

Le Fonds d'expérimentation pour la jeunesse et les politiques éducatives : premier retour d'expérience

Marc Gurgand,

École d'économie de Paris, CNRS

Mathieu Valdenaire,

Fonds d'expérimentation pour la jeunesse, DJEPVA, MENJVA

Le Fonds d'expérimentation pour la jeunesse (FEJ) soutient de nombreuses innovations de politique scolaire en les accompagnant d'une évaluation réalisée par une équipe indépendante. Si cette démarche d'expérimentation et d'évaluation n'est pas nouvelle dans l'Éducation nationale, le FEJ a porté depuis sa création une démarche ambitieuse au plan méthodologique. Parmi la diversité des méthodes mobilisées, l'accent mis sur les évaluations d'impact recourant à la comparaison de groupes-tests et de groupes-témoins déterminés par tirage au sort constitue probablement la principale originalité de ce programme. Les deux premières années d'existence du FEJ et le suivi des quelques 500 expérimentations lancées à ce jour ont permis de préciser les conditions de validité des démarches d'évaluation et tirer les enseignements de la mise en œuvre des différentes expérimentations. L'enjeu central de la période qui s'ouvre désormais est celui de la capitalisation des résultats issus des expérimentations, afin d'inscrire leurs enseignements dans la conception des politiques publiques.

Le Fonds d'expérimentation pour la jeunesse (FEJ) est une innovation dans le paysage institutionnel français. Créé par la loi du 1^{er} décembre 2008, il vise à soutenir, sur des contributions de l'État et de partenaires privés¹, des dispositifs expérimentaux en faveur des jeunes. Il ciblait initialement les jeunes âgés de 16 à 25 ans et abordait principalement les questions d'insertion sociale et professionnelle. Depuis 2010, et son intégration au sein du ministère de l'Éducation nationale, de la jeunesse et de la vie associative, le champ d'action du FEJ est étendu à tous les jeunes âgés de moins de 25 ans, partant, à toutes les dimensions des politiques scolaires.

Les expérimentations consistent à déployer à échelle réduite une action ou un dispositif que l'on souhaite tester avant d'envisager une généralisation ou, plus modestement, un essai-image ou une inflexion. L'évaluation est donc consubstantielle à cette démarche puisqu'elle vise à tirer les enseignements de l'expérimentation. Sur ce plan, le FEJ est particulièrement exigeant : avec le soutien d'un conseil scientifique, il s'assure que les évaluations seront rigoureuses et

apporteront des diagnostics raisonnés et fondés empiriquement.

Le FEJ adopte simultanément deux approches. D'une part, il suscite, à travers des appels à projets, l'initiative d'une multitude d'acteurs (allant des associations aux établissements publics, en passant par les communes ou les conseils généraux), qui doivent proposer des programmes originaux sur un thème d'intervention donné. En ceci, le Fonds encourage l'innovation et mobilise les acteurs autour de dispositifs qui, sans son soutien et les éléments d'appréciation apportés par l'évaluation, resteraient isolés et méconnus. Par exemple, de multiples initiatives de lutte contre le décrochage scolaire sont portées par des associations, des établissements, des villes, et plusieurs d'entre elles mériteraient d'être mieux connues et reprises par d'autres acteurs, ou encouragées par l'État ou les collectivités locales. Le FEJ a ainsi sélectionné plus de 40 initiatives sur ce seul thème, à travers des appels à projets. Au total, le FEJ soutient aujourd'hui près de 400 projets,

NOTE

1. Principalement la Fondation Total.

dont une partie importante dans le domaine scolaire.

D'autre part, le FEJ accompagne, financièrement et techniquement, la mise en place expérimentale de dispositifs conçus par les services de l'État qu'il est nécessaire de tester, soit parce que leur efficacité est difficile à établir *a priori*, soit parce que les conditions de leur déploiement sont complexes et doivent être travaillées à petite échelle. Dans ce cas, les porteurs de projet (qui sont typiquement des établissements scolaires) répondent à des appels à candidature et les évaluateurs sont recrutés par appel d'offres sur des cahiers des charges précis. C'est ainsi que l'on expérimente, par exemple, l'organisation du temps scolaire (« Sport l'après-midi »), le « livret de compétences expérimental » (qui vise à formaliser et valoriser tous les acquis des élèves, y compris ceux qui ne sont pas de nature scolaire au sens traditionnel du terme) ou encore des techniques d'apprentissage de la lecture susceptibles de mieux aider les élèves en difficulté.

Dans les deux cas, on peut dire que le FEJ contribue à renouveler les modalités de réflexion sur les politiques scolaires, soit en donnant une place nouvelle aux initiatives du terrain, soit en soumettant à un examen rigoureux et public des projets d'ambition nationale.

Cet article présente la démarche du FEJ, telle qu'elle s'est construite depuis environ trois ans, les difficultés concrètes qu'il a eu à surmonter et les défis qu'il lui reste à relever. On espère surtout montrer qu'à partir d'un programme ambitieux, il est possible d'aller assez loin dans le renouvellement de l'expérimentation et de l'évaluation.

LE FEJ DANS L'ÉDUCATION NATIONALE

Certes, l'expérimentation et l'évaluation ne sont pas des démarches nouvelles dans l'Éducation nationale. Il est habituel de rappeler que l'évaluation *des élèves* est une pratique quotidienne des enseignants, et qu'elle est devenue une pratique institutionnelle avec les évaluations nationales. Mais cette évaluation-là ne nourrit pas, ou très ponctuellement, l'évaluation *des politiques*. Pour que les évaluations nationales soient aussi la matière d'évaluations de politiques scolaires, il faut que certaines écoles soient dans un dispositif donné, d'autres en dehors, et que les unes et les autres soient par ailleurs assez comparables pour que l'on puisse interpréter des écarts de résultats aux évaluations nationales comme les effets de cette intervention. En fait, dans beaucoup de cas, l'évaluation des politiques scolaires en dehors d'une situation expérimentale est un exercice difficile.

L'expérimentation n'est pas non plus nouvelle. On peut même envisager le monde scolaire comme un univers très évolutif, constamment en recherche de solutions et qui expérimente, implicitement ou explicitement, presque en continu. La politique des ZEP en donne un exemple historique : comme l'a illustré le rapport Moisan-Simon [3], cette politique contenait, au-delà des moyens, un encouragement à l'innovation pédagogique et aux interactions plus fortes entre l'école et les structures locales, qui ont pris corps sur le terrain de façons extrêmement variables. Cette diversité des actions résultait de la volonté affichée de mobiliser les acteurs de terrain et

d'encourager leur initiative. Aussi, les ZEP sont apparues d'abord comme des terrains d'expérimentation et certains considèrent que les innovations ainsi produites ont bientôt été incorporées dans la *doxa* pédagogique et ont contribué à faire évoluer l'ensemble du système éducatif [1].

Cependant, une telle approche de l'expérimentation, imparfaitement formalisée, est privée d'évaluations conçues et mises en place d'emblée, qui permettraient d'en tirer les enseignements de façon explicite, transparente et ouverte au débat. La direction de l'évaluation, de la prospective et de la performance (DEPP) a une tradition d'expérimentation à des fins d'évaluation mais, d'une part, elle repose sur une commande politique qui est longtemps restée ponctuelle et, d'autre part, elle doit s'inscrire à l'intérieur d'un programme de travail dans lequel prime la production statistique et la mesure de la performance notamment la mesure des acquis des élèves. En outre, cette partie de son activité tend à rester peu visible, comme l'expérimentation de la réduction de la taille des classes au cours préparatoire menée en 2002-2003, opération méconnue et pourtant d'une ambition et d'une ampleur exceptionnelle. Aujourd'hui, le FEJ, à travers les financements qu'il apporte et la volonté politique qu'il traduit, permet de renforcer cette activité, et la DEPP mène par exemple l'évaluation du « livret de compétences expérimental ». Elle contribue également de façon décisive à l'évaluation des internats d'excellence, du dispositif *Coup de Pouce Clé*² ou

NOTE

2. Dispositif de prévention des décrochages précoces en lecture-écriture créé par l'association pour favoriser l'égalité des chances à l'école (Apféé).

encore de l'apprentissage de la lecture, toutes opérations soutenues par le FEJ. L'articulation de ces deux composantes du ministère en charge de l'Éducation nationale (DEPP et FEJ), qui peuvent apporter chacune leurs ressources et leur culture, est l'occasion de renforcer l'évaluation des politiques scolaires.

Autre signe d'une évolution de fond, l'expérimentation est de plus en plus formalisée comme telle, et fait l'objet d'une attention croissante au cœur même de l'institution. La loi d'orientation et de programme pour l'avenir de l'École du 23 avril 2005 mentionne ainsi dans son article 34 que les projets d'école ou d'établissement peuvent prévoir la réalisation d'expérimentations portant sur un vaste ensemble de domaines³. C'est également ce que reflète la création, au sein de la direction générale de l'enseignement scolaire (DGESCO) d'un département « recherche, développement, innovation et expérimentation » (DRDIE), et le fait que cette direction générale s'est récemment dotée d'un conseil scientifique. Comme le FEJ, le DRDIE adopte la double démarche d'identification et de soutien aux initiatives locales et d'expérimentations d'intérêt national. Il mobilise notamment un réseau d'animation dans les rectorats qui favorise les échanges entre le terrain et l'administration centrale. Il valorise aussi l'évaluation et mobilise la recherche en éducation dans le but

à la fois de la faire connaître et de l'orienter en fonction des besoins du système éducatif.

La partie de l'activité du FEJ qui porte sur les politiques scolaires s'inscrit donc dans une tradition d'expérimentation dont on perçoit un nouveau dynamisme. Et celui-ci n'est pas apparent uniquement dans l'administration centrale : ainsi, les Journées de l'innovation tenues les 31 mai et 1^{er} juin 2011 ont accueilli une multitude de stands d'établissements scolaires présentant des projets originaux, allant de l'entreprise junior à l'apprentissage des langues rares. D'une certaine manière leur émergence est directement liée à la notion de « projet d'établissement » (elle-même issue du foisonnement pédagogique des ZEP), mais ces projets sont, plus que jamais, valorisés par l'institution.

On peut comprendre cette attention de deux manières. Il peut s'agir de valoriser la diversité et d'encourager (par l'exemple) la différenciation, comme pour s'adapter à des publics spécifiques, ou pour mobiliser les équipes pédagogiques autour de projets originaux. Cet esprit n'est pas très éloigné de celui du programme ÉCLAIR (Écoles, collèges et lycées pour l'innovation, l'ambition et la réussite). Ou bien, il peut s'agir d'identifier des approches particulièrement pertinentes, afin de les reprendre de façon plus ou moins centralisée, de les diffuser, de les encourager. Elles alimentent alors une boîte à idée, dans laquelle l'État pourra puiser pour faire évoluer le système. Dans ce second esprit, l'évaluation est une étape nécessaire, car beaucoup d'initiatives sont séduisantes, mais il est difficile de savoir lesquelles sont les plus prometteuses. Et les acteurs eux-mêmes sont mal placés pour en juger,

ce qui plaide pour une évaluation qui soit, ou bien strictement extérieure, ou bien accompagnée d'un regard externe, et dont les conclusions sont, dans tous les cas, rendues publiques.

LES DÉMARCHES D'ÉVALUATION

Dans ce contexte, le renforcement de l'attention portée à l'évaluation est probablement l'apport spécifique du FEJ. D'abord, accompagner toute expérimentation d'une évaluation rigoureuse est, pour le FEJ, un principe de fonctionnement. Ensuite, il offre des financements dédiés et une expertise technique pour la définition et le suivi du protocole d'évaluation. Enfin, il soutient la diffusion des résultats en les publiant et en les mettant en débat.

L'évaluation des politiques publiques n'est pas une activité originale, même si le terme est extrêmement galvaudé. Mais l'évaluation, telle qu'elle est pratiquée, s'appuie trop rarement sur l'expérimentation. Or, l'expérimentation offre les variations de situations, les contrastes, qui révèlent les effets ou les conséquences d'une intervention, les différences spécifiques qu'elle produit. Par ailleurs, lorsqu'elle porte sur des interventions intégralement déployées, l'évaluation arrive souvent trop tard, car les capacités d'ajustement (ou de suppression) d'une politique en place sont généralement limitées. Dans la démarche expérimentale, les conditions de l'évaluation sont pensées dès la conception des projets : les règles du FEJ, en imposant des soumissions conjointes du projet lui-même et de son évaluation, suscitent d'emblée le dialogue du porteur et de l'évaluateur.

NOTE

3. La loi mentionne l'enseignement des disciplines, l'interdisciplinarité, l'organisation pédagogique de la classe, de l'école ou de l'établissement, la coopération avec les partenaires du système éducatif, les échanges et le jumelage avec des établissements étrangers d'enseignement scolaire.

Elles assurent aussi que des données (ce terme incluant ici les entretiens ou les monographies) pourront être produites, avant, pendant et « à côté » de l'expérimentation.

Les approches de l'évaluation soutenues par le FEJ sont multiples. Certaines analysent dans la durée le déroulement du processus de mise en œuvre de la politique considérée. À travers des analyses de cas ou des monographies, elles permettent de comprendre comment l'organisation sociale que constitue un dispositif peut s'ajuster ou se recomposer. Particulièrement importantes à cet égard sont les conduites, les contraintes et les stratégies des différents groupes d'acteurs. Les corps d'inspection ont ce type d'expertise et produisent des analyses souvent centrées sur les fonctionnements institutionnels. Ils interprètent moins souvent le système dans sa dimension sociologique, mais il n'y a pas de solution de continuité entre leur travail et les évaluations pointées par le FEJ, lequel a cherché à faire intervenir des évaluateurs plus extérieurs au système.

Par exemple, un grand nombre de projets de prévention et de lutte contre le décrochage scolaire financés par le FEJ organisent le repérage et le suivi des élèves décrocheurs à travers une collecte systématique d'informations partagées entre différentes institutions (Éducation nationale, missions locales, Pôle emploi, etc.). L'enjeu ici est de comprendre quelles solutions techniques et institutionnelles peuvent fonctionner, quels blocages se présentent et comment ils peuvent être levés. Autre exemple, la partie « qualitative » de l'évaluation du « livret de compétences expérimental », qui retient une approche moins

organisationnelle : elle s'intéresse à l'appropriation du dispositif par les différents acteurs (équipes éducatives, élèves, familles) et identifie les conditions d'appropriation, les éventuels phénomènes d'autocensure, et, de manière générale, comment le livret est mis à profit pour améliorer l'orientation des élèves. Elle étudie également la complémentarité des compétences scolaires et extrascolaires, dans une perspective d'égalité des chances : par exemple, les élèves en difficulté dans le milieu scolaire valorisent-ils leurs compétences extrascolaires ? Au contraire, certains élèves sont-ils stigmatisés ?

Un autre ensemble d'évaluations mesurent l'impact du dispositif expérimenté à travers une approche statistique. Elles visent à comparer la situation des bénéficiaires d'un dispositif avec la situation qu'ils auraient connue s'ils n'avaient pas bénéficié de ce dispositif. L'écart entre ces deux situations constitue bien la valeur ajoutée de l'intervention. Pourquoi financer ou encourager tel programme de soutien scolaire ou d'aide à l'apprentissage de la lecture si les élèves qui en bénéficient ne savent pas – où à peine – mieux lire et écrire à la fin de l'année ? Pour en juger, il faut davantage que le jugement informel des enseignants ou la conviction des bénévoles, qui estiment que les progrès des enfants « sont visibles ».

Comme on ne peut pas observer ce qu'il serait advenu des bénéficiaires en l'absence de dispositif, le défi méthodologique consiste à observer les situations d'individus non bénéficiaires mais parfaitement semblables. La difficulté tient à ce que les élèves qui sont naturellement retenus dans un dispositif donné sont *a priori* différents des autres élèves : leur situation

est en général un mauvais point de comparaison.

Il y a lieu de hiérarchiser les solutions qui peuvent être apportées à ce problème. La solution la plus fiable consiste à tirer au hasard les bénéficiaires et les non-bénéficiaires. On est alors assuré que toutes les caractéristiques initiales des deux groupes sont en moyenne semblables, à condition d'observer des effectifs suffisants. Par exemple, pour évaluer le dispositif *Coup de Pouce Clé*, qui prend en charge un groupe de 5 enfants tous les soirs après l'école afin de développer leur goût pour la lecture, on a tiré au sort les 72 écoles parmi 109 qui peuvent bénéficier du dispositif dès la première année, les 37 autres entrant dans le programme la deuxième année. Les élèves du second groupe fournissent un très bon point de comparaison pour les élèves du premier groupe. Ou bien, pour démontrer que les trois réunions organisées avec les parents d'élèves de sixième et l'équipe éducative en début d'année (sous le nom de « Mallette des parents ») sont susceptibles d'améliorer le comportement des élèves et de réduire leur absentéisme, on a tiré au sort les classes qui, à l'intérieur d'une trentaine de collèges, ont participé au programme la première année. La comparaison de ces écoles ou de ces classes à la fin de l'année donne une mesure très simple et très crédible des changements induits par le programme et par lui seul, et permet de les quantifier – dans les limites liées à la précision statistique.

Il est indéniable que le FEJ a mis une insistance nouvelle, dans le paysage français, sur les évaluations dites aléatoires. La visibilité de ces expérimentations dépasse sans doute leur poids quantitatif dans l'activité

du Fonds⁴, mais elles constituent probablement ce qu'il apporte de plus neuf dans le champ de l'évaluation. Il s'agit à nos yeux d'une approche dont le potentiel est très grand mais qui reste encore sous-utilisée.

Il est intéressant de noter que les sciences de l'éducation et notamment la psychologie de l'éducation ont été des champs pionniers dans l'utilisation de ces méthodes aléatoires. D'ailleurs, l'article de référence pour l'évaluation d'impact statistique a été publié dans *Journal of Educational Psychology* [5], même s'il a une portée beaucoup plus générale que ce seul champ. La recherche en sciences de l'éducation, principalement anglo-saxonne, qui utilise des expériences affectant aléatoirement des élèves ou des classes, porte généralement sur un très petit nombre de sujets et reste encore proche de la logique de laboratoire, connexe aux sciences cognitives. Par exemple, dans la métaanalyse des recherches sur les techniques d'apprentissage de la lecture réalisée aux États-Unis à la demande du Congrès [4], on peut voir que l'étude typique tire au sort deux classes parmi quatre, dans lesquelles elle met en place une intervention très ciblée et soigneusement contrôlée, parfois réalisée par les chercheurs eux-mêmes. Les expériences menées dans le cadre du FEJ portent sur des

politiques développées en grandeur réelle, et donc sujettes à toutes les difficultés de mise en œuvre, aux résistances et aux réactions du monde social. Elles ne mesurent pas seulement les processus cognitifs, elles testent aussi les conditions de leur déploiement sur le terrain avec des ressources et une capacité de contrôle limitées.

L'ÉVALUATION DES EXPÉRIMENTATIONS DANS LE CHAMP DE L'ÉDUCATION : PREMIER RETOUR SUR EXPÉRIENCE

Ainsi, le Fonds d'expérimentation est rapidement devenu un acteur important pour la réflexion sur les politiques scolaires. Il s'inscrit dans une tradition plus large, mais a aussi contribué à faire évoluer les pratiques dans des directions nouvelles. Le retour sur un peu plus de deux ans d'existence et près de 400 expérimentations lancées permet de tirer plusieurs enseignements.

Phase préparatoire et vertu du binôme porteur de projet-évaluateur

Une originalité du FEJ est de faire figurer la qualité de l'évaluation parmi les critères principaux de la sélection des projets financés, donnant à l'ensemble des acteurs de l'expérimentation des incitations à être ambitieux sur les protocoles d'évaluation. Le recours à des jurys rassemblant personnalités externes à l'administration et spécialistes de l'évaluation des politiques publiques a permis de rendre cette démarche

crédible et d'émettre des recommandations adaptées à chaque projet. Les procédures de sélection, comportant deux étapes, identifient tout d'abord les projets proposant de véritables politiques innovantes, puis retiennent parmi ces projets ceux dont le protocole d'évaluation est jugé convaincant. Cette double condition assure que les enseignements tirés de l'expérimentation seront à la fois pertinents et crédibles. Cette inscription de la qualité de l'évaluation dans les procédures de sélection, assortie de modalités de fonctionnement des jurys qui ont permis un réel débat autour des méthodes, apparaît centrale pour concrétiser l'ambition de protocoles d'évaluation fondés au plan scientifique.

Cette démarche de sélection des projets introduit une contrainte nouvelle pour les porteurs de projets. Faire place à l'évaluation, avec ses exigences d'observation, parfois d'introduction de populations *test* et *témoin*, est une complication pour des dispositifs parfois eux-mêmes en construction et en tâtonnement. Cependant, un constat qui peut-être paradoxal est celui d'une forte demande d'évaluation de la part des porteurs de projets eux-mêmes, une fois passées les réticences souvent exprimées aux premiers stades des discussions. Dans le cadre des appels à projets, les conditions de réponse qui prévoient que porteur de projet et évaluateur candidatent en binôme, se sont révélées utiles. En effet, dans le champ de l'évaluation des politiques publiques, la spécificité des expérimentations tient au fait que les modalités d'évaluation de la politique expérimentée font l'objet d'une discussion explicite dès la conception des projets, et sont validées par

NOTE

4. Parmi l'ensemble des expérimentations lancées depuis juillet 2009, environ 37 % impliquent un protocole d'évaluation avec assignation aléatoire. De manière plus générale, un peu plus de 55 % des évaluations procèdent par comparaison entre un groupe témoin et un groupe cible, déterminés de manière aléatoire ou non.

l'ensemble des parties prenantes. On pourrait craindre que cela déforme artificiellement la structure des projets. Mais cette naissance partenariale des expérimentations permet souvent de formaliser les dispositifs envisagés sans doute davantage qu'ils ne l'auraient été spontanément : la définition de critères d'évaluation nécessite d'exposer clairement la problématique de l'expérimentation et d'expliciter les objectifs poursuivis par le dispositif proposé, ce qui influe aussi en retour sur l'architecture d'ensemble du dispositif.

Inversement, ce travail préalable et conjoint entre porteur et évaluateur permet d'envisager des modalités d'évaluation qui seraient difficilement concevables en dehors du cadre expérimental. Dans le cadre d'évaluations quantitatives, elle permet ainsi de prévoir le recours à des groupes de contrôle rigoureusement définis et observés dans le temps de manière systématique, avec des outils de mesure adaptés aux objectifs du dispositif. L'évaluation conçue *après* le déploiement d'une politique permet rarement d'obtenir cette qualité de suivi. Ce point est particulièrement évident pour les évaluations aléatoires car la sélection aléatoire des bénéficiaires doit être intégrée d'emblée à la conception du projet et approuvée par l'ensemble des acteurs préalablement à la sélection. Les modalités acceptables d'une telle sélection (aussi bien d'un point de vue éthique que politique, et d'insertion dans la logique d'ensemble du projet) font l'objet d'un long travail préparatoire. Dans le cadre de méthodes qualitatives, la mobilisation des porteurs, dès la phase de candidature, permet une meilleure compréhension des modalités de l'évaluation, qui ne sont plus

perçues comme subies. L'accès aux terrains s'en trouve ensuite facilité et permet à l'évaluateur d'observer les conditions de mise en œuvre du projet très en amont de sa réalisation : à nouveau, le temps et les conditions de l'observation sont améliorés par la construction simultanée du projet et de son évaluation.

Ce travail conjoint est propre à la démarche d'appel à projets. Certaines expérimentations, cependant, sont conçues à une échelle nationale et les modalités d'évaluation définies par les équipes du FEJ en lien avec le responsable institutionnel du dispositif. Dans ce cas, les modalités d'évaluations peuvent être imposées aux porteurs effectifs. Par exemple, dans le projet « 10 000 permis pour réussir », des acteurs locaux (typiquement des missions locales) pouvaient candidater pour mettre en œuvre la subvention de 1 000 euros sur les cours de conduite proposée aux jeunes. Mais elles devaient accepter un dispositif de tirage au sort, dont seules certaines modalités de mise en œuvre pouvaient être discutées après coup. Il ne fait pas de doute que cette évaluation permettra de mesurer précisément l'utilité d'une démarche généreuse mais dont les effets, notamment sur l'insertion professionnelle, sont totalement inconnus ; mais on ne se situe pas ici dans le cadre où le travail en amont du porteur et de l'évaluateur permet de faire progresser l'ensemble de la démarche. Dans d'autres cas, par exemple celui de *Coup de pouce Clé* déjà évoqué plus haut, le dispositif d'évaluation a été défini avec les principaux responsables, l'association qui porte le projet, l'administration centrale de l'Éducation nationale et les rectorats. En revanche, il a été imposé aux écoles,

pourtant directement concernées. De toute évidence, un travail d'information et d'explicitation au plus près du terrain est nécessaire et le calendrier ne permet pas toujours de le mener dans de bonnes conditions.

L'offre d'évaluation externe

Une des spécificités importantes du FEJ consiste à confier l'évaluation des expérimentations à des équipes externes à la fois aux administrations concernées et aux structures porteuses des projets. Cette caractéristique apparaît centrale pour garantir l'indépendance du jugement des équipes d'évaluation. Ces équipes sont formées essentiellement d'universitaires ou de membres de cabinets de conseil et d'évaluation des politiques publiques. Au fil des vagues de nouvelles expérimentations, le nombre d'équipes d'évaluation impliquées dans la démarche du FEJ a cru de manière considérable, atteignant aujourd'hui près de 170 évaluateurs différents. Ces évaluateurs ont été sélectionnés soit en se portant candidats en binôme avec des porteurs de projet dans le cadre des appels à projets thématiques, soit en répondant aux appels d'offres d'évaluation. Le constat de cette ouverture considérable aux acteurs de l'évaluation des politiques publiques, tant universitaires que privés, s'accompagne néanmoins d'une relative difficulté à faire croître le nombre d'évaluateurs de qualité répondant aux sollicitations du FEJ. Cette difficulté s'incarne dans le nombre relativement limité de réponses aux appels d'offres d'évaluation publiés par le FEJ, particulièrement du fait d'équipes universitaires dont la veille semble moins systé-

matique, et la posture d'évaluateur embarqué auprès d'un porteur de projet moins naturelle. La mobilisation d'un vaste réseau d'évaluateurs possédant les compétences méthodologiques nécessaires à la diversité des missions d'évaluation définies par le FEJ semble ainsi une condition importante à la réussite des futures expérimentations.

Le choix des méthodes d'évaluation

Si l'ensemble des dispositifs à évaluer ne se prête pas à toutes les méthodes d'évaluation, le principe général des évaluations du FEJ est d'articuler méthodes quantitatives et qualitatives sur un même objet, afin d'évaluer à la fois l'impact de la politique expérimentée, et d'apprendre de ses conditions de mises en œuvre. Pour ce qui est des dispositifs nationaux, c'est par exemple le cas des évaluations du *livret de compétences expérimental* ou de *Coup de pouce clé* ; dans le cadre des appels à projets, l'expérimentation « Promotion de l'apprentissage et sécurisation des parcours des jeunes apprentis en France » portée par le Conseil national des missions locales, associe également les deux approches. Les deux volets de l'évaluation, quantitatif et qualitatif, sont parfois mis en œuvre par la même équipe, mais les compétences méthodologiques très différentes qui sont sollicitées peuvent également conduire à les faire réaliser par deux équipes différentes.

Le recours à des méthodes quantitatives d'évaluation connaît un préalable important, qui doit être souligné : la capacité à mesurer et à résumer statistiquement les principaux impacts du projet. Pour obtenir

une précision statistique suffisante, permettant notamment de discerner convenablement des différences éventuelles entre deux groupes ou des évolutions dans le temps, il faut que les dispositifs soient menés à une échelle suffisante, et sur des populations de bénéficiaires potentiels assez larges. Cela peut sembler contradictoire avec la notion même d'expérimentation, et c'est parfois le cas. Cependant, le FEJ finance des évaluations à très large échelle telles que « La Mallette des parents », le revenu contractualisé d'autonomie ou « 10 000 permis pour réussir », portant chacun sur plusieurs milliers de jeunes ou d'élèves. Mais d'autres dispositifs, dont on aurait souhaité connaître l'impact, n'ont pu bénéficier d'évaluations quantitatives. Ainsi, l'expérimentation « Sport l'après-midi » porte sur la question fondamentale des rythmes scolaires. Il aurait été extrêmement intéressant d'établir avec précision si non seulement les résultats, mais aussi l'engagement scolaire, la motivation, les comportements se trouvent améliorés lorsque des élèves bénéficient d'une telle réorganisation des emplois du temps. Malheureusement, le nombre d'établissements mobilisables rapidement était trop faible pour que des comparaisons entre élèves bénéficiaires ou non du dispositif aient la moindre signification. À défaut, le risque est que le protocole d'évaluation conduise systématiquement à déclarer le dispositif inefficace, car le trop petit nombre d'élèves ou d'établissements bénéficiaires ne permet pas d'obtenir une précision statistique suffisante pour détecter des différences de résultats, y compris importantes, entre groupes-test et témoin. Dans ce cas, l'évaluation se contentera d'analyser la mise en

œuvre de la réorganisation, les obstacles qu'elle rencontre, les modifications qu'elle induit dans le travail des différents acteurs – mais elle ne pourra pas déterminer d'impact à proprement parler sur des résultats finaux.

Une difficulté accessoire tient au recueil des données nécessaires. En milieu scolaire, le système d'information des établissements ne contient qu'une partie des données qui sont pertinentes, soit pour décrire l'intervention, soit pour mesurer des résultats ; et dans le premier degré, ces systèmes d'information sont pratiquement inexistantes. Par ailleurs, une grande partie des effets pertinents peuvent être mesurés par des tests, qu'ils soient cognitifs ou conatifs, et la passation de ces tests peut impliquer des contraintes lourdes, comme la banalisation d'heures de cours. Cet effort demandé aux établissements est généralement bien accepté à partir du moment où l'usage des tests est clairement expliqué : ils permettent de construire une connaissance précise sur l'impact de politiques scolaires, et sont assortis de règles de confidentialité très strictes dans leur manipulation, garantissant notamment qu'ils n'aient pas d'autre usage que l'évaluation d'un dispositif précis.

Enfin, il faut souligner que le travail quantitatif étant de nature statistique, il revient, pour le dire vite, à construire des moyennes. Cela implique que l'analyse de moyennes soit utile et intéressante. Ce n'est pas le cas lorsque de telles moyennes agrègent des déclinaisons extrêmement variables d'un projet. Certaines expérimentations constituent un temps de tâtonnement pour construire sur le terrain des dispositifs dont les principes généraux sont établis, mais

les moyens de les mettre en œuvre encore imprécis. Dans ce cas, seule l'analyse qualitative des projets est véritablement pertinente, et elle est particulièrement passionnante. Le travail statistique n'a de sens que lorsque les interventions sont assez précisément formalisées.

Contexte expérimental et faisabilité des évaluations aléatoires

Après un peu plus de deux ans de recul sur les expérimentations, la présence d'un contexte véritablement expérimental apparaît comme une condition *sine qua non* à la mise en place d'évaluations aléatoires. En son absence, des tentatives d'évaluations avec assignation aléatoire menées dans le cadre du FEJ se sont révélées infructueuses, mais néanmoins riches d'enseignement pour la mise en place de futures évaluations de cette nature.

D'un point de vue épistémologique, l'évaluation d'impact ne peut porter que sur des interventions qui sont *manipulables*, c'est-à-dire lorsqu'on peut concevoir qu'une même personne soit ou ne soit pas soumise à la mesure à évaluer. Dans la pratique, cela implique que les évaluations aléatoires, qui contrôlent les états de bénéficiaire et de non-bénéficiaire ne sont réalisables que lorsque le contexte est véritablement expérimental. *A contrario*, lorsqu'un dispositif se trouve déjà à un stade où il est largement étendu, ou bien lorsque le profilage de ses bénéficiaires est très fin et en limite considérablement le champ d'application, l'évaluation aléatoire s'avère très difficile à insérer. Par ailleurs, la co-construction, par le porteur et l'évaluateur, de l'ensemble du dispositif projet-évaluation, dont on a vu qu'elle était si importante, ne peut avoir lieu. Le travail mené en vue de l'évaluation aléatoire des *Écoles de la deuxième chance* ou de l'*Établisse-*

ment public d'insertion de la défense (EPIDe), très tôt dans l'existence du Fonds, a révélé cette difficulté (cf. encadré). L'évaluation de *Coup de Pouce Clé*, dispositif développé depuis plusieurs années, n'a pu être menée qu'à l'occasion d'une extension dans de nouvelles zones.

L'incertitude, en partie irréductible, quant à la possibilité d'évaluer rigoureusement certains dispositifs, plaide pour généraliser les études de faisabilité ou phases pilote préalables au déploiement des dispositifs expérimentaux. Dans le cadre du FEJ, ces études de faisabilité ont été menées au cas par cas. La plupart a confirmé la possibilité d'évaluer les expérimentations prévues, mais plusieurs, à l'image du très médiatisé projet de « bourse aux projets de classe » dans l'académie de Créteil ou de l'évaluation de l'EPIDe, ont conduit à l'abandon des expérimentations, les conditions d'une évaluation d'impact probante n'étant pas réunies.

Les évaluations portées par le Fonds d'expérimentation pour la jeunesse des *écoles de la deuxième chance* (E2C) et de l'*Établissement public d'insertion de la défense* (EPIDe), malgré des protocoles expérimentaux différents (le point commun étant le tirage au sort individuel des bénéficiaires), ont connu un destin commun, menant dans les deux cas à l'abandon des évaluations engagées. Si les options techniques retenues étaient différentes, ces deux évaluations prenaient place dans un **contexte institutionnel similaire** :

- tout d'abord, par l'**absence de caractère expérimental** : l'évaluation ne concernait pas un dispositif expérimental et conçu comme tel, qui soit nouveau pour les acteurs, et qui intègre d'emblée des modalités d'évaluation conçues avec la participation et le consentement de chacun des acteurs. Les dispositifs préexistants à l'évaluation, celle-ci s'est trouvée contrainte de se greffer sur des pratiques et des modes de fonctionnement solidement établis ;
- ensuite, par la **nature des dispositifs** et le ciblage de leurs bénéficiaires : des dispositifs répondant aux besoins de jeunes possédant des caractéristiques très spécifiques, dans des centres aux modes de recrutement pluriels et différents entre centres, avec une prescription par les missions locales (non directement concernées par le dispositif) prenant une place importante, mais non exclusive ;
- enfin, par le **manque de connaissances préalables**, non seulement quant à l'impact de ces dispositifs, mais aussi plus simplement quant aux caractéristiques et aux trajectoires des jeunes bénéficiaires et aux procédures de recrutement et de sélection des bénéficiaires.

Les évaluations soutenues par le FEJ dans ce contexte se proposaient de mieux définir le profil et les trajectoires des jeunes bénéficiaires des deux dispositifs, en développant des enquêtes dont la richesse permettait d'aller bien au-delà des données actuellement disponibles. Elles permettaient également de déterminer l'impact des deux dispositifs sur ces trajectoires, en comparant les trajectoires des jeunes bénéficiaires à celles de jeunes possédant des caractéristiques parfaitement similaires, déterminés de manière aléatoire. Les difficultés constatées dans la mise en place de ces protocoles d'évaluation avec assignation aléatoire initialement ont toutefois permis d'apprendre sur les conditions de faisabilité de telles évaluations et de mieux comprendre le contexte institutionnel dans lequel ces évaluations s'insèrent.

Enfin, exactement comme dans le contexte médical, les évaluations avec assignation aléatoire soulèvent des enjeux éthiques. Le principe général est d'accepter de priver provisoirement certaines personnes d'une innovation dans le but d'obtenir une information précise sur l'intérêt qu'elle présente. Comme dans un protocole médical, on ne saurait nier qu'il y a là un coût pour les personnes privées d'une intervention potentiellement utile. On considère ici que ce coût est justifié par le bien public (la connaissance) qui est produit.

Moins que le principe d'exclusion (qui est largement mis en œuvre dans les situations non-expérimentales), c'est la forme aléatoire qui est mal vécue. Ce n'est pas toujours le cas, cependant, et il n'est pas rare que les acteurs, notamment les jeunes, perçoivent une plus grande justice dans le tirage au sort que dans une assignation par un responsable (équipe pédagogique, directeur, recteur) qui peut être perçue comme arbitraire et répondant à des règles implicites et variables avec l'identité du responsable.

En tout état de cause, les dispositifs aléatoires doivent respecter un certain nombre de règles (y compris la validation par un comité d'éthique). Ainsi, le tirage au sort n'est justifié que s'il porte sur un dispositif qui vient s'ajouter au droit commun : en aucun cas l'expérimentation ne conduit à priver un élève ou un jeune du bénéfice d'avantages préexistants (en vérité, cette condition s'applique à toute forme d'expérimentation). En matière de pédagogie par exemple, il serait difficile de substituer radicalement une pédagogie à une autre : les dispositifs existants travaillent plutôt en superposant des modules

d'intervention. L'éthique médicale met en avant la notion de « consentement éclairé », qui a été conçue en réaction à des situations historiques où des substances potentiellement dangereuses ont été administrées à leur insu à des patients vulnérables. Dans les politiques sociales, la substance est rarement dissimulable et, dans le contexte scolaire, il peut revenir à l'autorité pédagogique d'en décider l'administration, si bien que ce principe ne s'applique pas de façon automatique. Mais certains protocoles peuvent prévoir une forme de consentement éclairé en demandant aux jeunes, aux familles ou aux établissements (selon les cas) de se porter volontaire pour entrer dans le dispositif. La sélection aléatoire s'applique par ailleurs facilement à des situations où il y a plus de volontaires que de places disponibles : le tirage au sort est alors une façon de gérer le rationnement qui se substitue à d'autres règles possibles. Dans le cas où il est possible d'identifier sans hésitation des publics extrêmement prioritaires pour bénéficier d'un dispositif, il est possible de prévoir qu'un nombre limité d'individus soient soustraits au tirage au sort.

IMPLICATION DE POLITIQUE PUBLIQUE : LA VALIDITÉ EXTERNE DES EXPÉRIMENTATIONS

Les notions de « généralisation » et « d'essai » sont souvent attachées à celle d'expérimentation. Elles impliquent de chercher à déterminer, à travers l'expérimentation et l'évaluation qui l'accompagne, si un dispositif doit être répliqué sur d'autres territoires. C'est là une des motivations

originelles du FEJ. Prise au pied de la lettre, cette idée pose cependant plusieurs problèmes.

D'une part, on expérimente souvent des interventions qui cherchent leur forme même. Dans certains cas, les dispositifs sont ancrés sur des territoires : ils ne sont pas exportables tels quels, donc pas simplement généralisables. L'évaluation doit permettre de comprendre quel est la nature du travail qui permet à un principe d'intervention de s'appliquer, elle doit analyser le précédent que constitue l'expérimentation, de manière non pas à la reproduire à l'identique, mais à identifier et faire connaître les difficultés à contourner et les démarches qui permettront de la décliner ailleurs. C'est typiquement ce qu'on pourrait attendre de l'évaluation de « Sport l'après-midi », tant il est probable que la déclinaison du principe par les établissements sera variable et sensible aux conditions locales.

D'autre part, lorsqu'un dispositif est précisément formalisé et facilement reproductible ailleurs, on peut être confronté au problème de la validité externe de l'évaluation. La validité interne signifie que l'évaluation démontre un effet qui s'interprète correctement comme l'impact du dispositif à l'intérieur de l'expérimentation, c'est-à-dire dans le contexte dans lequel elle a été développée et sur la population directement concernée. Mais, la démarche évaluative associée à l'expérimentation étant radicalement empirique, elle ne vaut, en toute rigueur, que pour les données qui ont été examinées. La question de la validité externe reste donc en principe ouverte : les mêmes effets se reproduiront-ils ailleurs et notamment dans les contextes, peut-être différents, d'une éventuelle généralisation ?

La validité externe potentielle d'une expérience doit être examinée au cas par cas. Par exemple, les clubs *Coup de pouce Clé* étant le fruit d'une organisation et d'une méthode dont le format est bien défini, il n'y a pas de raison de penser que leurs effets soient très différents lorsqu'on passe des élèves en difficulté de l'éducation prioritaire d'une académie, à ceux d'une autre académie. D'ailleurs, en l'espèce, l'expérimentation est menée sur trois académies et le profil des élèves y est remarquablement semblable. À l'autre extrémité se trouvent les actions qui, dans la généralisation, ont des effets systémiques. Un exemple classique est celui de la réduction de la taille des classes. L'expérimentation sur quelques dizaines de classes, dans lesquelles les élèves ont été affectés aléatoirement aux États-Unis (expérimentation STAR) a donné des résultats probants [2]. Généralisée, elle implique cependant des recrutements massifs d'enseignants, comme ce fut le cas dans certains états américains, inspirés par les résultats de cette expérimentation. Or, certains craignent que le niveau de formation plus bas de ces nouveaux enseignants ne contrebalance l'effet de réduction de la taille des classes. Il s'agit toutefois d'un exemple extrême et les politiques expérimentées par le FEJ ne sont pas susceptibles d'avoir des effets systémiques aussi importants.

Entre ces deux cas extrêmes, la difficulté la plus classique tient à la sélection des écoles ou des établissements : ils sont généralement volontaires et ont réfléchi à un projet en amont, ce qui implique que la mise en œuvre pourra être plus difficile ou rencontrer des obstacles différents si elle est étendue, voire imposée. Par

ailleurs, les effets peuvent être hétérogènes (par exemple il est bien documenté que l'impact de la réduction de la taille des classes est plus fort sur les élèves les plus faibles). Il est alors important de pouvoir évaluer des impacts potentiellement différents sur différentes sous-populations.

Selon leur nature, les expérimentations ont ainsi une validité externe plus ou moins grande, qui doit faire l'objet d'une discussion. Cela n'enlève rien ni à leur force, ni à leur utilité – à moins de considérer, radicalement, que l'on ne peut rien apprendre de l'expérience. Mais, pour le comprendre, il faut absolument renoncer à un usage naïf de l'évaluation : elle ne donne pas des réponses toutes faites, elle ne promeut pas des dispositifs « clé en main » que l'on peut dupliquer à la chaîne, bref elle n'est pas une technocratie. Au contraire, elle donne des éléments qui sont dans certains cas presque auto-suffisants, mais qui, dans d'autres cas, ne constituent qu'une pièce d'un mécano – pièce centrale dont on veut être assuré du bon fonctionnement, mais que l'on devra inclure dans un ensemble, qui reste à penser, avec ce que cela comporte de risque et de choix politique. Comment peut-on envisager la réduction de la taille des classes sans être d'abord passé par l'examen de ses effets à petite échelle ? Mais la réponse à cette question laisse ouvertes les modalités de construction complète d'une politique à large échelle.

On peut naturellement chercher à étendre le champ d'application de l'expérimentation. Ainsi, dans les pays en développement, on cherche de plus en plus souvent à reproduire en différents lieux telle expérimentation qui a démontré localement l'efficacité d'une politique. On peut

aussi changer d'échelle, expérimenter des interventions à l'échelle de « marchés » entiers, ou concevoir des protocoles expérimentaux qui révèlent l'impact d'une intervention sur les jeunes qui n'en sont pas directement bénéficiaires. Il y a de nombreux exemples de telles démarches, mais ils exigent d'avoir une réelle ambition pour l'évaluation.

Il reste que, dans beaucoup de cas, on ne peut pas éviter que l'usage qui est fait d'expérimentations situées implique une capacité d'extrapolation, la mise en jeu d'une intuition sur la pertinence d'une politique dans d'autres contextes. En démontrant qu'il est possible d'impliquer les parents dans le collège et que cette implication a des effets sur les élèves, on n'apprend pas seulement que les réunions de la « Mallette des parents » sont utiles. D'ailleurs, elles sont généralisées sous des formes assez souples, et sans doute variables d'un collège à un autre. On apprend surtout qu'il y a dans le lien entre l'école et les parents une ressource insuffisamment exploitée et qu'il peut être très efficace de mobiliser. Au fond, l'expérimentation se comprend surtout dans un esprit de recherche : elle révèle des mécanismes qui ont une valeur générique. Elle propose des outils de pensée plus que des réponses toutes faites. Mais ces outils ne sont pas abstraits et spéculatifs : ils sont fondés rigoureusement et empiriquement.

CONCLUSION

Le Fonds d'expérimentation pour la jeunesse est parvenu en très peu de temps à développer à une échelle nouvelle les expérimentations accompagnées d'évaluations. Il s'agit d'un formidable élan, qui s'inscrit assez

aisément dans la dynamique en cours dans l'Éducation nationale et contribue à la renforcer. Nous avons beaucoup appris sur la mise en œuvre des expérimentations et nous allons apprendre plus encore sur les politiques en faveur des jeunes.

Malgré les difficultés et les échecs passés et à venir, nous avons appris qu'il était possible d'être ambitieux en exigeant des expérimentations qu'elles soient l'occasion d'apprendre sur des bases transparentes et rigoureuses. Nous avons montré notamment qu'il était possible de mettre en œuvre des expériences contrôlées, dont les limites sont connues et doivent être bien posées, mais dont l'apport est très important et qui constituent une nouveauté radicale en France dans la période récente. Le travail du FEJ a contribué à les installer dans le paysage de l'évaluation, sans négliger ni sous-estimer les autres approches.

Aujourd'hui, le FEJ a devant lui le défi de la capitalisation. Les évaluations sont un bien public, et c'est ce qui justifie les efforts pour les produire. Cela implique d'abord que tous les rapports remis au FEJ doivent être immédiatement rendus publics et pouvoir faire l'objet de débats sur leur qualité et leurs implications. Mais, pour que les acteurs des politiques publiques, administrations, collectivités locales, associations, puissent s'approprier la masse d'information accumulée et s'en servir pour concevoir des projets futurs, il est nécessaire de les synthétiser tout en faisant ressortir les résultats des évaluations les plus probantes. Cela pose des problèmes de ressources (en raison de la masse très importante de projets) et des difficultés de méthode (car les approches de l'évaluation sont variées, les enseignements peuvent être contradictoires

ou liés à des spécificités insuffisamment apparentes).

Il reste à apprendre comment les acteurs vont se saisir des résultats et si ceux-ci vont influencer les politiques. On sait que la temporalité de l'expérimentation n'est pas toujours celle de la décision publique. On pourrait rêver d'un processus de conception des politiques publiques qui épouse un temps plus long, se donne celui de la réflexion et de l'expérimentation. La fréquence et la précarité des réformes scolaires plaident certainement pour un tel changement. Mais il y a d'autres conditions : il faut être capable de cumuler les résultats, de les articuler, de les rendre intelligibles. Il faut enfin qu'ils soient visibles, qu'on sache qu'ils existent, qu'on sache où les trouver. Il s'agit à nouveau d'un terrain sur lequel le FEJ s'apprête à expérimenter. ■

BIBLIOGRAPHIE

[1] **Kherroubi M. et Rochex J.-Y.** (2002), La recherche en éducation et les ZEP en France. 1^{ère} partie : Politique ZEP, objets, postures et orientations de recherche, *Revue française de pédagogie*, n° 140, pp. 103-131.

[2] **Krueger A.** (1999), Experimental Estimates of Education Production Function, *Quarterly Journal of Economics*, 114, pp. 497-532.

[3] **Moisan C. et Simon J.** (1997), *Les déterminants de la politique scolaire en zone d'éducation prioritaire*. La Documentation française, Paris.

[4] **National Institute of Child Health and Human Development** (2000), *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). U.S. Government Printing Office, Washington, DC.

[5] **Rubin D.B.** (1974), Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, vol. 66, pp. 688-701.