

# CYCLE DE SÉMINAIRES

## L'impact des politiques publiques

### COMPTE RENDU DE LA SÉANCE 1

#### Comment évaluer ?

Vendredi 17 novembre 2017

COMPTE RENDU

L'évaluation des politiques publiques s'est développée au cours des dernières années grâce notamment à un meilleur accès aux données confidentielles et à l'émergence de nouveaux acteurs. Selon **Fabrice LENGART** (commissaire général adjoint, France Stratégie), il reste cependant des progrès à faire pour accroître le nombre des évaluations d'impact des politiques publiques<sup>1</sup>, pour en faire comprendre les méthodes et les résultats et pour faire en sorte qu'elles influent sur les décisions et les pratiques.

L'idée du séminaire est de réunir chercheurs, praticiens et décideurs afin de croiser les points de vue. Le cycle de séminaires comprendra six séances, organisées chacune autour de deux volets : un volet thématique, en l'occurrence l'éducation aujourd'hui, et un volet transversal, à travers la démarche et les méthodes évaluatives lors de cette première séance.

#### SESSION 1 – FAUT-IL MOINS D'ÉLÈVES PAR CLASSE ?

Selon **Marc GURGAND** (directeur de recherche au CNRS), l'exemple de la réduction de la taille des classes est intéressant à plusieurs égards : il s'agit d'une politique bien définie ; l'évaluation de ses effets mobilise diverses méthodes qui permettent d'illustrer leur intérêt et leurs limites ; la littérature académique, française et étrangère, est relativement importante et convergente<sup>2</sup>.

1. Les évaluations d'impact cherchent à mesurer les effets des politiques publiques en essayant de reconstituer la situation contrefactuelle sans la politique.
2. Voir la note de l'Institut des politiques publiques (IPP) et les slides de la présentation de Marc GURGAND en ligne sur le site de France Stratégie : [www.strategie.gouv.fr/debats/seance-1-cycle-evaluation-dimpact-politiques-publiques-evaluer](http://www.strategie.gouv.fr/debats/seance-1-cycle-evaluation-dimpact-politiques-publiques-evaluer).

Fournir une réponse fiable à la question « la réduction de la taille des classes permet-elle d'améliorer la réussite scolaire des élèves ? » est cependant compliqué car les élèves les plus faibles sont souvent regroupés dans les classes plus petites. Plusieurs méthodes ont été utilisées pour corriger ces « biais de sélection ».

Les premières méthodes mises en œuvre consistaient à redresser les résultats des élèves par les différences de résultats à des tests initiaux, réalisés en début d'année (avant la mesure). Elles ne trouvaient que des effets faibles de la réduction de la taille des classes. Ces méthodes étaient cependant imparfaites car les tests initiaux ne prédisent que très partiellement les tests finaux : il y a donc toujours l'espace pour des biais de sélection substantiels.

De nouvelles méthodes plus robustes ont ensuite été utilisées : l'expérimentation aléatoire contrôlée caractérisée par un tirage au sort à la fois des élèves et des enseignants<sup>3</sup> ; les quasi-expériences qui exploitent des variations dans la taille des classes indépendantes du niveau initial des élèves, comme par exemple les seuils de déclenchement d'ouverture des classes.

La mobilisation de ces méthodes plus robustes permet de mettre en évidence des effets importants de la réduction de la taille des classes dans l'enseignement primaire :

- sur un ensemble de neuf évaluations réalisées dans cinq pays différents, sept évaluations trouvent un effet significatif ;
- passer de vingt-quatre à douze élèves par classe équivaldrait à réduire de moitié l'écart moyen de réussite à l'entrée en CP entre les enfants favorisés et défavorisés en France ;
- quelques études indiquent que cette mesure aurait également des effets positifs à terme sur l'insertion professionnelle (taux d'emploi, salaires), en Suède notamment.

Pour l'enseignement secondaire, les résultats sont moins probants.

La réduction de la taille des classes est une politique efficace en soi, qu'elle s'accompagne ou non d'un ajustement des pratiques pédagogiques.

Un calcul rapide de type coûts-bénéfices indique que pour que cette mesure d'un coût de 700 M€/an en REP/REP+ en CP/CE1 soit rentable<sup>4</sup>, il suffit que le salaire futur des bénéficiaires soit de 1 % supérieur à ce qu'ils percevraient sans la mesure.

**Frédérique ALEXANDRE-BAILLY** (rectrice de l'académie de Dijon) rappelle que la réduction de la taille des classes est effective en CP en REP+ depuis septembre 2017<sup>5</sup> et qu'elle sera évaluée par la Direction de l'évaluation, de la prospective et de la performance (DEPP) du ministère de l'Éducation nationale (MEN). Elle a été associée au recrutement de maîtres expérimentés qui doivent bénéficier d'une prime supplémentaire de 3 000 euros et à une politique de formation et d'accompagnement des professionnels (séminaire national, séminaire inter-académique en présence du ministre). **Frédérique ALEXANDRE-BAILLY** s'interroge sur ce que pourra montrer l'évaluation du dispositif dans la mesure où il s'agit d'une politique mixte combinant à la fois la réduction de la taille des classes et une formation importante des enseignants et qui implique d'ores et déjà beaucoup de réflexions pédagogiques au sein des circonscriptions.

3. Il n'existe qu'un seul exemple d'expérience aléatoire contrôlée sur ce sujet : il s'agit du programme STAR mis en œuvre aux États-Unis en 1986.

4. Les réseaux d'éducation prioritaire (REP).

5. Les modalités de mise en œuvre peuvent varier : il peut s'agir de classes à douze élèves avec un enseignant (dans 85 % des cas) ou de classes à 24 élèves avec deux enseignants (dans 15 % des cas).



L'extension de la mesure aux REP l'an prochain se fera dans des conditions différentes, avec des moyens notamment plus faibles qu'en REP+.

Dans l'académie de Dijon, qui ne compte qu'un seul REP+ (passage de 5 classes de CP à 12 classes dans 4 écoles), la mise en œuvre des CP dédoublés a suscité des discussions avec les maires et les enseignants et des interrogations notamment sur la pérennité de la mesure (sur l'opportunité de construire de nouveaux locaux si la politique doit être par la suite arrêtée) et sur son articulation avec le dispositif « Plus de maîtres que de classes » (dont certains seront placés en CP dédoublés). Son déploiement s'est accompagné de formations pour les enseignants des classes de CP dédoublés (notamment pour qu'ils utilisent davantage les manuels scolaires et sachent mieux les discriminer selon leur qualité). Certains temps de regroupement des élèves ont aussi été prévus dans certaines disciplines (en sport ou en art) pour développer les interactions sociales entre les élèves et la dimension collective. Les enseignants ainsi libérés peuvent venir en appui de leurs collègues pendant ces temps (recréant de ce fait le dispositif des maîtres plus).

Selon **Frédérique ALEXANDRE-BAILLY**, l'évaluation de la politique des CP dédoublés soulève plusieurs défis, notamment ceux d'apprécier « la prise en main managériale » (c'est-à-dire la façon dont les acteurs s'en sont emparés), de prendre en compte les externalités négatives – liées à une baisse des moyens dans les classes de niveau supérieur (à partir du CE2) – ainsi que les coûts évités grâce à la politique (coût de remédiation et coût médical).

**Olivier COSNEFROY** (DEPP, MEN) précise le dispositif d'évaluation prévu par la DEPP en association avec l'École d'économie de Paris. Le dispositif comprendra trois volets : le premier sur les performances scolaires des élèves (à partir de trois échantillons d'élèves en REP+, en REP et hors REP) ; le deuxième sur les pratiques des enseignants et la façon dont ils se sont emparés du dispositif ; le troisième sur les relations dans les classes dédoublées entre les élèves et entre élèves et enseignants.

## SESSION 2 – L'ÉVALUATION DES POLITIQUES PUBLIQUES, QUESTIONS DE MÉTHODES

Dans son introduction liminaire, **Pauline GIVORD** (Insee)<sup>6</sup> rappelle la définition de l'évaluation des politiques publiques telle qu'elle figure dans le décret de 1998<sup>7</sup> : « L'évaluation d'une politique publique [...] a pour objet d'apprécier [...] l'efficacité de cette politique en comparant ses résultats aux objectifs assignés et aux moyens mis en œuvre ». L'évaluation ne se donc limite pas au pilotage des politiques (qui consiste à compter le nombre de bénéficiaires et à chiffrer le coût), mais cherche à répondre aux questions suivantes :

- la politique a-t-elle répondu aux objectifs fixés ? (critère d'efficacité) ;
- a-t-elle un impact significatif sur ses bénéficiaires ? Par exemple, les dispositifs d'éducation prioritaire permettent-ils de réduire l'échec scolaire ?

De simples comparaisons entre bénéficiaires et non-bénéficiaires ne permettent pas en général de mesurer l'efficacité d'un dispositif en raison des effets de sélection (liés aux différences entre les populations) et des effets de contexte.

6. Voir les slides de la présentation de Pauline Givord sur le site de France Stratégie, [www.strategie.gouv.fr/debats/seance-1-cycle-evaluation-dimpact-politiques-publiques-evaluer](http://www.strategie.gouv.fr/debats/seance-1-cycle-evaluation-dimpact-politiques-publiques-evaluer).

7. Décret n° 98-1048 du 18 novembre 1998 relatif à l'évaluation des politiques publiques.



L'évaluation d'impact des politiques publiques utilise la notion de contrefactuel, qui sert de point de référence et mesure la situation « hypothétique » qui aurait été observée sans la mise en place du dispositif.

Quatre catégories de méthodes quantitatives permettent d'estimer ce contrefactuel dans le domaine de l'éducation et plus généralement :

- les expérimentations aléatoires (méthode de référence) : elles consistent en un test à échelle réduite à partir d'une sélection au hasard des bénéficiaires et des non-bénéficiaires du dispositif. Leurs résultats peuvent alors être directement comparés au cours du temps. Cette méthode est cependant parfois difficile à mettre en œuvre (coût élevé, considérations éthiques) ;
- les régressions sur discontinuités : elles comparent des bénéficiaires et des non-bénéficiaires de part et d'autre des « seuils d'éligibilité » aux dispositifs. Cette méthode ne mesure que l'effet du dispositif à la marge (sur une petite partie des bénéficiaires), et ne permet donc pas de généraliser les résultats à l'ensemble des bénéficiaires ;
- le matching : il compare les bénéficiaires avec des non-bénéficiaires quasi identiques au regard de certaines caractéristiques observables (origine sociale, établissement, sexe, etc.). Cette méthode réduit une part du biais de sélection, mais ne tient pas compte des caractéristiques inobservables (la motivation par exemple) ;
- la double différence : elle compare l'évolution de la situation entre les bénéficiaires et les non-bénéficiaires, avant et après la mise en place de la politique. Elle suppose que l'évolution de chacun des deux groupes aurait été similaire sans la politique, ce qui n'est que rarement le cas.

Une bonne démarche d'évaluation d'impact requiert de :

- expliciter les objectifs : augmenter le niveau scolaire ? Réduire la part de décrocheurs ? Améliorer le climat scolaire ? Former des citoyens éclairés ? Les objectifs ne sont cependant pas toujours clairement énoncés dans la loi ;
- décliner ces objectifs en indicateurs quantitatifs lorsque cela est possible ;
- déterminer la méthode empirique à utiliser selon la nature du dispositif et les moyens disponibles ;
- recueillir les données : cette étape doit être anticipée. On ne peut pas se contenter de suivre uniquement les bénéficiaires ;
- exploiter et interpréter les résultats, et assurer leur validation par les pairs.

Enfin, une évaluation d'impact nécessite du temps dont ne dispose pas forcément le politique.

Les évaluations d'impact n'apportent pas de réponse définitive :

- ce sont des réponses statistiques. Il n'est pas toujours possible de statuer sur l'effet significatif de la politique, lorsque les intervalles de confiance sont larges ;
- les réponses concernent une question posée, un contexte et un public. Elles n'apportent donc pas de réponse globale. Il est donc nécessaire de s'appuyer sur la connaissance accumulée.

Les réponses des évaluations d'impact permettent toutefois de nourrir le débat public et d'éclairer et d'objectiver les choix politiques qui ont été faits.



La table ronde a ensuite réuni trois intervenants, **André BARBÉ**, **Jérôme TEILLARD** et **Laura LITVINE**.

**André BARBÉ** (conseiller maître à la Cour des comptes) porte un regard critique sur la place de l'évaluation dans la sphère publique, qu'il s'agisse des études d'impact des projets de loi – obligatoires mais de mauvaise qualité ou inexistantes comme par exemple sur la politique d'accueil des grands événements sportifs ; des évaluations ex post – rarement prévues à l'exception de la loi sur la refondation de l'école qui a mis en place un Comité de suivi chargé d'évaluer l'atteinte des objectifs ; des expérimentations – pour lesquelles le principe d'une différenciation de traitement des élèves pendant plusieurs années est difficilement accepté et la généralisation bien souvent décidée avant la disponibilité des résultats des évaluations (comme pour les internats d'excellence).

**André BARBÉ** regrette également l'absence en France d'un émetteur de l'évaluation suffisamment puissant (à l'instar de l'OCDE pour PISA). La puissance de l'émetteur est à la fois nécessaire et indispensable, et dépend de son indépendance. Sa pérennité est également essentielle pour élaborer une doctrine et promouvoir une culture de l'évaluation (comme la Haute Autorité de santé a pu le faire). Deux autres dimensions font obstacle à l'évaluation en France : le manque de clarté des objectifs associés aux lois et des indicateurs mal renseignés<sup>8</sup>.

La Cour des comptes pour sa part a reçu la mission constitutionnelle d'évaluer les politiques publiques. Elle peut l'exercer de manière indépendante en choisissant ses thèmes, ou être saisie par le gouvernement et le Parlement. La Cour se trouve dans une situation complexe car elle doit à la fois construire une culture de l'évaluation en s'appuyant sur la recherche (l'évaluation n'étant pas son cœur de métier) et concilier ses missions de contrôle et d'évaluation, qui peuvent parfois entrer en conflit.

L'évaluation à la Cour des comptes repose à la fois sur : le maintien de l'activité de contrôle ; la prise en compte des parties prenantes rassemblées dans des comités d'accompagnement (experts, chercheurs, représentants de l'administration) ; la combinaison d'évaluations qualitatives (outils habituellement utilisés par la Cour) et quantitatives (techniques nouvelles pour la Cour).

**Jérôme TEILLARD** (IGEN) relate son expérience passée en tant que promoteur de politiques publiques. Les évaluations d'impact se sont développées en France il y a environ dix ans. Le Fonds d'expérimentation pour la jeunesse (FEJ) a permis d'accélérer le changement et de franchir une étape. Le principe d'égalité de traitement des élèves est très marqué en France et sa rupture est mal acceptée par les parents. La justification de l'expérimentation a été faite avec la promesse d'un dispositif rodé d'évaluation d'impact. L'objectif était de démontrer rapidement l'efficacité ou non d'un dispositif de politique publique.

La « mallette des parents » est un exemple d'expérimentation réussie où il a été possible d'expérimenter et d'évaluer avant de généraliser (en classe de sixième d'abord puis en classe de troisième). Pour d'autres dispositifs, comme les internats d'excellence, cette continuité n'a pas pu exister.

Si les expérimentations et les évaluations d'impact sont importantes, elles comportent aussi des limites liées à l'acceptabilité des expérimentations<sup>9</sup> ; au passage à l'échelle quand le dispositif est étendu ; au temps nécessaire à leur réalisation. Néanmoins, selon **Jérôme TEILLARD**, on assiste à une prise de conscience croissante de l'importance d'un progrès basé sur la preuve scientifique. Cela pose la question de la relation avec la recherche en éducation et des processus d'alliance à imaginer.

8. Par exemple, si on regarde plus spécifiquement la Mission interministérielle enseignement scolaire (MIES), il en ressort que les objectifs ne sont pas stables au cours du temps et que plusieurs indicateurs ne sont pas renseignés.

9. Par exemple, l'expérimentation sur le décrochage scolaire a été critiquée au motif qu'elle rémunérait l'absentéisme.



**Laura LITVINE** (experte au Behavioural Insights Team, BIT, au Royaume-Uni) indique que le BIT a été créé en 2010 au cabinet de David Cameron à la fois pour appliquer les sciences comportementales (les « nudges » ou coups de pouce) aux politiques publiques et évaluer l'impact de dispositifs publics (ou de leurs modalités). Le BIT s'est séparé du gouvernement en 2014, mais ce dernier reste toutefois actionnaire principal. Le BIT travaille avec de nombreux ministères ou agences publiques au Royaume-Uni (par le biais d'un contrat cadre avec le gouvernement qui peut faire appel au BIT lorsqu'il le souhaite), et dans de nombreux pays (une trentaine).

Le BIT essaie de maintenir une rigueur importante dans ses évaluations tout en essayant de réduire le temps de leur réalisation et leur coût (par exemple en utilisant des données existantes ou en s'inscrivant dans des structures existantes). Aujourd'hui, plus de trois cent soixante évaluations randomisées ont été conduites pour le compte du gouvernement anglais, et une école sur trois environ a participé à une évaluation randomisée (dans les groupes tests ou de contrôle). Les What Work Centres (structures à financement public) jouent un rôle important dans la production de méta-analyses et la mise à disposition d'outils de diffusion et de vulgarisation de la production scientifique à destination des enseignants<sup>10</sup>.



10. Voir les toolkits du WWC Education Endowment Foundation (EEF) disponibles à l'adresse suivante : <https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/>.

