répètent que celui-ci produit des effets négatifs et qu'il faut le supprimer. Mais les taux de redoublement ne baissent pas. On peut donc affirmer que les réformes n'ont pas eu les effets escomptés aux yeux des principaux responsables politiques de l'enseignement. À cet égard, des politiciens et chercheurs belges ont identifié une série de causes à l'échec de la politique : il existerait une culture de l'échec scolaire et du redoublement en FWB ; les réformes auraient envoyé un mauvais signal aux enseignants en leur disant que les années supplémentaires étaient utiles pour réussir le diplôme ; les enseignants seraient favorables au redoublement et croiraient en ses effets bénéfiques, ce qui traduirait leur méconnaissance de la recherche sur le sujet (Crahay, 2019 ; Désir, 2014).

**Tableau : Législations visant à réduire le redoublement dans l'enseignement secondaire en FWB depuis 1990.**

Nom du texte	Contenu
Arrêté du Gouvernement de la Communauté française du 19 juillet 1993 modifiant l'arrêté royal du 29 juin 1984 relatif à l'organisation de l'enseignement secondaire.	Passage automatique de la première à la deuxième année de l'enseignement secondaire, suppression du redoublement pour les deux premières années de l'enseignement secondaire. À partir de la rentrée 1995-1996, le premier diplôme doit être obtenu en 3 ans maximum, avec la possibilité d'une année supplémentaire à la fin de la deuxième année qui est censée être différente du simple redoublement.
Décret relatif à l'organisation du 1 <sup>er</sup> degré de l'enseignement secondaire du 19/07/2001 (M.B. du 23/08/2001).	Abandon du principe de promotion automatique, introduction d'une année complémentaire qui « ne peut en aucun cas constituer une répétition de l'année précédente » (p. 1). À partir de la rentrée 2001-2002, un élève peut compléter le 1 <sup>er</sup> degré en 3 ans maximum et a la possibilité de faire des années complémentaires à la fin de la première année (1AC/1S) ou après la deuxième année (2CC/2S).
Circulaire n° 5795 du 30/06/2016 Circulaire générale relative à l'organisation de l'enseignement secondaire ordinaire et à la sanction des études.	Suppression de la 1 <sup>re</sup> année complémentaire dans l'enseignement secondaire à partir de la rentrée scolaire 2015-2016. Les élèves peuvent encore effectuer une deuxième année complémentaire (2S) à la fin de leur deuxième année d'enseignement secondaire.

<sup>3</sup> Un schéma similaire peut être observé dans l'enseignement primaire. Cependant, nous limitons ici notre propos aux trois premières années de l'enseignement secondaire, car c'est là que l'on trouve les taux de redoublement les plus élevés. En outre, c'est surtout sur les deux premières années de l'enseignement secondaire que se sont concentrées les tentatives de réduction du redoublement. Nos propres recherches empiriques ont donc porté sur ce segment de la scolarité.

L'étude approfondie de la conception et de la réception des réformes menées<sup>4</sup> montre que les enseignants ne sont pas clairement en faveur du redoublement (De Pascale et Draelants, 2022) et que, si les politiques publiques belges n'ont pas atteint leur objectif de réduction du redoublement et d'amélioration des résultats de l'apprentissage, c'est – pour partie – parce qu'elles se sont trop exclusivement appuyées sur des méta-analyses au détriment d'autres savoirs disponibles. Ce faisant, l'expérience belge peut être vue comme un cas d'école d'élaboration d'une politique fondée sur certaines preuves (ce à quoi se réduit une politique dite « *evidence-based* »), bien plus que d'une politique informée par les connaissances (*knowledge-informed policy*). Dans le premier cas, c'est un éventail limité de prétendues « meilleures preuves » qui détermine ce qu'une politique devrait être, tandis que dans le second, c'est un éventail plus large de sources qui documente la politique, incluant les divers types de recherches et les savoirs professionnels (Draelants et Revaz, 2022).

#### **Les méta-analyses dans les rapports d'expertise au fondement de la politique**

Avant de revenir sur ce qui a été ignoré par l'action publique lorsqu'elle se fondait sur les méta-analyses, il nous faut expliquer comment ces dernières ont été mobilisées concrètement pour bâtir une politique de lutte contre le redoublement. Précisons que le lien avec cette politique n'est bien entendu pas direct : jamais les débats parlementaires n'évoquent les méta-analyses en tant que telles. Le lien est toujours médié. Les responsables politiques évoquent plutôt « la recherche », les rapports qu'ils ont commandés pour s'informer de l'état de la recherche sur le sujet, ou le nom des experts mandatés, auteurs de ces rapports. Nous ne soutenons pas non plus que la recherche en éducation, par le biais des méta-analyses sur le redoublement, soit la seule cause d'inefficacité de la politique de réduction du redoublement. Le contexte dans lequel cette politique a été menée et les motivations budgétaires

qui furent premières dans la décision de remplacer le redoublement par la promotion automatique ont également joué un rôle important. Le ministre de l'Éducation reconnaissait le fait sans ambages. Dans la presse de l'époque, il déclarait : « C'est au seuil du secondaire que l'échec sévit le plus ; si l'on veut faire des économies, c'est là qu'il faut agir en priorité » (interview donnée au journal *Le Soir*, 21 juin 1994).

En 1991, un rapport réalisé par des experts belges pour l'OCDE dans la série « Examen des politiques nationales d'éducation » et intitulé *Les systèmes éducatifs en Belgique : similitudes et divergences*, comparait la situation de la Belgique francophone avec celle de ses voisins. La FWB y apparaît comme la championne du redoublement. Ce rapport est important, car il fut parmi les premiers à dénoncer les taux de redoublements particulièrement élevés et à mettre en évidence de notables disparités de performance en fonction des établissements, mais surtout car il fit grand bruit dans la presse à l'époque. Des rapports internationaux comme le *Rapport mondial sur l'éducation 1993* de l'UNESCO confirmeront la singularité de la FWB à cet égard. Par ailleurs, d'autres écrits d'origine nationale ont contribué à problématiser la situation du système scolaire belge francophone. Un document crucial à cet égard fut la *Radioscopie de l'enseignement* menée par une équipe interuniversitaire et parue en mars 1992. Selon la *Radioscopie*, le gonflement de la population scolaire imputable au redoublement s'élève au minimum à 8,17 %, soit environ 50 000 élèves, ce qui équivaut à une classe d'âge complète ajoutée à l'effectif scolaire. La *Radioscopie* met dès lors en évidence qu'il est possible de réaliser des économies significatives en supprimant le redoublement. Cet argument du coût du redoublement va, d'après les observateurs de l'époque, peser lourd dans la décision en raison de la situation financière délicate de la FWB qui, dans un contexte de crise budgétaire, cherche à tout prix à faire des économies. Intervient alors l'avis des chercheurs en éducation et un rapport de 1993 signé par Marcel Crahay (voir encadré 1).

4 Cette étude s'est notamment appuyée sur des entretiens avec des enseignants et des observations sur les pratiques de redoublement dans les écoles.

## Encadré 1 : L'important rôle de médiation joué par les chercheurs, le cas de Marcel Crahay

Parmi les experts ayant joué un rôle important dans le débat public sur le redoublement durant la période considérée, citons en particulier Marcel Crahay. C'est à travers lui et ses écrits que les méta-analyses vont influencer la politique éducative. Marcel Crahay, à l'époque directeur du Service de pédagogie expérimentale de l'université de Liège, est l'expert le plus souvent mentionné dans les discussions au parlement et en commission de l'éducation. Il est également intervenu comme l'un des cinq grands experts en pédagogie chargés de former les formateurs d'enseignants ; enfin, il a participé avec son équipe aux premières évaluations de la mise en place de la réforme du premier degré. Par ailleurs, il intervient personnellement de manière régulière au fil de la controverse à travers la publication de rapports, d'ouvrages scientifiques, d'essais, de tribunes et d'interviews dans les journaux pour dénoncer le redoublement. Mobilisé par les uns pour justifier le bien-fondé de la réforme, mis en cause par d'autres comme le prototype du pédagogue déconnecté des réalités, il a été un acteur très présent qui a joué des rôles divers dans le déroulement de la réforme initiale (celle de 1993) et des débats qui ont pu l'entourer. En effet, Crahay est un des seuls scientifiques à avoir été actif pratiquement à toutes les étapes du processus politique de changement. Son rapport, *Échec des élèves ? Échec de l'école ? La Communauté française de Belgique en échec scolaire* (1993), a joué un rôle important dans les premiers pas du processus et figure parmi les rapports déclencheurs ayant attiré l'attention publique sur les méfaits du redoublement.

Nous ne voudrions cependant pas laisser croire que Marcel Crahay a été le seul expert impliqué. D'autres chercheurs en éducation, souvent collaborateurs de ce dernier, sont intervenus à l'époque et plus encore dans les années 2010, notamment pour coécrire un autre rapport d'expertise sur le redoublement, intitulé *Le redoublement en Fédération Wallonie-Bruxelles* et paru en 2014 (le rapport est signé par Ariane Baye, Florent Chenu, Marcel Crahay, Dominique Lafontaine et Christian Monseur). Soulignons aussi que Marcel Crahay, malgré son rôle d'expert, n'est pas dénué de réflexivité scientifique et de lucidité sur les rapports entre science et politique. Il est conscient des limites des connaissances scientifiques pour transformer l'école. Ainsi, il souligne avec justesse qu'« en éducation et, plus largement, dans les affaires humaines, les faits objectifs ont moins de réalité que les états psychologiques » (Crahay, 2012, p. 476). Il met également en garde contre le scientisme et les dérives technocratiques qui pourraient en découler. Il rappelle ainsi que « les questions soulevées ne sont pas uniquement scientifiques. L'éducation est inéluctablement le lieu de choix axiologiques et, partant, d'affrontements idéologiques. Ce serait « verser dans l'obscurantisme que de les taire » (Crahay, 2012, p. 33). Par ailleurs, Crahay, comme les auteurs de la *Radioscopie*, plaide pour que les économies engendrées par la suppression du redoublement soient réinvesties dans l'organisation de remédiation, de soutien pédagogique préventif. Sur ce point, les responsables politiques n'ont pas suivi les recommandations des experts (Draelants, 2009), ce qui atteste bien un usage sélectif par le politique des preuves et recommandations formulées par les experts, comme nous l'évoquons en introduction.

Que contient ce rapport ? D'abord une reprise de la comparaison internationale : on y lit que « La Belgique a le triste privilège de présenter le taux de retard scolaire le plus élevé de l'Europe des Douze, mais aussi très probablement de l'Europe entière » (Crahay, 1993, p. 3). Ensuite, y est mentionné le « surcoût lié aux redoublements dans l'enseignement secondaire » qui « dépasse les 4 milliards de francs belges » (*Ibid.*, p. 37). Enfin, le rapport reprend les résultats de certaines méta-analyses centrées sur le contexte américain, depuis décriées par la littérature scientifique, allant dans le sens de la suppression du redoublement (voir encadré 2).

Il mobilise aussi deux autres études, l'une menée par Seibel en France, une autre réalisée par Bain dans le canton de Genève. Ce qui signifie que le rapport rédigé par Crahay, qui aboutit à une disqualification sans nuance du redoublement, se fonde uniquement sur des études menées dans d'autres contextes nationaux.

La même optique est également retenue dans le rapport Baye *et al.* de 2014, qui appuiera la réforme de 2016. Bien que mobilisant une bibliographie plus étoffée que le rapport de 1993, ses auteurs continuent d'une part d'invoquer la comparaison

internationale et d'autre part de se fier essentiellement à certaines études quantitatives et aux méta-analyses centrées sur le cas américain, en

citant cette fois les trois méta-analyses les plus anciennes (voir encadré 2), pour tirer des conclusions sur le phénomène du redoublement.

## Encadré 2 : Les méta-analyses sur le redoublement, des résultats scientifiques nuancés

Sur le thème du redoublement, cinq articles utilisant la méta-analyse ont été publiés au cours des cinquante dernières années : Allen *et al.* 2009 ; Goos *et al.* 2021 ; Holmes, 1989 ; Holmes et Matthews, 1984 ; Jimerson, 2001. En outre, l'article de Jackson (1975), encore influent dans les synthèses plus récentes bien qu'il s'agisse d'une synthèse narrative, constitue la première grande synthèse de tous les travaux réalisés sur le redoublement depuis la fin des années 1920 aux États-Unis (Draelants, 2018). En politique belge, parmi ces six références, quatre ont une visibilité particulière dans les débats et les textes politiques : la synthèse de Jackson (1975) et les trois plus anciennes méta-analyses (Holmes, 1989 ; Holmes & Matthews, 1984 ; Jimerson, 2001). Le rapport Crahay se réfère notamment explicitement à Jackson (1975) et à Holmes et Matthews (1984).

Ces trois anciennes méta-analyses ont pourtant été critiquées pour avoir intégré des études de qualité disparate (Lorence, 2006), contribuant à donner une image biaisée des recherches sur le redoublement (Allen *et al.*, 2009 ; Draelants, 2018 ; Lorence, 2006) et dissimulant que « la littérature est mixte et non pas unilatérale » (Alexander *et al.*, 2003, p. 30). En l'occurrence, ces méta-analyses ont diffusé l'idée que les effets du redoublement seraient globalement négatifs, alors même que d'autres études, certes moins nombreuses, ont mis en évidence les effets positifs du redoublement. En particulier, l'étude sociologique d'Alexander, Enwtisle et Dauber (2003) se distingue pour son caractère exceptionnel sur le plan méthodologique<sup>5</sup>.

Les deux dernières méta-analyses en date réalisées sur la question du redoublement (Allen *et al.* 2009 ; Goos *et al.* 2021), ont prêté davantage attention à l'importance de la qualité des études incluses dans leur synthèse. Elles ont aussi montré que les méta-analyses antérieures étaient basées sur des recherches aux techniques analytiques médiocres, ce qui avait eu pour effet de renforcer les résultats en défaveur du redoublement. Dès lors que les études sont modérées en fonction de leur qualité, l'effet du redoublement apparaît globalement nul au lieu d'être négatif (Allen *et al.*, 2009). La méta-analyse réalisée par Goos et ses collègues (2021) aboutit à un résultat proche. Elle précise en outre que, lorsque le redoublement est accompagné d'une remédiation spécifique, son effet est alors positif sur les résultats scolaires (avec l'effet le plus important recensé dans la méta-analyse). Bien qu'elles donnent une image fautive et désormais caduque des recherches sur le redoublement, les trois anciennes méta-analyses seront mobilisées tour à tour pour légitimer les réformes, et ce, sans que les critiques ultérieures ne soient mentionnées<sup>6</sup>.

Aucune recherche conduite en Belgique sur le redoublement n'est citée dans ces rapports. Cette ignorance du contexte, qui postule implicitement une validité universelle des résultats obtenus dans ces études, est pour le moins regrettable. En effet, dans les résultats scientifiques ambivalents des méta-analyses quant aux effets positifs, négatifs

ou nuls du redoublement, nos lunettes de sociologues voient des signaux d'alerte d'hétérogénéités inobservables qu'il faudrait creuser davantage. Nous faisons l'hypothèse que cela masque des effets différenciés que l'on observe sur le terrain (De Pascale et Draelants, 2022), mais également entre États. En effet, il est intéressant de

5 Notamment par la qualité de son échantillonnage de type aléatoire, son *design* longitudinal et les nombreuses données disponibles sur les élèves et leur histoire scolaire. La recherche est aussi inédite par sa durée : entamée en 1982, elle s'est achevée en 2002.

6 Ce constat est aussi vérifié au sein de la communauté scientifique. Entre 2010 et 2020, nous avons calculé que ces trois méta-analyses obtiennent 83,8 % des citations au regard de la dernière méta-analyse alors en date (Allen C. S. *et al.*, 2009) qui tenait compte des critiques énoncées sur les recherches antérieures. L'article présentant la méta-analyse de Jimerson se classe même au 46<sup>e</sup> rang des articles les plus cités en psychologie scolaire (McNicholas *et al.*, 2022).

se pencher sur l'origine géographique des publications portant sur le redoublement (celles qui vont être synthétisées dans les méta-analyses). Historiquement, les recherches étaient presque exclusivement produites aux États-Unis ; sur la dernière décennie, il s'agit encore de la majorité (53,9 %). Le pourcentage d'études réalisées en Belgique augmente, il est passé de 0 % durant les années 1990 à 8,9 % (soit 16 études) entre 2014 et 2023<sup>7</sup>. Or, si l'on peut considérer que ces 16 études portant sur le cas belge sont suffisantes (et même importantes relativement à la taille du pays), l'examen de l'origine de ces travaux montre qu'elles utilisent essentiellement des données produites dans la partie flamande du pays. Or, en Belgique, chaque communauté linguistique organise différemment son enseignement. Ces données ne sont pas forcément utilisables dans d'autres contextes, ce que précisent d'ailleurs régulièrement ces travaux (voir e.a. Baert & Picchio, 2021 ; Grabau & Van Damme, 2023 ; Van Canegem et al., 2021).

Retenons de tout ceci qu'il est difficile de tirer des conclusions définitives sur les effets du redoublement à partir des seules méta-analyses, et qu'il serait nécessaire de pousser plus loin les recherches, en mobilisant aussi des approches qualitatives avant de pouvoir fonder une politique pertinente en la matière. Si l'aspect pédagogique de la réforme a été en partie perçu par les enseignants comme un vernis masquant la réelle finalité du projet – réaliser des économies (Draelants, 2009) –, l'échec relatif de la politique de lutte contre le redoublement menée en Belgique francophone des années 1990 à nos jours s'explique aussi par les impensés d'une politique publique justifiée par les méta-analyses.

### Les méta-analyses confrontées au terrain : mise en évidence d'une déconnexion des pratiques<sup>8</sup>

Le recours aux méta-analyses sur le redoublement, qui a vocation à établir la légitimité de la réforme, apparaît aux enseignants comme une manœuvre politique, un prétexte plutôt qu'un fondement authentique. Les entretiens menés à l'époque (au cours de l'année scolaire 2002-2003) montrent bien que les enseignants ne sont pas dupes de l'argument pédagogique : ils pensent que la réforme vise avant tout à faire des économies, en renonçant dès lors à accompagner la promotion automatique de

mesures qui auraient permis de remédier efficacement aux difficultés d'apprentissage des élèves en échec scolaire.

*« Je trouve, moi, personnellement, que toutes ces réformes ont amené un nivellement par le bas. J'estime que tout le monde a sa chance, mais, pour moi, ça n'a jamais été une réforme sur le principe de l'égalité ! Ça a été une réforme économique ! Donc ça, c'est déjà une chose. Qu'on fasse attention à l'élève, qu'on essaie de lui trouver l'enseignement qui lui convient le mieux, c'est une bonne chose. Mais je pense que ça coûterait beaucoup plus d'argent que ce qu'ils nous ont donné... » (Enseignante de mathématiques et de sciences, Collège A)*

*« Politiquement parlant, ça n'a jamais été décidé que pour faire des économies. Point final. Et puis on a trouvé la façon la plus stratégique de nous faire passer ça. N'importe quoi ! Oui, je suis contre un redoublement abusif. Mais alors qu'on réfléchisse ! Et qu'on trouve une autre façon de faire. » (Enseignant de géographie, Collège B)*

Lorsqu'on analyse en détail les raisons pour lesquelles les enseignants continuent d'adhérer au redoublement, malgré le discours politique et académique décrivant ses effets délétères sur les élèves, on constate que le redoublement remplissait diverses « fonctions latentes » (Draelants, 2009) : gérer l'hétérogénéité des élèves en les « triant » au sein des établissements ; réguler l'ordre scolaire et le comportement des élèves ; se distinguer des écoles voisines ; et maintenir l'autonomie professionnelle des enseignants qui décident du sort des élèves en fin d'année.

Le fait d'ignorer ces fonctions latentes a eu des conséquences néfastes sur la mise en œuvre de la réforme, rejetée avec véhémence par les enseignants. Considérant qu'elle était irréalisable, ils ont affirmé qu'ils n'étaient pas tenus d'agir loyalement dans le cadre de la réforme. Ils ont revendiqué leur autonomie pédagogique et déclaré leur refus de changer leur façon de faire la classe ou d'adapter leurs méthodes d'enseignement.

*« En première année du secondaire, on doit théoriquement remettre tout le monde sur un pied d'égalité, mais c'est impossible. Vu les classes super-nombreuses, c'est impossible de travailler*

<sup>7</sup> Calculs réalisés en mars 2024 à partir de la base de données Scopus.

<sup>8</sup> Ce paragraphe est fondé sur une première enquête, menée en 2002-2003 qui avait pour objectif de comprendre les résistances aux réformes pédagogiques visant à réduire le redoublement. Les extraits sont issus d'entretiens menés à cette période auprès d'enseignants issus de trois établissements d'enseignement secondaire contrastés de la Fédération Wallonie-Bruxelles (Draelants, 2009).

*individuellement pour les remettre à flot. Ces réformes, c'est du bla-bla : sur papier, c'est facile de changer les choses, mais sur le terrain, là, on se rend compte des contraintes qui pèsent sur les enseignants ! Ceux qui pondent ces décrets, ils ne savent pas ce que c'est, hein ! » (Enseignant d'éducation physique, Collège A)*

*« Il y a des gens qui sont payés pour penser. Bon ben ils pensent et ils essaient de réaliser des choses, qui parfois sont bien, parfois qui sont irréalisables sur le terrain. Ça c'est très beau les discours, mais il faut venir dans les classes pour voir ça. » (Enseignante de géographie et d'éducation scientifique, Collège C)*

L'on observe alors que les enseignants n'adhèrent pas aux conclusions de la recherche issue des méta-analyses sur le redoublement dont ils ont entendu parler, ou que des conseillers pédagogiques leur ont présentées. Ils ont particulièrement du mal à croire que le redoublement est toujours négatif et que toutes les études sont unanimes à ce sujet, ce qui, on l'a vu, n'est en réalité pas le cas, mais qui tend à être présenté comme tel dans la mesure où les méta-analyses masquent les désaccords existants en mettant en avant l'effet moyen du redoublement. Dans certaines situations, ils estiment le redoublement utile, notamment pour se maintenir dans l'enseignement général et éviter une réorientation vers une filière technique ou professionnelle non souhaitée. Dans ce cas précis, la suppression du redoublement peut conduire à une relégation directe des élèves en échec, en empêchant les enseignants de leur donner une seconde chance.

*« Je me souviens d'une élève qui est arrivée chez nous en deuxième. Elle a fait une première 2<sup>e</sup> nulle, une deuxième 2<sup>e</sup> faible et a trouvé le moyen de faire une troisième 2<sup>e</sup> puis n'a plus jamais eu aucun problème scolaire ! Alors, quand je pense à cela, je me dis, voilà une enfant qui [dans la situation actuelle de suppression du redoublement] aurait été en 3<sup>e</sup> P [3<sup>e</sup> année d'enseignement professionnel] et qui avait autre chose dans ses mains que pour faire une 3<sup>e</sup> P ! Ça, c'est un cas exceptionnel peut-être. Une gamine qui n'a eu sa maturité de fin de seconde qu'à quinze ans, quinze ans et demi. Mais ceux-là, c'est fini maintenant. On ne peut même plus [leur donner une seconde chance de se maintenir*

*dans l'enseignement général], même si on le voulait ! On ne peut pas ! » (Enseignante de mathématiques, physique et sciences économiques, Collège A)*

Si les enseignants ont du mal à comprendre la recherche, c'est aussi parce que la manière de concevoir le redoublement et de mesurer son efficacité dans les études que synthétisent les méta-analyses est déconnectée de celle qui prévaut sur le terrain. En effet, lorsque la réforme a été implantée, les spécificités des usages du redoublement dans le cas belge n'étaient pas connues. En conséquence, aucune alternative pour gérer l'hétérogénéité des classes n'a été prévue. En outre, le recours à la note chiffrée et aux évaluations a été maintenu, et les programmes n'ont pas été changés. De la sorte, on a subitement demandé aux élèves qui travaillaient pour la note, parce qu'il y avait un bulletin, de s'intéresser intrinsèquement aux apprentissages. Cela n'a pas fonctionné<sup>9</sup>.

Concernant le redoublement en tant que tel, les enquêtes quantitatives individualisent toujours la pratique. Tout se passe comme si chaque redoublement recouvrait le même sens, quel que soit le contexte dans lequel il intervient, quel que soit l'élève, l'établissement ou le niveau d'étude envisagé. Pourtant, sur le terrain, le redoublement revêt des sens différents. Il vise par exemple tantôt à apprendre des méthodes de travail, à améliorer l'estime de soi, ou à éviter une réorientation. Nos recherches récentes, qui étudient le redoublement qualitativement et du point de vue des acteurs concernés, montrent que la finalité qu'élèves et enseignants lui attribuent varie et qu'entre deux élèves qui sont maintenus dans l'enseignement général ou professionnel, les causes et donc aussi les expériences du redoublement ont peu de points communs (De Pascale & Draelants, 2022).

## Conclusion

L'abstraction est inévitable en science, mais on peut légitimement se demander si celle qui consiste à subsumer sous un même terme des réalités disparates est une bonne abstraction, ou si nous n'avons pas besoin de concepts plus fins pour distinguer différents types de redoublements. C'est donc la question de l'hétérogénéité des redoublements, de ses effets différenciés qui sont éludés par les

<sup>9</sup> Les enseignants constatent et déplorent qu'il devient très difficile de mettre les élèves au travail quand la menace que constituait le redoublement disparaît. Ayant développé un rapport instrumental à la scolarité, les élèves travaillent en effet essentiellement « pour les notes » et « pour passer de classe ».

méta-analyses. Choses que certaines enquêtes quantitatives cherchent désormais à appréhender en recourant à des méthodes toujours plus sophistiquées (Gary-Bobo *et al.*, 2016), mais que l'on peut aussi étudier qualitativement et que l'on perçoit alors de manière évidente lorsque l'on prend la peine d'interroger les acteurs scolaires concernés par le phénomène (les enseignants, les élèves).

Concernant la méthode employée dans la plupart des études quantitatives, on peut là aussi constater un décalage entre le raisonnement des chercheurs et le raisonnement naturel. Dans la plupart des études<sup>10</sup>, on procède d'une manière quasi-expérimentale : les élèves qui redoublent sont soit comparés à des élèves du même âge que le leur, mais qui ont été promus (et qui sont donc scolarisés dans l'année supérieure), soit comparés à des élèves plus jeunes, mais qui fréquentent la même classe qu'eux (celle qu'ils redoublent). Autrement dit, les chercheurs adoptent un raisonnement contrefactuel. La méthode qu'ils utilisent pour mesurer l'efficacité du redoublement revient à se demander comment l'élève aurait progressé dans ses apprentissages s'il avait été promu plutôt que maintenu<sup>11</sup>. Ici aussi on peut légitimement s'interroger sur la pertinence de l'abstraction employée, car cette façon de faire diffère de celle qui compare un élève non pas à un autre élève, mais à lui-même. Ce raisonnement, qui peut être qualifié de longitudinal, correspond davantage au mode de raisonnement auquel recourent spontanément les acteurs de terrain. La première méthode n'est pas forcément mauvaise, mais elle produit une abstraction en décalage avec le raisonnement naturel. Il est donc intéressant d'employer les deux méthodes et de comparer les résultats. En l'occurrence, les deux méthodes aboutissent à des conclusions différentes. Alors que la première tend à montrer des effets globalement nuls du redoublement (comme le montrent les dernières méta-analyses), la seconde adoptée dans certaines recherches, comme celle d'Alexander *et al.* (2003), met en évidence de nets progrès pour les redoublants. Pour leur permettre de rebondir et de réduire l'écart qui les sépare de leurs condisciples, le redoublement apparaît alors pertinent. Les études qualitatives suggèrent que cette manière de procéder est plus proche du raisonnement naturel, car

sur le terrain l'objectif que les enseignants, mais aussi les familles et les élèves eux-mêmes, assignent communément au redoublement est de permettre au redoublant de se relancer et de s'améliorer (même si cela ne signifie pas la même chose pour chaque élève). Il s'agit moins de se comparer aux autres que de se comparer à soi.

Cette observation permet d'attirer l'attention sur le fait que les choix méthodologiques sur lesquels repose la science (dominante) du redoublement, celle synthétisée dans les méta-analyses, ne sont pas neutres. La question de l'efficacité du redoublement n'est neutre que d'apparence, en réalité elle n'est pas indépendante de valeurs puisque, pour répondre à la question, il faut d'abord définir de quelle efficacité il s'agit. Mais faire valoir la pluralité des définitions possibles de l'efficacité est difficile. En effet, le point de vue et les savoirs des enseignants ne sont pas considérés comme ayant de la valeur dans la politique fondée sur les preuves. C'est une chose que ressentent bien les enseignants qui s'estiment dépossédés de leur professionnalité quand les experts monopolisent la manière de définir et de poser les problèmes (Draelants et Revaz, 2022).

Les méta-analyses possèdent des particularités qui les rendent attractives pour les décideurs politiques et pour les scientifiques qui veulent se faire experts et intervenir dans l'arène politique. Elles semblent donc être d'efficaces outils de légitimation pour l'action publique, mais en réalité, comme nous l'avons vu, ce n'est pas le cas. Autrement dit, cette attractivité des méta-analyses est en quelque sorte un cadeau empoisonné. Celles-ci sont efficaces pour faciliter la prise de décision, mais ne le sont pas pour construire de l'acceptabilité, pour légitimer la réforme auprès des acteurs. Lorsqu'elles sont le seul type d'étude pris en considération, les méta-analyses renforcent plutôt la méfiance des acteurs scolaires envers la politique et l'expertise, et le sentiment que les experts sont déconnectés du réel.

En plaçant les savoirs produits par les méta-analyses au-dessus des autres et en légitimant la politique de lutte contre le redoublement essentiellement par une connaissance fondée sur des méta-analyses, qui plus est relativement anciennes

**10** Toutes les recherches quantitatives ne procèdent pas de la sorte, c'était notamment le cas de celle conduite par Alexander *et al.* (2003) qui procédait à des comparaisons différentes, de type « pré-post » où les sujets ne sont pas comparés à d'autres élèves, mais à eux-mêmes. Leur étude étant menée sur une longue durée, les élèves pouvaient en effet être leurs propres contrôles puisque leurs résultats sont connus avant l'année de redoublement, pendant celle-ci et après celle-ci.

**11** Les deux comparaisons sont par principe affectées d'un biais d'échantillonnage. La première comparaison est défavorable au redoublant et risque de sous-estimer les effets du redoublement (les élèves du même âge sont plus loin dans leur scolarité et suivent donc un programme scolaire plus avancé), la seconde lui est favorable et risque de surestimer les effets du redoublement (le redoublant a déjà pu se familiariser une première fois avec la matière, ce que n'ont pu faire ses condisciples)

et discutées, le risque est que les experts donnent un sentiment de fausse certitude aux décideurs. Cela laisse penser qu'il n'est pas nécessaire de nourrir l'information d'autres méthodes scientifiques (plus qualitatives) et d'autres savoirs (notamment professionnels). L'on peut supposer que les décideurs sont alors convaincus de disposer de données solides sur lesquelles fonder leur action plutôt que de préjugés idéologiques, or faits et valeurs politiques ne sont pas dissociables si aisément. Les choix méthodologiques sur lesquels la science du redoublement repose ne sont pas neutres.

Faire droit aux autres formes de preuve et cesser de placer les méta-analyses et les méthodes

sur lesquelles elles se fondent sur un piédestal serait l'approcher à retenir. D'autres épistémologies, constructivistes ou post-positivistes, qui appréhendent différemment le monde social, et d'autres méthodologies, qualitatives notamment, sont nécessaires pour embrasser de manière plus large tout phénomène social, le situer dans son contexte et le comprendre en profondeur. Loin d'être antiscience, une telle démarche est au contraire plus exigeante d'un point de vue épistémologique. Faute de pluralisme épistémologique dans les données qui l'informent, l'action publique en éducation, mais également dans tout autre domaine, risque fort d'être confrontée à des déconvenues comme l'illustre l'échec de trente ans de lutte contre le redoublement en FWB.

**Caroline De Pascale** est doctorante sous mandat d'assistante à l'UCLouvain, membre du Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation (GIRSEF).

**Hugues Draelants** est professeur à l'UCLouvain, membre du GIRSEF.

**Jo Helgetun** est chargé de recherches F.R.S-FNRS, membre du GIRSEF.

## Références

- Ahn S., Ames A. et Myers N. (2012)**,  
 "A Review of Meta-Analyses in Education."  
*Review of Educational Research*, 82, pp. 436-476.
- Airi G. B. (1861)**,  
*On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.*
- Alexander K., Entwisle D. et Dauber S. (2003)**,  
*On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades, Second Edition*, Cambridge University Press.
- Allen C. S., Chen Q., Willson V. et Hughes J. N. (2009)**,  
 "Quality of Research Design Moderates Effects of Grade Retention on Achievement: A Meta-Analytic, Multilevel Analysis",  
*Educational Evaluation and Policy Analysis*, 31, pp. 480-499.
- Baert S. et Picchio M. (2021)**,  
 "A signal of (Train)ability? Grade repetition and hiring chances", *Journal of Economic Behavior & Organization*, 188, pp. 867-878.
- Baye A., Chenu F., Crahay M., Lafontaine D. et Monseur C. (2014)**,  
*Le redoublement en Fédération Wallonie-Bruxelles. Rapport d'expertise* commandité par le groupe parlementaire du Parti Socialiste.
- Bensaude-Vincent B. et Dorthe G. (2023)**,  
*Les sciences dans la mêlée*, Seuil.
- Biesta G. (2007)**,  
 "Why 'What Works' won't work: Evidence-based practice and the democratic deficit in educational research",  
*Educational Theory*, 57, pp. 1-22.
- Bourdieu P. (1984)**,  
*Homo Academicus*, Minuit.
- Chung J. (2016)**,  
 "The (mis)use of the Finnish teacher education model: 'policy-based evidence-making'?", *Educational Research*, 58(2), pp. 207-219.

- Cochran W. G. (1937),**  
 “Problems Arising in the Analysis of a Series of Similar Experiments”, *Supplement to the Journal of the Royal Statistical Society*, 4(1), pp. 102-118.
- Crahay M. (2019),**  
*Peut-on lutter contre l'échec scolaire ?*, De Boeck.
- Crahay M. (2012),**  
*L'école peut-elle être juste et efficace ? De l'égalité des chances à l'égalité des acquis*, De Boeck.
- Crahay M. (1993),**  
 « Échec des élèves, échec de l'école ? », *Recherche en éducation. Théorie et Pratique*, 11.
- Demeulenaere P. (2006),**  
 « Présentation : le problème de l'abstraction en sociologie », *L'Année sociologique*, 56(1), pp. 253-262.
- De Pascale C. et Draelants H. (2022),**  
 « L'enseignant et le redoublement. Au-delà des idées reçues », *Les Cahiers de Recherche du Girsef*, 128.
- Désir C. (2014),**  
*La remédiation à l'échec scolaire*. Rapport déposé au Parlement de la Communauté française.
- Draelants H. (2009),**  
*Réforme pédagogique et légitimation. Le cas d'une politique de lutte contre le redoublement*, De Boeck.
- Draelants H. (2018),**  
 « Le redoublement est-il vraiment moins efficace que la promotion automatique ? Une évidence à réinterroger », *Les Cahiers de Recherche du GIRSEF*, 113, pp. 1-22.
- Draelants H. et Revaz S. (2022),**  
*L'évidence des faits. La politique des preuves en éducation*, PUF.
- Dunlop G. (1933),**  
 “Methods of Experimentation in Animal Nutrition”, *The Journal of Agricultural Science*, 23(4), pp. 580-614.
- Fisher R. A. (1934),**  
*Statistical methods for research workers*, 5<sup>e</sup> édition, Oliver & Boyd.
- Gary-Bobo R. J., Goussé M. et Robin J. M. (2016),**  
 “Grade retention and unobserved heterogeneity”, *Quantitative Economics*, 7(3), pp. 781-820.
- Glass G. V. (1976),**  
 “Primary, Secondary, and Meta-Analysis of Research”, *Educational Researcher*, 5, pp. 3-8.
- Glass G. V. (2015),**  
 “Meta-analysis at middle age: a personal history”, *Research Synthesis Methods*, 6(3), pp. 221-231.
- Goldacre B. (2013),**  
*Building Evidence into Education [Rapport]*, pp. 1-19.
- Goos M., Pipa J. C. et Peixoto F. (2021),**  
 “Effectiveness of grade retention: A systematic review and meta-analysis”, *Educational Research Review*, 34, pp. 1-14.
- Grabau L. J. et Van Damme J. (2023),**  
 “School Stratification and Science Climate in Early Secondary Education in Ireland and Flanders: Associations with Students' Science Dispositions and Science Literacy”, *International Journal of Science and Mathematics Education*.
- Gurevitch J., Koricheva J., Nakagawa S. et Stewart G. (2018),**  
 “Meta-analysis and the science of research synthesis”, *Nature*, 555(7695), pp. 175-182.
- Haidich A. B. (2010),**  
 “Meta-analysis in medical research”, *Hippokratia*, 14(Suppl. 1), pp. 29-37.
- Helgetun J. et Menter I. (2022),**  
 “From an age of measurement to an evidence era? Policy-making in teacher education in England”, *Journal of Education Policy*, 37(1), pp. 88-105.
- Helgetun J. B. (2021),**  
*Maintaining and changing institutionalized environments: teacher education policymaking in England and France compared*, Université de Louvain.
- Holmes C. T. (1989),**  
 “Grade-level retention effects: A meta-analysis of research studies”, in L. A. S. M. L. Smith (Ed.), *Flunking grades: Research and policies on retention* (pp. 16-33), The Falmer Press.
- Holmes C. T. et Matthews K. M. (1984),**  
 “The Effects of Nonpromotion on Elementary and Junior High School Pupils: A Meta-Analysis”, *Review of Educational Research*, 54, pp. 225-236.

- Ioannidis J. P. A. (2016)**,  
 “The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses”, *The Milbank quarterly*, 94(3), pp. 485-514.
- Jackson G. B. (1975)**,  
 “The Research Evidence on the Effects of Grade Retention”, *Review of Educational Research*, 45(4), pp. 613-635.
- Jimerson S. (2001)**,  
 “Meta-analysis of Grade Retention Research: Implications for Practice in the 21<sup>st</sup> Century”, *School Psychology Review*, 30, pp. 420-437.
- Kitsiou S., Pare G., Jaana M. et Gerber B. (2017)**,  
 “Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews”, *PLoS One*, 12(3), e0173160.
- Knogler M., Hetmanek A. et Seidel T. (2022)**,  
 “Determining an Evidence Base for Particular Fields of Educational Practice: A Systematic Review of Meta-Analyses on Effective Mathematics and Science Teaching”, *Frontiers in Psychology*, 13.
- Laroche P. (2015)**,  
*La méta-analyse : Méthodes et applications en sciences sociales*, De Boeck.
- Latour B. (2002)**,  
 « Et si l'on parlait un peu politique ? », *Politix*, 15(58), pp. 143-165.
- Latour B. (2004)**,  
 “Why has critique run out of steam? From matters of fact to matters of concern”, *Critical inquiry*, 30(2), pp. 225-248.
- Lorence J. (2006)**,  
 “Retention and Academic Achievement Research Revisited from a United States Perspective”, *International education journal*, 7, pp. 731-777.
- Lush J. L. (1931)**,  
 “Predicting gains in feeder cattle and pigs”, *Journal of Agricultural Research*, 42, pp. 853-881.
- Lynch J., Rowlands J., Gale T. et Parker S. (2020)**,  
 “An outline of a theory of practice methodologies: Education research as an expansive activist endeavour”, in J. Lynch, J. Rowlands, T. Gale et S. Parker (Eds.), *Practice Methodologies in Education Research*, Routledge.
- Newcomb S. (1886)**,  
 “A Generalized Theory of the Combination of Observations so as to Obtain the Best Result”, *American Journal of Mathematics*, 8(4), pp. 343-366.
- Pearson K. (1904)**,  
 “Report on Certain Enteric Fever Inoculation Statistics”, *British Medical Journal*(3), pp. 1243-1246.
- Pearson K. (1933)**,  
 “On a Method of Determining Whether a Sample of Size n Supposed to Have Been Drawn from a Parent Population Having a Known Probability Integral has Probably Been Drawn at Random”, *Biometrika*, 25(3/4), pp. 379-410.
- Peirce B. (1852)**,  
 “Criterion for the rejection of doubtful observations”, *Astronomical Journal*, 2(45), pp. 161-163.
- Rosenthal R. (1991)**,  
*Meta-Analytic Procedures for Social Research*, Sage Publications.
- Smith M. L. et Glass G. V. (1977)**,  
 “Meta-analysis of psychotherapy outcome studies”, *The American psychologist*, 32/9, pp. 752-760.
- Snedecor G. W. (1937)**,  
*Statistical Methods applied to experiments in Agriculture and Biology*, Collegiate Press.
- Tröhler D. (2015)**,  
 “The medicalization of current educational research and its effects on education policy and school reforms”, *Discourse Studies in the Cultural Politics of Education*, 36, pp. 749-764.
- Van Canegem T., Van Houtte M. et Demanet J. (2021)**,  
 “Grade retention and academic self-concept: A multilevel analysis of the effects of schools’ retention composition”, *British Educational Research Journal*, 47(5), pp. 1340-1360.
- Weingart P. (1999)**,  
 “Scientific expertise and political accountability: paradoxes of science in politics”, *Science and public policy*, 26(3), pp. 151-161.

# Note réactive

## États-Unis



Publiée dans le cadre de l'activité de veille internationale en gestion publique du bureau de la recherche de l'IGPDE, la *Note réactive* décrit des expériences administratives réalisées dans d'autres pays du monde dans les cinq grands domaines du management public (budget et performance, gouvernance, relation à l'utilisateur, emploi public, transformation numérique).



# États-Unis : l'échelle de preuves, un nouvel outil pour justifier les décisions budgétaires

Amicie de Tanoüarn

**Face à la quantité importante de données à traiter chaque année lors des exercices de programmation budgétaire, l'Office of Management and Budget de l'État de Rhode Island aux États-Unis décide en 2020 de reconfigurer et uniformiser les formulaires de demande budgétaire et crée une échelle des preuves à partir desquelles évaluer la pertinence des demandes. Bien qu'encore imparfaite, cette approche semble avoir d'une part, permis un meilleur ciblage et une expression plus claire des demandes, et d'autre part, de rendre des arbitrages budgétaires mieux documentés et davantage discutés.**

À chaque cycle budgétaire, tous les décideurs publics ont à étudier des centaines, voire des milliers de propositions budgétaires nouvelles ou actualisées en un temps restreint. Ces propositions ne sont que rarement accompagnées d'évaluation de leurs résultats espérés ; tout au plus contiennent-elles de nombreuses informations que les décideurs doivent rapidement assimiler et analyser pour juger de leur pertinence. La question de la preuve de l'efficacité des programmes budgétaires est donc une donnée essentielle à la prise de décision politique afin de garantir le « bon » usage des deniers publics.

Face à ce constat, l'État de Rhode Island aux États-Unis a décidé de refondre l'architecture des documents à utiliser pour soumettre une proposition budgétaire : ceux-ci sont uniformisés sous un même format commun à toutes les agences et services administratifs. Adoptant une approche centrée sur les utilisateurs, ce modèle s'adosse à un nouvel outil, une échelle de preuves, qui doit permettre aux responsables politiques de saisir rapidement si chaque proposition budgétaire est soutenue par un faisceau de preuves susceptibles de garantir en théorie l'atteinte des objectifs présentés.

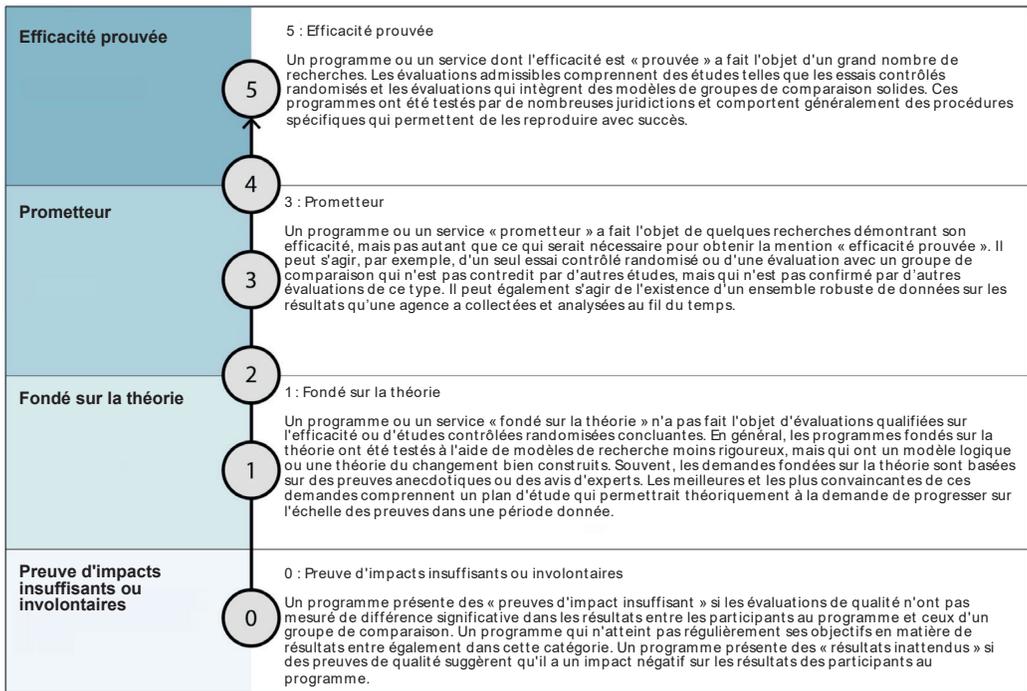
## Une échelle de preuves pour objectiver les propositions budgétaires

L'Office of Management and Budget (OMB) a été créé en 2012 par l'État américain de Rhode Island afin de réfléchir aux conditions de financement, d'organisation et de déploiement du budget de la collectivité. Dès ses premiers pas, ce département se rend compte de la quantité importante de propositions budgétaires qu'il a à traiter lors d'un processus budgétaire d'une durée relativement contrainte. Devant analyser plus de 500 dossiers provenant de 55 agences différentes en l'espace de 16 semaines, les équipes budgétaires prennent acte par ailleurs de la variété des thématiques inhérentes à l'action publique de l'État (aide financière aux petites entreprises, programme d'analyse à quai des toxines des crustacés, prise en charge par Medicaid des soins dispensés par les doulas, infrastructures de cybersécurité, remplacement du mobilier...), tout autant que du degré extrêmement variable des méthodes et des données techniques utilisées pour justifier la nécessité de la dépense. L'omission de détails-clés et de preuves permettant de juger de l'opportunité de financer certains projets ou programmes est alors considérée par l'OMB comme un frein majeur à une décision politique éclairée.

Face à cette hétérogénéité de l'information budgétaire, autant quantitative que qualitative, l'OMB, après avoir consulté le bureau du Gouverneur, décide, en janvier 2020, de lancer un projet de reconfiguration et d'uniformisation des formulaires de demande budgétaire soumis par les agences et les départements techniques de l'État. Dans l'objectif d'améliorer l'évaluation *ex ante* des propositions budgétaires opérée par les demandeurs, il s'associe avec le *Policy lab* de la *Brown University* pour remanier le design du *decision package*<sup>1</sup>, formulaire utilisé pour les demandes budgétaires. En adoptant une

approche centrée sur l'utilisateur<sup>2</sup> et en s'inspirant des « bonnes » pratiques développées par d'autres centres de recherche américains spécialisés dans les *evidence-based policies*<sup>3</sup>, le laboratoire procède à un état des lieux des connaissances sur les différents processus budgétaires existants aux États-Unis qu'il transmet à l'OMB. Au-delà de l'uniformisation du formulaire de propositions budgétaires, l'OMB décide de créer une échelle de preuves à partir de laquelle les agences et les départements techniques doivent évaluer la pertinence et l'efficacité de leurs propositions (voir figure).

**Figure – L'échelle de preuves de l'État de Rhode Island**



Source : <https://thepolicylab.brown.edu/projects/ri-budget-template>. Document traduit par la rédaction.

1 <https://thepolicylab.github.io/budget-template/>

2 <https://thepolicylab.brown.edu/projects/ri-budget-template>

3 Le *Policy lab* s'appuie notamment sur l'expertise du *Pew Research Center*, un centre de recherche situé à Washington, et du fonds philanthropique *Arnold Ventures*.

Dans le même temps, l'OMB met en œuvre des formations et des réunions d'échanges avec les utilisateurs pour les acculturer au nouvel outil et les aider à formuler des propositions budgétaires les plus claires et probantes possibles.

## Une dynamique en voie de pérennisation et d'expansion

Alors que le nouveau formulaire et l'échelle de preuves sont désormais utilisés depuis trois cycles budgétaires, les premiers résultats semblent encourageants<sup>4</sup>. D'une part, les preuves apportées semblent plus nombreuses et plus solides, utilisant des méthodologies plus développées, comme celle des analyses coûts-bénéfices. Les demandeurs semblent à ce titre avoir procédé à un écrémage préalable de leurs propositions budgétaires pour ne présenter à l'OMB que celles dont les éléments de preuve sont les plus convaincants. D'autre part, le nouveau formulaire, qui cadre plus explicitement les attendus de l'OMB, semble avoir permis aux agences et départements d'exprimer plus clairement leurs demandes, alors même que le

nombre d'informations que requiert le formulaire est plus élevé qu'auparavant. Enfin, les preuves, plus nombreuses et détaillées, ont aidé le bureau du Gouverneur à discuter plus en profondeur les arbitrages budgétaires et les décisions politiques qui en résultent.

Toutefois, certains risques demeurent<sup>5</sup>. Dans la mesure où chaque agence est tributaire de l'évaluation des propositions budgétaires qu'elle soumet, l'OMB avait identifié un risque d'opportunisme et de surévaluation de leur qualité. Il a alors soumis à des évaluateurs indépendants l'ensemble des propositions budgétaires reçues, révélant qu'un tiers d'entre elles étaient effectivement trop optimistes par rapport aux résultats budgétaires attendus. Sont par ailleurs toujours en service des formulaires où l'échelle des preuves n'est pas utilisée pour renseigner sur la qualité des demandes budgétaires. Cela laisse penser qu'un effort de formation supplémentaire est nécessaire pour acclimater l'ensemble des agents à l'utilisation de ce nouvel outil. En dépit de ces premiers écueils, l'innovation fait des émules, comme en témoigne sa répllication récente dans cinq autres États américains : le Colorado, le Minnesota, la Caroline du Nord, le Mississippi et le Tennessee.

**Amicie de Tanoüarn** est apprentie « veille internationale et recherche en gestion publique comparée » au sein du bureau de la Recherche de l'IGPDE.

<sup>4</sup> Pour plus de détails sur ce point, se reporter au site de l'OPSI : <https://oecd-opsi.org/innovations/rhode-island-evidence-scale-for-budgeting/>.

<sup>5</sup> Pour plus de détails sur ce point, se reporter au site de l'OPSI : <https://oecd-opsi.org/innovations/rhode-island-evidence-scale-for-budgeting/>.



# L'Œil du chercheur

*L'Œil du chercheur* propose une présentation synthétique de thèses récemment soutenues dans le champ de la gestion publique ainsi que des résumés d'articles marquants publiés dans un bouquet de revues spécialisées en sciences de gestion. Des informations d'ordre général peuvent également figurer sur les publications et les événements de l'IGPDE.



## Revue des articles

## La Cour des comptes européenne : une championne sous-évaluée de l'évaluation ?

Antonin Thyraud

Article paru dans *Politiques & management public*, vol. 40, n° 1, janvier 2023, pp. 5-28, et accessible en ligne : [https://www.jle.com/fr/revues/pmp/e-docs/01\\_la\\_cour\\_des\\_comptes\\_europeenne\\_une\\_championne\\_sous\\_evaluee\\_de\\_levaluation\\_\\_346584/article.phtml](https://www.jle.com/fr/revues/pmp/e-docs/01_la_cour_des_comptes_europeenne_une_championne_sous_evaluee_de_levaluation__346584/article.phtml).



### Le thème

Cet article présente une socio-histoire de la Cour des Comptes européenne (CCE) depuis sa création en 1977. Orientée aujourd'hui vers l'audit de performance, la CCE a néanmoins, dès ses premières années d'existence, ancré son activité vers le volet évaluatif. En jouant sur son mandat officiel de garante de la « bonne gestion financière » et grâce à l'action de l'un de ses membres chargé des méthodes, André Middelhoek, la CCE se place dès ses origines comme un acteur incontournable de la progressive montée en puissance d'une culture de l'évaluation au sein des institutions européennes.

### Les données

Pour alimenter sa démarche socio-historique, l'auteur s'est appuyé sur deux types de sources : des archives et des entretiens semi-directifs. D'une part, les données collectées proviennent des fonds d'archives de la CCE, conservés aux Archives historiques de l'Union européenne à Florence, et, pour les documents plus récents, des ressources en ligne mises à disposition par l'institution. D'autre part, 15 entretiens semi-directifs ont été réalisés avec des membres de la CCE, de la Commission européenne, et des évaluateurs privés. La variété de ce matériel recueilli dans le cadre d'une thèse

(soutenue depuis à l'EHESS) a permis d'étudier les pratiques évaluatives de la CCE sous de larges aspects : travaux d'audit, formation professionnelle des auditeurs, relations de la Cour avec ses équivalents étrangers...

### Les résultats

L'article met en évidence une présence ancienne de l'évaluation au sein de la Cour, marquée par l'influence du déploiement dans des pays membres et dans les pays anglo-saxons de pratiques évaluatives en gestation dans des institutions de contrôle. Il revient notamment sur le rôle de l'un de ses membres, André Middelhoek, dont l'action au sein de la Cour participe à structurer les méthodologies d'évaluation issues de l'analyse économique coûts-avantages, et à penser la manière de les diffuser au sein des institutions européennes. Enfin, l'auteur montre de quelles manières ce positionnement évaluatif se solidifie au cours des années 1980 et 1990 par différents biais : production de « rapports spéciaux », action de « méta-évaluation » des politiques proposées par la Commission, partenariat avec d'autres institutions internationales de contrôle, déploiement de formations professionnelles pour les auditeurs de la CCE centrées sur les analyses coût-avantages.

## Revue des articles

# Évaluation des politiques publiques : vers une pratique intégrée au pilotage de l'action publique

Françoise Goter et Samia Khenniche

Article paru dans *Gestion et management public*, vol. 10, n° 3, 2022, pp. 35-56, et accessible en ligne : <https://doi.org/10.3917/gmp.103.0035>



## Le thème

Cet article étudie la possibilité d'intégrer l'évaluation dans le pilotage de l'action publique pour en améliorer la performance. Les auteures partent du constat que la notion de performance dans le cas de l'action publique est délicate à définir en raison de sa double dimension : interne (moyens, résultats) et externe (impacts, effets). L'évaluation, contrairement au contrôle de gestion, doit permettre de prendre en compte ces deux dimensions, ce qui explique que bien qu'elle soit fréquemment promue comme nécessaire, elle soit aussi peu diffusée et lorsqu'elle l'est, aussi peu mobilisée dans les débats.

## Les données

Cet article prend appui sur une recherche-intervention qui répondait à la demande d'accompagnement de la direction du développement territorial d'une mairie dans la perspective du renouvellement par les autorités locales de son contrat local de santé (CLS), et plus spécifiquement des ateliers santé ville (ASV). Le CLS est un dispositif porté conjointement par l'agence régionale de santé (ARS) et une collectivité territoriale dans le but de réduire les inégalités territoriales et sociales de santé en déployant une offre de santé de proximité. Les ASV visent plus spécifiquement les quartiers prioritaires déterminés dans le cadre de la politique de la ville, et cherchent à développer une approche transversale de la santé qui prenne en compte des enjeux sociaux, éducatifs et d'insertion professionnelle, à travers l'organisation d'actions et d'événements sur le territoire. Afin de répondre à la demande de la ville, les auteurs ont

mis en œuvre entre 2018 et 2020 un protocole de recherche proche d'une évaluation *in itinere*, devant permettre d'obtenir des preuves exploitables en cours d'action. Les ASV de deux quartiers prioritaires de la ville ont été étudiés grâce à 50 entretiens qualitatifs et 33 entretiens quantitatifs. Les premiers ont permis de faire ressortir des « phrases-témoins », c'est-à-dire des expressions-clés partagées par les acteurs et relatives aux impacts sociaux et économiques de l'ASV. À partir de ces phrases-témoins, les seconds ont permis de mesurer la valeur accordée à chaque impact par les acteurs et d'élaborer des indicateurs d'impact.

## Les résultats

L'évaluation *in itinere* réalisée a permis de croiser les informations de chaque partie prenante du dispositif et se concrétise, selon les auteurs, comme une solution pertinente pour intégrer l'évaluation au pilotage. Elle fait ressortir des jugements globalement positifs sur les ASV. Ceux-ci apparaissent notamment susceptibles de contribuer, au niveau social, à une meilleure connaissance entre partenaires, à une circulation de l'information plus rapide et congruente entre partenaires de santé et associatifs. Ils permettent également d'améliorer les compétences en ingénierie de projet en matière de santé et de mieux aligner les actions sur les besoins des usagers. Au niveau économique, l'évaluation met en évidence une meilleure mutualisation des ressources humaines et matérielles, des actions de prévention efficaces et des gains d'ingénierie. La recherche-intervention a aussi permis d'établir plusieurs « indicateurs de coût-performance cachés », c'est-à-dire non encore mesurés dans les systèmes de pilotage en place.

## Revue des articles

## La Garantie jeunes en pratique dans les missions locales : une expérimentation encadrée

Nicolas Farvaque et Carole Tuchsirer

Article paru dans *Travail et Emploi*, n° 153, janvier-mars 2018, et accessible en ligne : <http://journals.openedition.org/travailemloi/7831>.



### Le thème

La Garantie jeunes (GJ) est un dispositif d'accompagnement des jeunes visant à leur assurer une meilleure insertion sur le marché du travail. Prenant la forme d'une expérimentation encadrée, sa mise en œuvre a été déléguée aux missions locales, organismes de terrain créés en 1982 selon une vision large de l'insertion traitant simultanément des questions relatives à la famille, au logement, à la santé, à la culture, à la formation, à l'emploi et à l'accès à la citoyenneté. L'article questionne les décalages potentiels entre le cadre défini *a priori* par les concepteurs de la GJ et la réalité de sa mise en œuvre sur le terrain. Les auteurs s'interrogent à la fois sur la manière dont le modèle d'intermédiation des missions locales entre jeunes et employeurs imaginé dans le cadrage expérimental a pu évoluer lors de sa mise en pratique, et sur la façon dont ce modèle a fait évoluer les missions locales elles-mêmes.

### Les données

Les données de l'article sont issues d'une évaluation réalisée par la Dares entre novembre 2015 et avril 2016, à la demande du Comité scientifique de la Garantie jeunes. L'étude est scindée en deux axes : analyse de la mise en œuvre locale de la GJ d'une part, et évolution des pratiques d'accompagnement au sein des missions locales d'autre part. Les missions locales étudiées se situent dans

trois départements distincts, appartenant à trois types de territoires différents (rural, semi-urbain et métropolitain). Elles sont de tailles différentes, de 18 salariés à une centaine. Une soixantaine d'entretiens ont été réalisés et des observations d'ateliers ont été faites avec les parties prenantes de chaque mission locale (responsables des missions locales, employeurs, partenaires locaux...). Des entretiens ont été réalisés dans cinq autres missions locales de ces départements dans une perspective comparative.

### Les résultats

L'article montre que la GJ a été bien accueillie par les missions locales, même si l'aspect expérimental en a compliqué l'assimilation, et cela en dépit d'outils structurants de pilotage (guides, indicateurs...) qui équipaient le dispositif. Les organisations du travail des missions locales ont été fortement impactées par la nécessité d'assurer un accueil continu de cohortes au sein des locaux, d'intégrer de nouveaux conseillers ou d'assurer la coordination du dispositif. En centrant l'offre de services sur l'insertion des jeunes, la GJ a aussi permis d'améliorer la relation des missions locales avec ces derniers en développant leur suivi collectif. Néanmoins, l'article montre que la relation aux entreprises n'a pas été profondément bouleversée, alors même que l'un des objectifs de la GJ était de développer une médiation active entre missions locales et employeurs.

## Revue des articles

# Expérimentation, enquête, expérience : les politiques publiques à l'épreuve de la science

Mathias Béjean

Article paru dans les *Cahiers du GRM*, n° 16, 2020, et accessible en ligne : <http://journals.openedition.org/rgm/2251>.



## Le thème

Cet article part du constat du déploiement croissant d'un « gouvernement expérimental » depuis l'inscription dans la Constitution d'un droit à l'expérimentation en 2003. Rappelant l'inscription de ces méthodes expérimentales dans le domaine plus large des *evidence-based policies*, l'auteur étudie leurs éléments les plus structurants, en particulier les motivations, les croyances et les formes qu'elles peuvent prendre lorsqu'elles sont mobilisées par l'action publique. Face à certains impensés, l'auteur discute de l'utilisation de l'enquête comme type d'expérimentation pouvant permettre d'aller au-delà des limites de l'expérimentation.

## Les données

L'article se fonde sur une large revue de littérature académique et, dans une moindre mesure, sur certains rapports institutionnels qui questionnent ou promeuvent l'utilisation des expérimentations par les acteurs publics. L'auteur développe son

argumentation autour d'une discussion des principes qui fondent l'approche expérimentale à partir d'une référence aux auteurs canoniques en la matière.

## Les résultats

L'article met en évidence une limite importante de la démarche expérimentale résidant dans son incapacité à se déployer dans des situations relativement incertaines, inattendues ou complexes, ce qu'il nomme les *wicked-problems* de la science sociale. Pour dépasser cette limite, l'auteur s'inspire des travaux de John Dewey et de sa méthodologie par l'enquête présentée comme « un processus, temporalisé et localisé, au cours duquel une situation indéterminée est progressivement déterminée par un groupe social, la *communauté d'enquête* ». À ce titre, substituer l'enquête à l'expérimentation permettrait de passer d'une logique expérimentale à une logique expérimentielle, plus proche du terrain, comme l'illustre le déploiement de certains *living labs*.

## Revue des articles

# Gouverner par la communication comportementale en temps de crise sanitaire : l'argument comportementaliste comme nouveau paradigme de l'action publique en santé

Béa Arruabarrena

Article paru dans *Approches théoriques en Information-Communication (ATIC)*, n° 5, 2022/2, pp. 79-90, et accessible en ligne : <https://doi.org/10.3917/atic.005.0079>.



## Le thème

Cet article s'intéresse à l'usage des sciences comportementales dans les éléments de communication utilisés par le gouvernement français durant la pandémie de Covid-19. En posant comme constat que la communication comportementale, c'est-à-dire le recours à une communication privilégiant notamment les émotions pour influencer les comportements des individus et leur faire accepter certaines mesures, est devenue un paradigme dominant de l'argumentation politique en santé, l'auteure interroge ses effets sur les citoyens et leur participation à l'action publique.

## Les données

L'article développe dans un premier temps une approche qui questionne les théories sous-jacentes de la communication comportementale. Pour ce faire, il met en discussion un ensemble d'articles académiques et scientifiques sur les sciences comportementales. Cette première revue de littérature, dont l'auteure précise qu'elle sera approfondie dans le cadre d'un projet de recherche ultérieur porté par le Cnam/CNRS et recourant à une approche méthodologique quantitative (questionnaire) et qualitative (enquête de terrain par des ateliers citoyens), est complétée par une analyse du contexte sociopolitique de développement du *design* des politiques publiques en santé, notamment déployée à partir de travaux et de rapports institutionnels produits par la DITP, le Sénat ou des instituts de sondage.

## Les résultats

L'article aboutit à la conclusion que le recours à des dispositifs d'information et de communication comportementaux s'inscrit dans une gestion néo-libérale et biopolitique de l'action publique en santé. En mettant au cœur de la communication des éléments qui cherchent à agir sur les émotions des individus, notamment la peur, pour les persuader du bien-fondé des politiques conduites, ce type de communication cherche à obtenir des citoyens une forme de consentement en contournant leur raisonnement. Selon l'auteure, il pose donc des problèmes éthiques et ne laisse pas de place à la participation des individus dans la mesure où la communication ne se fonde pas sur la cohérence de discours raisonnés, obérant par là-même les possibilités de débat entre l'État et les citoyens. Enfin, l'article montre que la communication comportementale, en brouillant l'intention et le contenu des messages de communication, produit une communication dysfonctionnelle qui peut apparaître comme contradictoire aux citoyens. Selon l'auteure, l'exemple de l'attestation de sortie est emblématique, dans la mesure où ce document visait à responsabiliser les individus pour qu'ils aménagent leurs sorties, alors même que son *design* reposait sur un nombre important d'options possibles pour rendre cognitivement le choix plus difficile et dissuader les individus de sortir.

## Thèses et rapports institutionnels

# Guide de l'évaluation socioéconomique des investissements publics

France Stratégie

Rapport paru en septembre 2023 et accessible en ligne :

<https://www.strategie.gouv.fr/publications/guide-de-levaluation-socioeconomique-investissements-publics-edition-2023>

## Le thème

Ce guide débute par le constat que les investissements publics représentent aujourd'hui dans la plupart des pays développés une part importante de l'investissement. Souvent de long terme et engageant des ressources publiques, ces investissements impliquent que les acteurs publics évaluent leur valeur socio-économique. C'est d'ailleurs en ce sens que l'État français, par la loi de programmation pluriannuelle des finances publiques du 31 décembre 2012, a imposé une évaluation socioéconomique préalable pour tous les projets d'investissements civils financés par l'État, ses établissements publics, les établissements publics de santé ou les structures de coopération sanitaire. Ce guide doit permettre d'aider les services publics à réaliser les évaluations requises par la loi, en proposant des principes, concepts et méthodes opérationnelles qui peuvent être utilisés dans toutes les administrations concernées pour mettre en regard les gains de bien-être pour la collectivité avec ses coûts sociaux, économiques, financiers, sociétaux et environnementaux.

## Les données

Alors que France Stratégie avait publié un premier rapport méthodologique en 2013, l'instance s'associe début 2017 au Secrétariat général pour l'investissement (SGPI) afin d'instaurer un comité d'experts des méthodes d'évaluation socioéconomique des investissements publics. Ce comité, constitué des administrations concernées (Direction générale du Trésor, Commissariat général au développement durable, SGPI...) et

d'économistes formés aux techniques d'évaluation les plus récentes, est présidé par Roger Guesnerie, professeur d'économie émérite au Collège de France et président honoraire de la *Paris School of Economics*. Il publie en 2017 une première version de ce guide dont l'édition de 2023 est enrichie de différents compléments opérationnels. Au-delà des expertises mobilisées directement au sein du comité, ce guide s'appuie sur une revue de la littérature académique et institutionnelle.

## Les résultats

Ce guide se déploie en deux temps. Dans une première partie, après avoir présenté la démarche générale que suppose l'évaluation des investissements publics, il insiste plus spécifiquement sur l'importance d'établir un cadrage général de l'évaluation et différentes options de calcul, puis détaille de manière approfondie certaines bonnes pratiques liées au déroulé de l'évaluation. C'est notamment dans cette partie que les aspects méthodologiques sur l'objectivation des effets non marchands, la valorisation monétaire des effets marchands et non marchands, l'analyse coût-bénéfice, les risques et incertitudes et les effets non monétisables des investissements sont présentés. Le guide insiste aussi sur la nécessité de développer une analyse financière de l'investissement et de compléter l'ensemble par une évaluation *ex post*. Dans une seconde partie, différents compléments opérationnels sont détaillés, en lien par exemple avec le paramétrage macroéconomique des investissements ou celui des valeurs pour rendre compte des effets marchands et non marchands.

## Thèses et rapports institutionnels

## Les *Living Labs*, une perspective territoriale

Quentin Marron

Thèse de doctorat en Sciences du territoire sous la direction de Bernard Pecqueur soutenue le 31 mai 2022 à l'université Grenoble Alpes, et accessible en ligne : <https://theses.fr/2022GRALH008>

Cette thèse part du constat d'une crise des politiques publiques et, au-delà, du système actuel de démocratie représentative. Si le niveau local était pendant plusieurs décennies relativement protégé, selon l'auteur, il ne l'est plus car les citoyens ont le sentiment d'être mis à distance, mesurant mal les enjeux, les compétences et les acteurs en place. Surtout, ils estiment que les pouvoirs publics ne répondent que partiellement aux difficultés qu'ils rencontrent. En écho, émergent partout des initiatives novatrices dans lesquelles les parties prenantes cherchent à résoudre directement les problèmes qui leur incombent. Parmi les multiples entités existantes, la thèse prend les *Living Labs* comme objet. Initialement issus du milieu universitaire et du monde privé, ces laboratoires vont progressivement s'étendre aux politiques publiques. Des collectivités vont s'en saisir, voyant une opportunité intéressante pour constituer des formes d'action collective et impliquer les citoyens dans la gestion locale. En positionnant les usagers comme co-constructeurs, ces *Living Labs*, que l'auteur qualifie de « territoriaux », remettent en cause le principe d'une linéarité descendante et peuvent être considérés comme une réponse à la crise de l'action publique locale, en changeant les manières de la concevoir et de la conduire. La thèse s'attache à qualifier leur caractère éventuellement nouveau, différencié et transformateur, en analysant la réalité de

l'ancrage géographique local et la prise en compte de spécificités territoriales qu'ils affichent. Face à un objet de recherche relativement nouveau et mouvant, cette recherche déploie un bagage théorique composite recourant à d'autres disciplines telles que la sociologie et la science politique. Dans un premier temps, elle établit une comparaison nationale avant de se concentrer, dans un deuxième temps, sur trois dispositifs : le Lab'01, Brie'Nov et Autonom'lab. Le protocole méthodologique repose sur l'analyse de documents et la réalisation d'entretiens. La thèse montre que la réponse apportée par les *Living Labs* locaux à la crise de l'action publique locale est relative, tout comme leur caractère transformateur. Leur capacité à changer les manières de faire reste marginale. Étant utilisés de manière intermittente, ils ne donnent pas lieu à une densification organisationnelle. Le manque de durabilité des initiatives limite la capacité à se diffuser et à créer des effets de contagion. En outre, ils sont dans l'incapacité de changer d'échelle, en dépassant le niveau local, pour agir de manière structurelle. Ce sont des dispositifs encore « discrets », à tel point que l'on peut parler de « transformations silencieuses » et de « changements graduels transformateurs », notamment en matière de démocratie locale et d'imprégnation de nouveaux principes (expérimentation, droit à l'échec, culture de la co-construction) auprès des institutions publiques.

## Thèses et rapports institutionnels

# Analyse du processus d'institutionnalisation d'une expérimentation pour l'emploi : le cas de « Territoires zéro chômeur de longue durée » à Jouques

Camille Retsin

Thèse de doctorat en Sciences économiques sous la direction de Fransesca Petrella soutenue le 7 juillet 2022 à l'université d'Aix-Marseille, et accessible en ligne : <https://theses.fr/2022AIXM0253>

L'objectif de cette thèse est de comprendre le processus d'institutionnalisation d'une dynamique de coopération territoriale d'acteurs autour de l'expérimentation Territoires zéro chômeur de longue durée (TZCLD). Déclinée à trois niveaux (national, territorial et organisationnel) autour d'une entreprise à but d'emploi (EBE), cette expérimentation a pour objectif de mettre en œuvre le droit à l'emploi inscrit dans la Constitution de 1946 en créant de l'activité non-concurrentielle au sein d'une EBE, en CDI, à temps choisi et rémunéré sur la base d'un Smic sans opérer de sélection. Cette thèse s'ancre dans une démarche qualitative d'étude de cas unique composé de sous-unités interreliées et se place dans une posture d'abduction permettant une construction

progressive du cadre théorique. Deux hypothèses sont formulées : la première est que l'expérimentation représente une innovation sociale dans le champ des politiques publiques de l'emploi ; la seconde est qu'elle a vocation, en cas de réussite du modèle, à s'institutionnaliser. Ce travail repose sur la construction d'un cadre théorique composé de trois approches permettant de comprendre les ressorts de la coopération entre une diversité d'acteurs sur un territoire : celle de la gouvernance, des proximités et des logiques institutionnelles. Au travers d'une analyse processuelle de la dynamique de coopération territoriale des acteurs, il fait émerger de chaque approche un ensemble de facteurs participant à l'institutionnalisation de l'expérimentation.

## Thèses et rapports institutionnels

## La preuve de concept comme outil de développement des capacités de générativité collective : modélisation, expérimentation et conditions de performance

Caroline Jobin

Thèse de doctorat en Sciences de gestion sous la direction de Pascal Le Masson et de Sophie Hooge, soutenue le 24 mars 2022 à l'université Paris Sciences et Lettres (PSL), et accessible en ligne : <https://theses.fr/2022UPSLM012>

Conçue dans les années 1960 à la Nasa dans le contexte du programme Apollo, la preuve de concept – ou *proof of concept* (POC) en anglais – a connu un succès grandissant au fil des décennies auprès des praticiens de l'innovation. Davantage encore qu'à l'époque, le régime d'innovation contemporain requiert d'explorer collectivement l'inconnu. Les processus de conception se sont complexifiés : l'identité des objets et des collectifs est rarement prédéterminée. Ils doivent être conçus collectivement et simultanément. Dans ce contexte, la thèse étudie comment la preuve de concept peut devenir un outil de développement des capacités de générativité collective. Tout d'abord, la recherche confirme par une étude généalogique que le POC était et est encore le symptôme d'un régime d'innovation se devant de fédérer et faire travailler ensemble des collectifs d'exploration. Ces collectifs comprennent des acteurs aux expertises et légitimités hétérogènes, souvent issus d'institutions variées. Le POC est un des rares outils de gestion permettant d'organiser l'exploration collective de l'inconnu. Il est particulièrement adapté à ces situations parce qu'il porte une double logique de preuve : une « preuve de connu », processus expérimental permettant de générer de nouvelles connaissances et de les valider (c'est-à-dire de les reconnaître collectivement comme vraies ou fausses) et une « preuve

d'inconnu », processus expérimental permettant de générer de nouveaux concepts et de guider leur exploration (c'est-à-dire de nommer collectivement un inconnu interprétable). Ensuite, grâce à l'étude rétrospective de preuves de concept conçues et réalisées par l'agence de *design* les Sismo, la thèse identifie qu'il existe différents arrangements entre preuve de connu et preuve d'inconnu et qu'une cohabitation est possible et bénéfique. De manière contre-intuitive, elle montre que plus la preuve de connu est recherchée et de qualité, plus elle permet de faire émerger de manière simultanée la preuve d'inconnu, et vice-versa. La question des compétences de l'expérimentateur ressort comme l'élément critique pour permettre cette cohabitation, et avoir un POC à fort pouvoir génératif. Enfin, la thèse propose une étude de cas longitudinale d'une preuve de concept conduite par la Chaire de philosophie à l'hôpital et les Sismo, qui portait explicitement une ambition de plus forte générativité. Ces travaux confirment l'hypothèse que le POC doit être étudié, et donc évalué, non plus seulement comme un outil de test d'un concept mais, également ou dorénavant, comme un outil de développement de capacités de conception collective. La recherche identifie des conditions de gouvernance critiques pour que ce changement de paradigme puisse s'opérer au profit de la société.

Cette rubrique a été préparée par **Amicie de Tanouärn**, apprentie « veille internationale et recherche en gestion publique comparée » au sein du bureau de la recherche de l'IGPDE, et **Edoardo Ferlazzo**, chef du département Gestion publique comparée.



SOMMAIRE DU NUMÉRO PRÉCÉDENT

## **N° 20 (2024/1) – PATRIMOINE IMMOBILIER PUBLIC ET TRANSITION ÉCOLOGIQUE**

### **Éditorial**

MARIE NIEDERGANG

### **[Regards croisés] Patrimoine immobilier public : quels enjeux notamment dans le contexte de la transition écologique ?**

ENTRETIEN ENTRE YOANN QUEYROI ET ALAIN RESPLANDY-BERNARD

### **[Étude] La gestion du patrimoine immobilier public local face aux enjeux écologiques : quelles modalités de pilotage et de gouvernance ?**

SÉBASTIEN DONY ET YOANN QUEYROI

### **[Étude] La gestion de l'immobilier public face aux enjeux écologiques : étude comparative des stratégies déployées aux niveaux étatique et local**

FATEM-ZAHRA EL FASSI, AVEC LA PARTICIPATION DE BRUNO FILLIARD ET ALEXANDRE MOREAU

### **[Note réactive] Helsinki 3D : un outil dynamique et collaboratif pour la gestion du patrimoine bâti**

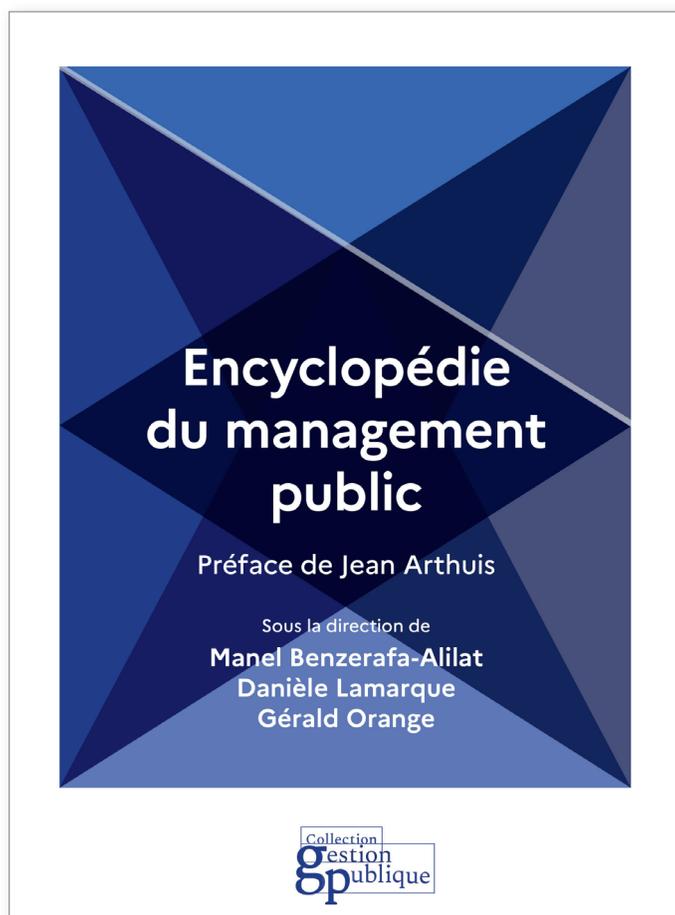
PASCAL FRUCQUET

### **[L'Œil du chercheur] Revue d'articles et de thèses**

Disponible en librairie

INSTITUT DE LA GESTION PUBLIQUE ET DU DÉVELOPPEMENT ÉCONOMIQUE

## Gestion publique



ISSBN : 978-2-11-162105-3

752 pages

28 €

<https://www.economie.gouv.fr/igpde-editions-publications>

Pour se procurer l'ouvrage auprès de l'éditeur :  
[recherche.igpde@finances.gouv.fr](mailto:recherche.igpde@finances.gouv.fr)

Édition numérique accessible sur  
<https://books.openedition.org/igpde/15291>

 **OpenEdition**  
Books