

n° 81 [mars 2012]



Éducation *formations*

**Expérimentations :
innovation,
méthodologie,
évaluation**



n° 81 [mars 2012]

Éducation *formations*



**Expérimentations :
innovation,
méthodologie,
évaluation**

Directeur de la publication

Michel Quéré

Rédactrice en chef de ce numéro

Caroline Simonis-Sueur

Secrétaire de rédaction

Marie Zilberman

PAO

Linéal

Photo couverture

©Caroline Lucas/MENJVA

Éducation & formations

Ministère de l'éducation nationale,
de la jeunesse et de la vie associative
Direction de l'évaluation, de la prospective
et de la performance (DEPP)
61-65, rue Dutot – 75732 Paris Cedex 15

Impression

Ovation

Tarif 2011

Le numéro France : 13 euros

Vente au numéro

Ministère de l'éducation nationale,
de la jeunesse et de la vie associative
DEPP/Département de la valorisation et de l'édition
61-65 rue Dutot – 75732 PARIS CEDEX 15

☎ : 01 55 55 72 04

Éducation & formations accepte bien volontiers d'ouvrir régulièrement ses colonnes à des spécialistes n'appartenant pas à la Direction de l'évaluation, de la prospective et de la performance.

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux-mêmes, et non les institutions auxquelles ils appartiennent, ni, a fortiori, la Direction de l'évaluation, de la prospective et de la performance.

Ce numéro thématique d'*Éducation & formations* rassemble des contributions d'origines diverses dont le regroupement vient éclairer les multiples facettes de l'importance prise par la dynamique des expérimentations dans le système éducatif français. La variété des auteurs répond volontairement à la variété des acteurs qui façonnent la transformation de ce système. Chercheurs universitaires, agents de l'administration centrale du ministère en charge de l'éducation et agents des services académiques forment ainsi un ensemble éditorial qui rend compte de façon riche et complémentaire de l'importance progressivement prise par les expérimentations dans les transformations observables du système éducatif.

Force est de constater que l'expérimentation et l'évaluation sont devenues, ces dernières années, des pratiques quotidiennes des acteurs de l'École que nos administrations respectives s'efforcent de stimuler, d'encourager et d'analyser, de façon à promouvoir « ce qui marche le mieux » et de contribuer progressivement aux changements en profondeur qui se jouent, à petite échelle, au travers des expérimentations et de leurs évaluations.

Ainsi, cet ensemble de contributions constitue moins un état des lieux qu'une sorte de cadre de référence auquel chaque acteur du système éducatif, qu'il soit parent d'élève, enseignant ou décideur au sein de ce système, peut trouver illustration, voire enseignement, pour sa propre réflexion et expérience.

Cette ambition du numéro est d'autant plus importante qu'elle s'inscrit dans un contexte national qui, jusqu'à un passé récent, n'a que peu d'expérience de l'expérimentation des politiques publiques et de son évaluation. Dans le champ éducatif, la loi du 23 avril 2005 d'orientation et de programme pour l'avenir de l'École est certainement un tournant. Outre l'introduction du socle commun de connaissances et de compétences qui fournit une nouvelle orientation et référence éducative, l'expérimentation connaît, depuis cette loi, un essor conséquent auquel a répondu la diversification des outils et méthodologies pour en mesurer les effets sur les pratiques enseignantes, en évaluer l'impact sur la réussite des élèves et apporter aux décideurs publics les éléments d'appréciation nécessaires au pilotage d'un système complexe confronté à une demande d'efficacité toujours plus forte.

La création du Fonds d'expérimentation pour la jeunesse à la fin 2008 a permis un nouvel élan de cette transformation en bouleversant les modalités traditionnelles de l'administration de mission

« à la française ». Actions plus déconcentrées, plus diffuses sur l'ensemble du territoire national, assorties par construction d'une procédure d'évaluation spécifique à chaque projet engagé, le Fonds d'expérimentation pour la jeunesse a constitué un support naturel des expérimentations dans le champ de l'éducation.

Ainsi, c'est de l'expérience accumulée par ces changements de contexte que ce numéro veut témoigner. Il le fait selon différents angles d'attaque que sont le recensement du périmètre des types d'actions expérimentées, les jeux d'acteurs dans la construction de ces expérimentations, les enjeux méthodologiques assortis à l'élaboration ou à la mesure de la performance relative de ces projets.

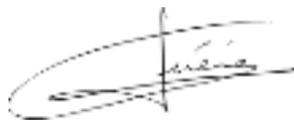
Cette variété des contributions entend aider à ce que des observateurs, acteurs ou partenaires du système éducatif, puissent mieux connaître les ressorts de l'expérimentation en éducation, les expériences engagées ici et là au service de la transmission des savoirs et des compétences en direction des jeunes de notre pays.

Le directeur général de
l'enseignement scolaire



Jean-Michel BLANQUER

Le directeur de l'évaluation,
de la performance et de la
prospective



Michel QUÉRÉ

Le directeur de la jeunesse,
de l'éducation populaire et
de la vie associative



Yann DYÈVRE

n° 81 [mars 2012]

Expérimentations : innovation, méthodologie, éducation

1^{re} partie – Expérimentations : conception et pilotage

-
- 9 L'expérimentation comme instrument d'action publique en éducation
B. Robert et J. Teillard
-
- 19 L'expérimentation *bottom up*, l'exemple ultra-marin de l'académie de la Réunion
P. Chabernaud
-
- 27 Le Fonds d'expérimentation pour la jeunesse et les politiques éducatives :
premier retour d'expérience
M. Gurgand et M. Valdenaire
-

2^e partie – Évaluations et méthodologie d'évaluation

-
- 41 Les méthodes d'expérimentation en question
D. Fougère
-
- 49 Petit précis méthodologique d'évaluation qualitative
S. Divay
-
- 55 Typologie des élèves de Poitou-Charentes âgées de 11, 13 et 15 ans.
Valorisation régionale de l'enquête internationale HBSC
S. Robin *et al.*
-
- 67 L'autoévaluation pour accompagner l'expérimentation et le projet d'établissement ?
Quelques pistes à partir d'expériences anglo-saxonnes
R. Normand
-

3^e partie – Expérimentations et résultats d'évaluations

-
- 79 Évaluation de l'effet du dispositif d'enseignement intégré de science et technologie (EIST).
Premiers résultats de l'analyse des progressions des élèves sur trois temps de mesure
M. Le Cam et T. Rocher
-
- 91 PACEM : une expérimentation sur l'utilisation d'évaluations standardisées des acquis
des élèves par les enseignants
J-F. Chesné et S. Prost
-

101 L'accompagnement des collégiens dans leur recherche de stage :
une évaluation expérimentale
Y. L'Horty, E. Duguet, P. Petit

109 Améliorer l'accès des jeunes ayant moins d'opportunités à la mobilité internationale.
Le cas appliqué d'une stratégie territoriale
S. Carel, F. Déloye

Expérimentations : conception et pilotage

L'expérimentation comme instrument d'action publique en éducation

Bénédicte Robert,

Jérôme Teillard,

DRDIE, DGESCO, MENJVA

Dans un contexte d'attente sociale plus forte sur les résultats de l'École et de personnalisation des parcours, l'expérimentation se présente comme un mode d'action pour dépasser certaines limites de l'action publique dans une société complexe, et conduire le changement, en se fondant sur des travaux de recherche et le retour de l'expérience conduite. Technique développée en dehors du champ de l'action éducative, l'approche expérimentale se trouve, de fait, influencée par ce terrain d'expérience qu'elle contribue à façonner en lui apportant de nouvelles perspectives.

Définie comme un processus collectif, organisé et évalué, permettant d'organiser l'exploration de réponses nouvelles pour satisfaire à l'exigence de réussite scolaire des élèves fixée par la société, l'expérimentation dans le champ de l'éducation fait, ces derniers temps, l'objet d'attentions et d'interrogations renouvelées.

Les uns en soulignent l'intérêt autant que l'urgence comme méthode de réforme de l'École ; d'autres en dénoncent l'usage excessif, voire les limites, pour produire, en pratique, des enseignements suffisants pour être généralisés à l'ensemble du système éducatif¹. Au cœur de l'actualité aujourd'hui, la question de l'intérêt et de l'utilisation de l'expérimentation pour initier et conduire le changement dans les politiques publiques doit être remise en perspective tant historiquement qu'en l'élargissant à d'autres contextes.

L'expérimentation en éducation n'est pas, loin s'en faut, une préoccupation exclusivement nationale : l'existence d'un sommet mondial de l'Innovation en éducation, la prise en compte de cette dimension dans le cadre européen ou encore l'organisation d'un pilotage national de l'innovation même dans les pays fédéraux

comme les États-Unis, attestent de l'acuité de la question pour les acteurs et responsables des différents systèmes éducatifs².

Par ailleurs, pour s'en tenir à la France, il faut relever que la question de l'expérimentation n'est ni totalement inédite, ni véritablement spécifique à l'École. Elle n'est pas davantage déconnectée des débats plus généraux sur les principes et réflexions mobilisés pour conduire l'action publique.

Ainsi, le législateur avait déjà, par la loi du 11 juillet 1975 relative

NOTES

1. *Une ambition pour dix ans*, rapport de la Commission pour la libération de la croissance présidée par Jacques Attali, octobre 2010, La Documentation française ; rapport de la mission commune d'information sur l'organisation territoriale du système scolaire et sur l'évaluation des expérimentations locales en matière d'éducation « *De la pyramide aux réseaux : une nouvelle architecture pour l'école* », Jean-Claude CARLE, Sénat, juin 2011.

2. Le Sommet mondial de l'innovation en éducation (WISE) qui s'est tenu à Doha (Qatar) du 1^{er} au 3 novembre 2011 s'est donné comme thème de réflexion « *Changing Societies, Changing Education* » ; conférence « Expérimentons ensemble contre le décrochage scolaire » du séminaire de contact européen Comenius consacré à la lutte contre le décrochage scolaire du 15 au 18 novembre 2011 à Amiens.

à l'éducation, dite loi Haby, retenu la possibilité de permettre aux établissements publics et privés d'enseignement de déroger aux dispositions du code de l'éducation pour conduire des « expériences pédagogiques » et ce, pour l'ensemble des niveaux scolaires³. Tout en mettant en place une réforme nationale de l'organisation scolaire, le législateur permettait donc que s'engagent, à une échelle et pour une durée limitées, et dans un cadre contrôlé par le ministre, des actions particulières de recherche et d'expérimentation pédagogiques.

Ce faisant, les politiques d'éducation s'inscrivaient dans un mouvement plus général d'ouverture du droit public à l'expérimentation, dont on attribue généralement la genèse à la loi Veil du 17 janvier 1975 relative à l'interruption volontaire de la grossesse. Ce mouvement a trouvé un aboutissement dans la révision constitutionnelle de 2003⁴, dont est issu l'article 34 de la loi du 23 avril 2005 d'orientation et de programme pour l'avenir de l'École promouvant l'expérimentation en éducation.

Comme en d'autres champs de l'action publique, l'expérimentation en éducation doit s'articuler avec le principe d'égalité, décliné, en l'espèce, sous la forme de l'égal accès à l'instruction. Elle y apporte un aménagement, limité dans l'espace et le temps, qui justifie, s'agissant d'un service public national, que l'expérimentation soit à la fois autorisée en amont et évaluée en aval par l'autorité académique.

Surtout, l'expérimentation s'inscrit dans une logique d'égalité de résultats que l'on pourrait définir comme une déclinaison exigeante de l'égalité des chances, objectif volontariste du système éducatif⁵. En effet, l'ex-

périmentation vise à vérifier, jusque dans ses résultats finaux (réussite scolaire, insertion sociale et professionnelle, absentéisme, etc.), la pertinence d'une approche renouvelée de la politique éducative. D'une certaine façon, l'expérimentation rejoint la logique d'une « discrimination positive à la française » (Conseil d'État, 1998) : sa légitimité à rompre avec l'égalité de traitement est d'œuvrer à une égalité de résultats, de porter plus haut le principe d'égalité pour passer de l'accès de tous à l'école à la réussite individuelle de chacun.

Si on se recentre sur le contexte présent du système éducatif, caractérisé par des demandes éducatives plus nombreuses et plus variées, en lien avec une différenciation sociale croissante, qui rendent plus complexe la mise en œuvre des politiques d'éducation, on voit que l'expérimentation constitue un instrument d'action publique permettant de construire des politiques d'éducation fondées sur des travaux de recherche, le retour de l'expérience, et prenant en compte, plus généralement, la contribution des acteurs du système éducatif, à la réussite des politiques d'éducation. S'inscrivant dans le champ de l'action publique, le recours à l'expérimentation se trouve également nourri et déterminé par cet environnement.

LE PROCESSUS COMPLEXE DE MISE EN ŒUVRE DES POLITIQUES D'ÉDUCATION

L'analyse des politiques publiques a considérablement contribué au renouvellement de la façon de concevoir l'action publique : le jeu des acteurs et des organisations, la diversité des niveaux d'action, des instru-

ments d'action publique, le poids des habitudes et cultures professionnelles, le rôle de l'évaluation expliquent que, même dans les pays aux traditions administratives anciennes, l'action publique ne se limite pas à la mise en œuvre des règles administratives et juridiques [13, 14].

Parmi les travaux qui ont aidé à comprendre le décalage parfois observé entre la politique telle qu'elle est énoncée et sa réalisation auprès de ses bénéficiaires, ceux sur la mise en œuvre occupent une place centrale.

La forme contemporaine des systèmes éducatifs des pays occidentaux remonte au XIX^e siècle [2, 7]. L'approche néo-institutionnaliste a insisté sur le fait que ces systèmes éducatifs sont régis par un ensemble de coutumes, habitudes ou routines, formelles ou informelles, auxquelles se conforment les acteurs. Les sociologues John Meyer et Bryan Rowan [11] sont allés jusqu'à comparer ces règles à des « mythes » que les organisations éducatives adoptent de manière « cérémonieuse », au fur et à mesure qu'elles se développent et que le champ organisationnel auquel elles appartiennent se structure. Les études comparatives ont montré que ces règles sont d'ailleurs relativement partagées entre les pays [1]. Pour Meyer et Rowan [12], l'enseignement s'intéresse moins à

NOTES

3. Article 18 de la loi n° 75-620 du 11 juillet 1975 relative à l'éducation, dite « Loi Haby » (codifié à l'article L. 314-2 du code de l'éducation).

4. Articles 3 et 5 de la loi constitutionnelle n° 2003-276 du 28 mars 2003 relative à l'organisation décentralisée de la République.

5. Article L. 111-1 du code de l'éducation notamment.

ce que les élèves peuvent apprendre qu'aux modalités d'organisation de ces apprentissages : « *L'éducation finit par être définie selon la règle de l'enseignement : l'éducation, c'est un enseignant diplômé qui enseigne une matière standardisée à un élève inscrit dans un établissement agréé* ».

Ces routines et ces règles sont intégrées par l'ensemble des acteurs du système éducatif, créant des effets d'anticipation de ce qui est attendu [9, p. 98]. Elles permettent de simplifier l'activité complexe qu'est l'enseignement. Elles font partie intégrante de l'organisation des établissements scolaires, au point de passer inaperçues. De ce fait, elles sont difficiles à changer et constituent une contrainte des processus de réforme.

L'existence de ces règles ne doit cependant pas faire oublier le rôle des acteurs. La métaphore du jeu d'échecs, citée par le sociologue allemand Norbert Elias [3], est à cet égard, éclairante : les joueurs sont contraints par l'existence de règles déterminant les ressources initiales de chacun et le cadre dans lequel le jeu peut évoluer ; les joueurs intériorisent ces règles qu'ils essaient d'utiliser à leur profit, mais ils ne jouent pas simplement avec les règles. En effet, aux échecs, l'action d'un joueur détermine celle de son adversaire, les joueurs sont donc en situation d'interdépendance. Les interactions entre les acteurs de la mise en œuvre et le public cible de la politique permettent d'adapter la politique aux spécificités locales.

Les recherches consacrées aux effets des réformes éducatives insistent sur les possibilités de réappropriation par les acteurs locaux de politiques impulsées « du haut » (*top-down*) [4, 5, 6, 10, 17]. L'ouvrage de Michael Lipsky, *Street-level bureau-*

cracy, analysait les conditions de travail des « fonctionnaires au guichet », comme les policiers, les juges de proximité et les enseignants. Le terme de « guichet » caractérisait non pas strictement les personnels qui travaillent derrière un guichet, mais les fonctionnaires qui sont en contact régulier avec les usagers des services publics. Contrairement à une vision linéaire (et *top-down*) des politiques publiques, la mise en œuvre est un processus d'« adaptation mutuelle » [15], rendu possible par le pouvoir discrétionnaire dont disposent ceux qui mettent en œuvre la politique. L'existence de règles et de contraintes formelles ne limite pas nécessairement ce pouvoir ; en effet, l'accumulation de ces règles peut être un facteur d'autonomie, car toutes les règles ne pouvant être appliquées, l'agent choisit celles auxquelles il se réfère. Le pouvoir discrétionnaire est une caractéristique des « fonctionnaires au guichet » parce que les situations auxquelles ils sont confrontés sont trop complexes pour pouvoir être résolues selon des procédures déterminées par avance. Cette complexité provient notamment du fait qu'ils s'adressent à des sujets humains et qu'on attend d'eux qu'ils prennent en compte cette dimension humaine dans leur action : « On veut que les enseignants soient sensibles au potentiel de chaque enfant. D'une certaine façon, la société n'attend pas seulement de ses services publics qu'ils soient impartiaux, mais aussi qu'ils montrent de la compassion à l'égard de circonstances particulières et de la souplesse dans leur traitement » [10, p. 15].

La marge de manœuvre dont disposent les acteurs peut être interprétée comme un obstacle à la mise en œuvre

de la politique, mais aussi comme un facteur de succès de cette politique. La mise en œuvre peut être conceptualisée comme une « exploration » au cours de laquelle l'enseignant a la possibilité de rapprocher la politique de son environnement, surmontant ainsi les défauts et les inadéquations de la politique [15].

Lipsky montre cependant que ces processus d'exploration et d'apprentissage dépendent des ressources dont disposent les acteurs. En effet, les organisations en contact avec les usagers sont soumises à un nombre croissant de demandes ; l'apport de ressources supplémentaires est immédiatement absorbé. Cette situation de forte contrainte explique que les « fonctionnaires au guichet » déclarent souvent ne pas faire « parfaitement » leur travail, mais le faire « correctement » compte tenu des conditions dans lesquels ils travaillent. Ce faisant, ils orientent leur activité moins vers un fonctionnement optimal que vers un fonctionnement satisfaisant [10, p. 82].

Pour cela, les acteurs développent des routines permettant de limiter les demandes, de maximiser les ressources dont ils disposent et de faire en sorte que les usagers obéissent aux procédures de l'organisation. Cela se traduit par exemple par une modification de la conception qu'ils ont de leur travail ; ils diminuent leurs objectifs pour les rendre plus compatibles avec les ressources dont ils disposent. Ils peuvent également modifier leur conception des usagers pour que leurs résultats soient plus conformes aux objectifs. Par exemple, un enseignant peut choisir de se concentrer sur un objectif de maintien de la discipline, se spécialiser et se cantonner à son

domaine de spécialité, ou encore porter une attention particulière aux bons élèves, qui correspondent à sa conception idéale du métier, pour que l'écart entre le métier tel qu'il l'imagine et ce qu'il est, soit le plus faible possible. Ces routines rencontrent d'autant moins d'obstacles que les usagers de ces services sont une population captive ; de plus, le cadre de cette interaction met les usagers en situation d'infériorité : les élèves obéissent à l'enseignant, les patients attendent le médecin, les accusés parlent avec respect au juge.

Ces éléments ont amené le politiste américain Richard Elmore à développer le concept de « politique à rebours » [5]. À l'inverse d'une approche « top-down » peu sensible aux interprétations et adaptations inhérentes à la mise en œuvre de la politique, l'approche de la politique à rebours (*backward mapping*) invite les décideurs à anticiper ces adaptations par une « théorie de l'action » détaillant l'ambition du projet, la façon dont le dispositif prévu est à même de répondre aux besoins des acteurs, ou du moins de s'intégrer à l'ensemble de leur activité, voire de la modifier pour mieux atteindre les objectifs assignés à la politique.

La méthode expérimentale est, entre autres, l'illustration de cette politique à rebours : la politique envisagée est testée et observée dans un certain nombre d'écoles et d'établissements. L'observation permet notamment de dégager les freins et leviers de la mise en œuvre, qui seront ensuite constitutifs de la politique. Elle procède par établissement de la preuve (*evidence-based*) et relève d'une démarche pragmatique et incrémentale de l'action publique.

L'EXPÉRIMENTATION : POUR UNE APPROCHE PRAGMATIQUE ET INCRÉMENTALE DU CHANGEMENT

Transposition d'une démarche scientifique dans l'action publique

La transposition des règles de l'expérimentation du champ des sciences fondamentales, aux sciences humaines, puis au champ de l'action publique ne s'est évidemment pas faite sans une phase d'adaptation aux principes et corpus de règles propres à ces domaines. Cependant, dans un cas comme dans l'autre, des principes communs demeurent : l'attachement à la rigueur du déroulement de l'expérimentation et la volonté de tirer des conclusions de portée générale à partir des résultats.

En s'appuyant sur une méthodologie scientifique, les concepteurs de l'expérimentation en définissent les hypothèses, les conditions de réalisations et d'évaluation, et les rendent les plus robustes possible, recourant par exemple à un conseil scientifique, représentant une diversité de points de vue⁶. C'est l'occasion d'adosser l'expérimentation à des travaux de recherche.

Le cahier des charges de l'expérimentation précise les conditions dans lesquelles cette hypothèse est expérimentée : les objectifs visés, le type d'écoles ou d'établissements, le nombre d'élèves, les professionnels concernés par l'expérimentation, le temps qui lui est consacré, les étapes de l'expérimentation et les bilans intermédiaires, les modalités d'accompagnement de l'expérimentation, etc.

Le cahier des charges précise également, dès le début, la nature de l'évaluation de l'expérimentation ainsi que son protocole. L'évaluation peut s'appuyer sur différentes méthodologies selon que l'on cherche à mieux appréhender les conditions de mise en œuvre d'une réforme (méthodologie qualitative) et/ou à mesurer l'impact de l'hypothèse (méthodologie quantitative), l'une n'excluant pas l'autre (*voir les articles de la deuxième partie de ce numéro*). Quelle que soit la méthodologie retenue, elle doit être la plus robuste possible de façon à permettre de tirer des enseignements de l'expérimentation quant à l'efficacité du dispositif expérimenté, sa faisabilité, sa pertinence, ses problèmes de mise en œuvre. Un ensemble d'enseignement en est attendu, dont le facteur commun est qu'il doit éclairer le responsable administratif ou politique, dont la responsabilité est d'initier l'expérimentation et de décider des suites à lui donner.

L'expérimentation relève d'une approche incrémentale et « apprenante » du changement, qui privilégie en particulier la phase de la mise en œuvre : comment les enseignants peuvent-ils s'approprier l'hypothèse expérimentée ? Les corps d'inspection sont-ils outillés pour les accompagner ? Quel type de formation faudrait-il mettre en œuvre ? En quoi cela modifierait-il l'organisation d'une école ou d'un établissement ?

L'enjeu réside, en effet, dans la capacité à extrapoler, à partir de la

NOTE

6. Voir par exemple le comité national d'évaluation, présidé par François Bourguignon, directeur de l'École d'économie de Paris, et constitué pour suivre l'expérimentation du revenu de solidarité active (RSA).

mise en œuvre, leviers et difficultés, principes et modes d'organisation, dans la perspective de leur extension. La dialectique de l'expérimentation se trouve là résumée : fonder l'expérience à partir des pratiques de quelques acteurs locaux ; s'adosser à une base scientifique suffisamment solide pour s'affranchir du contexte local propre à l'expérience conduite et en tirer des enseignements valables à une plus grande échelle.

Avec un objectif : plus d'efficience dans la mise en œuvre

Consubstantiel à l'acte éducatif lui-même, la démarche d'innovation et d'expérimentation apparaît comme une méthode particulièrement adaptée à un contexte de mise sous tension des résultats de l'École et, plus globalement, de demande d'une efficacité accrue de l'action publique. En l'espèce, le contexte est celui d'une École interrogée dans sa capacité à ancrer la réussite de tous les élèves dans la réalité de l'enseignement scolaire, vécu par les élèves, les parents et les enseignants, et à réussir le pari de l'égalité des chances.

La démarche expérimentale permet en effet, par le truchement d'actions construites à petite échelle, d'apporter progressivement des éléments objectifs pour fonder, dans un second temps, une action publique à plus large échelle. Ce faisant, elle peut permettre davantage qu'une action de grande ampleur envisagée sans le recul de l'expérience de tester des solutions nouvelles, c'est-à-dire issues d'une action réflexive des acteurs de terrain ou d'hypothèses formulées par la recherche, et de dépasser, par l'approche pragmatique et volontaire des acteurs, des intuitions ou subjectivités.

On relèvera d'ailleurs que dans les conclusions de son rapport de mai 2010⁷, la Cour des comptes a recommandé par rapport à une réforme décrétée d'en haut, de privilégier « une approche locale des questions à résoudre et (...) l'expérimentation de projets adaptés aux besoins diversifiés des élèves ».

Au-delà de l'École, la réponse par l'innovation s'applique également à d'autres champs de l'action publique dont l'efficience est mise en cause ou, à tout le moins, éprouvée. Il en est ainsi des systèmes de protection sociale que la crise économique tend à la fois à fragiliser, par la pression qu'elle impose sur les États en matière de contrôle de la dépense publique, et à mettre sous tension en accroissant la demande sociale.

Il est, à cet égard, particulièrement significatif que la Présidence polonaise de l'Union européenne ait, en lien avec la Commission européenne, tenu, à la fin du mois de septembre 2011, une conférence dédiée aux réponses innovantes face aux implications sociales de la crise. La problématique de cette conférence tenait précisément dans la nécessaire promotion de l'expérimentation comme moyen, dans un contexte de crise, d'améliorer la qualité et l'efficience économique des politiques sociales.

Répondre à trois limites de l'action publique

Au-delà de la contrainte d'efficience imposée par les attentes de résultat, la démarche expérimentale s'analyse aussi, ainsi que l'observait le Conseil d'État dans son rapport public de 1996 [18], comme la réaction à une triple limite de l'action publique moderne.

La première limite tient au fait que les décisions publiques lorsqu'elles sont perçues comme imposées sans concertation préalable se heurtent à des résistances voire à la défiance de ceux qui doivent les mettre en œuvre. En privilégiant la logique de conviction ou pour reprendre la terminologie anglaise, la logique de « *evidence-based policy* » plutôt que la logique d'adhésion, l'expérimentation s'inscrit dans une perspective pragmatique de l'action publique davantage propice à convaincre.

Dans le champ de l'éducation, l'expérimentation de l'enseignement intégré de science et technologie (EIST) s'apparente à une telle démarche. S'inscrivant résolument dans une perspective de décloisonnement des disciplines et de mise en cohérence de la science, elle offre, depuis 2006, la possibilité d'un autre modèle d'enseignement des sciences, dans le respect des objectifs et des programmes nationaux. Depuis la rentrée 2011, cette expérimentation est conduite dans 105 collèges, dont 38 relèvent du programme Éclair (écoles, collèges, lycées, ambition, innovation, réussite).

La deuxième limite réside dans la relative inadaptation de la norme générale et indifférenciée face à la grande variété de situations. La progressivité d'application induite par la démarche expérimentale apparaît face à cette complexité, comme un facteur d'efficacité de l'action publique.

Reprenant les propos du Doyen Boulouis, le rapporteur de l'assemblée nationale en charge de la loi

NOTE

7. *L'Éducation nationale face à l'objectif de la réussite de tous les élèves*, rapport thématique, mai 2010, p. 171.

organique relative à l'expérimentation notait que l'expérimentation constitue « une procédure de mise au point sur un échantillon » qui, s'intégrant à un processus de réforme, en constitue « la première phase opératoire dont la fonction demeure limitée et dont l'objet est de faire l'économie des inconvénients inhérents à l'application immédiate de mesures généralisées »⁸.

Dans notre champ d'observation, pour répondre à la priorité donnée à la personnalisation des parcours, la logique de l'expérimentation paraît effectivement bien adaptée car elle permet à la fois de laisser la place à un vaste champ d'initiatives locales et, par des bilans d'étape, de s'assurer que toutes les données du problème et toute la complexité des situations individuelles ont bien été prises en compte.

De la même manière, le « livret de compétences expérimental » prévu par l'article 11 de la loi du 24 novembre 2009, a précisément eu recours à cette démarche expérimentale en partant de l'analyse que tous les paramètres d'une véritable prise en compte par l'École des aptitudes et habilités acquises hors du cadre scolaire, n'étaient pas encore réunis. En initiant cette démarche à l'échelle locale autour de partenariats construits entre les établissements scolaires et leurs partenaires, le législateur a confié aux acteurs locaux le soin d'apporter la preuve de l'opportunité de cette évolution, initiée par le socle commun, et d'en déterminer les modalités (contenu de livret, méthode de reconnaissance des engagements et d'acquis non scolaires, modalités d'accompagnement des élèves et appropriation du livret par les élèves, les parents, les équipes et les conseils de classe).

On notera que cette orientation générale pour une démarche progressive ne signifie pas une absence de norme ou de toute intervention du niveau national : elle la renouvelle, conduit à privilégier la démarche de responsabilisation et d'autonomie accompagnée. À la norme précisée se substitue la définition d'un cadre général, l'accompagnement (par des guides, des conseils, de la formation), la production d'outils et de synthèse des « bonnes pratiques » et la diffusion de « ce qui marche ».

Enfin, la troisième limite réside dans le fait que la norme nationale unique paraît parfois incertaine pour prendre en compte toutes les potentialités offertes par les technologies, notamment dans le champ de l'information et de la communication. Le principe de l'expérimentation, que l'on retrouve également sous le terme de « démonstrateurs » employé dans le récent appel à projets du programme d'investissements d'avenir⁹, permet quant à lui de tester les potentialités offertes par les technologies, d'en apprécier les usages possibles au service de la réussite des élèves et d'apporter ensuite, sur la base de l'expérience, des recommandations pour un usage généralisé.

Au lieu d'être décrétées, les potentialités du numérique doivent être mesurées à l'aune des usages. Qu'il s'agisse des manuels *via* les environnements numériques de travail (ENT) ou des tablettes, l'ère du numérique et des réseaux, ainsi que la rapidité des évolutions techniques induisent des modifications profondes dans l'enseignement et les apprentissages, rendent les choix complexes, et au final, commandent des approches à petite échelle avant d'engager des généralisations qui peuvent être coûteuses.

Une inscription dans le droit

Depuis 2005, cette démarche est inscrite dans le cadre d'action du système éducatif. Au plan juridique, le droit à l'expérimentation a été consacré par la loi d'orientation et de programme pour l'avenir de l'École du 23 avril 2005, article 34 (*encadré*).

Cet article décline dans le code de l'éducation l'article 37-1 de la Constitution qui prévoit la faculté pour le législateur et le pouvoir réglementaire de prévoir, pour un objet et une durée déterminés, des dispositions à caractère expérimental.

Le Conseil constitutionnel ayant jugé que les dispositions de l'article 34 avaient un caractère réglementaire, cet article permet la mise en œuvre d'expérimentations qui ne dérogent pas à la loi. Ainsi, s'il n'est pas possible de déroger au contenu des programmes de caractère national, l'article autorise notamment des modulations dans les grilles horaires des enseignements, l'organisation pédagogique de la classe ou de l'établissement.

Pour autant, en pratique, l'expérimentation ne se limite pas au seul

NOTES

8. J. Boulouis « Note sur l'utilisation de la "méthode expérimentale" en matière de réformes », mélanges Trotabas, LGDJ, 1970, cité par Michel Piron dans le rapport n° 955, au nom de la commission des lois constitutionnelles, de la législation et de l'administration générale de la République sur le projet de loi organique, (n° 855), relatif à l'expérimentation par les collectivités territoriales, Assemblée nationale, juin 2003.

9. Arrêté du 26 décembre 2011 relatif à l'approbation du cahier des charges « Appel à projets e-éducation n° 2 – services numériques innovants pour l'e-éducation ».

Loi d'orientation et de programme pour l'avenir de l'École du 23 avril 2005

Article 34 (codifié L 401-1)

« Dans chaque école et établissement d'enseignement scolaire public, un projet d'école ou d'établissement est élaboré avec les représentants de la communauté éducative. Le projet est adopté, pour une durée comprise entre trois et cinq ans, par le conseil d'école ou le conseil d'administration, sur proposition de l'équipe pédagogique de l'école ou du conseil pédagogique de l'établissement pour ce qui concerne sa partie pédagogique.

Le projet d'école ou d'établissement définit les modalités particulières de mise en œuvre des objectifs et des programmes nationaux et précise les activités scolaires et périscolaires qui y concourent. Il précise les voies et moyens qui sont mis en œuvre pour assurer la réussite de tous les élèves et pour associer les parents à cette fin. Il détermine également les modalités d'évaluation des résultats atteints.

Sous réserve de l'autorisation préalable des autorités académiques, le projet d'école ou d'établissement peut prévoir la réalisation d'expérimentations, pour une durée maximum de cinq ans, portant sur 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 ou le jumelage avec des établissements étrangers d'enseignement scolaire. Ces expérimentations font l'objet d'une évaluation annuelle.

Le Haut conseil de l'éducation établit chaque année un bilan des expérimentations menées en application du présent article. »

champ de l'article 34. Dans le langage usuel et de communication, « l'expérimentation » prend un sens beaucoup plus souple, n'incluant pas nécessairement de dérogation, en tout cas explicite; elle incarne une démarche collective de changement, d'une certaine ampleur, structurée par des objectifs, des indicateurs, la préoccupation de l'évaluation et au-delà, par un arbitrage à effectuer sur l'intérêt ou non de modifier la norme et d'établir de nouvelles règles communes.

Des dispositions organisationnelles

En complément de cette reconnaissance juridique, l'expérimentation est également une réalité institutionnelle matérialisée par la création du département recherche-développement, innovation et expérimentation (DRDIE) au sein de la direction générale de l'enseignement scolaire (DGESCO) et la constitution d'un réseau d'acteurs de l'innovation, les conseillers académiques recherche-développement, innovation et expérimentation (CARDIE).

Conseillers techniques du recteur, ils ont des fonctions de veille de l'in-

novation et sont donc à l'origine et en soutien des formulations d'hypothèses expérimentales, de mise en œuvre et d'évaluation d'expérimentations, qu'elles aient un ancrage local ou académique ou qu'elles relèvent d'une impulsion nationale. Ils assurent également un lien structurel avec la recherche, puisqu'ils sont notamment les correspondants privilégiés de l'Institut français de l'éducation de l'école normale supérieure, de Lyon (Ifé-ENSL), qui a pris la suite de l'Institut national de la recherche pédagogique (INRP).

Le réseau des CARDIE fait l'objet d'une animation nationale (lettre d'information, séminaire professionnel, réseau RESPIRE (<http://respire-education.fr>). Les « Journées de l'innovation » qui se sont tenues les 31 mai et 1^{er} juin 2011, et dont la seconde édition se tiendra les 28 et 29 mars 2012 rendent visible ce réseau d'acteurs et sa valeur ajoutée pour le système éducatif.

Ce réseau a développé un certain nombre d'outils avec l'objectif de renforcer la dimension scientifique de la démarche expérimentale en éducation. Un vade-mecum « Innover pour une école des réussites » apporte

des outils et méthodes aux acteurs de l'innovation tandis que la base de données des expérimentations, Expérithèque¹⁰, recense les actions innovantes dans une perspective de mutualisation. Elle est adossée à un outil d'auto-évaluation centrée sur l'état des pratiques de l'établissement dans cinq domaines stratégiques : le pilotage éducatif et pédagogique, la priorité donnée aux apprentissages fondamentaux, les modalités d'évaluation des élèves, le climat scolaire, et l'ambition scolaire. Présentés de façon synthétique, ces cinq piliers sont identifiés par de nombreux travaux de recherche comme contribuant à l'efficacité de l'enseignement [8, 16].

Une collection de cartes mentales permet de détailler les gestes professionnels qui participent de l'innovation, de l'évaluation des élèves, des usages pédagogiques du numérique¹¹....

NOTES

10. www.eduscol.education.fr/expirithèque

11. Voir notamment www.eduscol.education.fr/innover

Enfin, une nouvelle collection du Centre national de documentation pédagogique se centre sur un exposé des facteurs de réussite en matière éducative, tant à partir d'un retour d'expérience d'équipes éducatives qu'à partir de revues de la littérature scientifique.

DES EXPÉRIMENTATIONS ÉGALEMENT DÉTERMINÉES PAR L'ENVIRONNEMENT DE L'ACTION PUBLIQUE

Plusieurs exemples d'expérimentation d'intérêt national permettent d'illustrer la manière dont l'initiation et le devenir des expérimentations sont également déterminés par le contexte de l'action publique.

Si l'expérimentation est un acte d'engagement et de conviction qui impose un choix à ses décideurs lorsqu'elle prend fin, on ne saurait résumer ledit choix à une simple alternative binaire : généraliser ou abandonner l'action conduite. La légitimité de la décision politique sur les enjeux éducatifs par delà les réalités statistiques, la prise en compte de l'autonomie des établissements ou encore de la complexité croissante des processus de réforme, pèsent sur les choix opérés au sortir des expérimentations. Loin d'invalider la démarche, cette situation oblige les acteurs de l'expérimentation à l'anticiper, à imaginer des scénarios multiples pour les suites des expérimentations.

De l'innovation à l'expérimentation

L'interrogation régulière, si ce n'est permanente, sur les frontières entre innovation et expérimentation, est légitime puisqu'elle conduit à préciser le cadre d'action des équipes locales.

Elle n'en constitue pas moins parfois un obstacle inutile à l'encouragement à la dynamique d'expérimentation et d'innovation. En guise de repère et sans valeur de doctrine, on désignera par expérimentations, les projets qui dérogent à une norme juridique et se fondent sur l'article 34 de la loi du 23 avril 2005, ainsi que ceux qui, sans déroger à une telle norme, s'inscrivent dans un dispositif structuré et construit à une échelle académique ou nationale qui justifie l'inscription dans cette catégorie et le rattachement à l'article 34 ; les innovations sont définies *a contrario*.

Cette distinction peut prendre la forme d'un prolongement lorsque les actions innovantes servent d'appui à une expérimentation. Ainsi, l'expérimentation « Cours le matin, sport l'après-midi », conduite dans 124 établissements à la rentrée 2010, tire ses origines d'une action innovante dans un établissement. La conduite de l'expérimentation permet de tester à plus grande échelle les freins et leviers de la mise en œuvre de ce type de politique. L'évaluation a montré que, grâce à cette expérimentation, un nombre plus important d'élèves fréquentait l'infirmerie, non pour des problèmes de santé, mais pour un échange avec l'infirmière sur leur santé, l'hygiène et le bien-être. Par ailleurs, peu nombreux sont les chefs d'établissement à pointer la difficulté de concentrer les apprentissages fondamentaux le matin et les activités sportives et culturelles l'après-midi, tout en doublant la pratique sportive pour les élèves.

De l'expérimentation à l'essai

L'expérimentation de la « Mallette des parents » est emblématique

d'une expérimentation académique, conduite de façon scientifique, et reprise nationalement, à grande échelle, sans pour autant devenir une norme obligatoire pour tous les établissements. Initiée dans l'académie de Créteil à partir de septembre 2008, elle est, dès le début, couplée à une évaluation par assignation aléatoire, réalisée par l'École d'économie de Paris. À partir de l'idée de coéducation, les parents d'élèves de sixième sont invités à venir échanger avec les enseignants autour de sujets centraux pour la scolarité d'un élève de sixième. Le contenu de ces échanges est cadré par le conseil scientifique de l'expérimentation.

La publication du rapport d'évaluation, en janvier 2010, est décisive pour la sortie d'expérimentation qui prend deux formes complémentaires :

- sur la base dudit rapport, une extension de ce principe d'action à 1 300 collègues à la rentrée 2010 est décidée en retenant le principe de l'inscription dans le cadre de l'autonomie des établissements. Le dispositif « mallette des parents » relève de la politique de l'établissement, contractualisée avec l'académie, et pour laquelle l'échelon national apporte une méthode et des outils « clés en main », validés par l'évaluation ;

- en écho aux échanges des équipes participant à l'expérience de l'académie de Créteil, il est décidé d'engager, sur les mêmes bases méthodologiques et scientifiques, une deuxième expérimentation « mallette des parents » autour de la problématique de l'orientation *post* troisième, dans l'académie de Versailles. Elle est également évaluée selon une méthodologie similaire (randomisation).

Ce faisant, cette expérimentation constitue une sorte de modèle « parfait » : le temps laissé à l'expérimentation et à l'évaluation a permis que des décisions de niveau national soient prises, intégrant une extension à la fois horizontale mais aussi verticale.

On notera que l'essai peut également consister à étendre l'expérimentation elle-même, permettant d'en amplifier le champ, moyennant une évaluation également étendue. Tel est le cas de l'expérimentation « Cours le matin, sport l'après-midi », pour laquelle le nombre d'établissements impliqués a été accru à la rentrée 2011 (de 141 à 209).

De l'expérimentation à l'anticipation de la généralisation

Premier cas concret d'expérimentation législative mise en œuvre par les collectivités territoriales et se revendiquant d'une approche scientifique, notamment par son mode d'évaluation externe, l'expérimentation du revenu de solidarité active a pu laisser chez certains le sentiment d'inachevé, la loi de généralisation intervenant avant que les résultats définitifs de l'évaluation ne soient connus¹².

Dans le champ éducatif, un exemple d'une telle pratique nous est donnée par le dispositif de prise en charge des jeunes sortis sans diplôme du système de formation initiale pour lequel la circulaire interministérielle du 22 juillet 2009 avait initié, dans une douzaine d'académies, une phase d'expérimentations – dont la plupart s'achèvent en fin d'année 2011 – tandis que le plan *Agir pour la jeunesse* annoncé à la fin de septembre 2009 – et décliné dans la loi, puis dans la circulaire du 9 février 2011 – en généralisait le principe pour l'ensemble du territoire.

En l'espèce, la décision de généraliser le principe d'action retenu pour les expérimentations n'a pas mis en cause les exigences méthodologiques destinées à garantir la validité interne de chacune des expérimentations. Chacune de ces évaluations est d'ailleurs utilisée par les acteurs locaux eux-mêmes pour bâtir le dispositif généralisé et les enseignements seront diffusés à tous.

Ce faisant, cette expérience illustre également une autre réalité, la difficulté de combiner parfois la logique de l'expérimentation, qui suppose une démarche volontaire et différenciée, et la conduite de politiques publiques pour lesquelles il est important et, dans certains cas, indispensable, que l'ensemble des acteurs locaux soient mobilisés en même temps.

CONCLUSION : LES BÉNÉFICES DE L'EXPÉRIMENTATION

Dans de nombreux pays, l'innovation dans le domaine éducatif se focalise sur les technologies de l'information et de la communication. La démarche présentée ici est d'une autre nature. Elle part du principe que le numérique n'est pas, par essence, innovant ; il l'est lorsqu'il implique un changement de pratiques pédagogiques profond chez l'enseignant.

Surtout, la démarche présentée ici se centre moins sur l'innovation que sur l'expérimentation comme démarche scientifique dans le domaine de l'action publique.

L'un de nos enjeux est de penser l'innovation et le changement de façon systémique : comment les pratiques des enseignants changent-elles ? Pourquoi

certaines écoles ou établissements font-ils mieux réussir leurs élèves que d'autres, à cadre institutionnel constant ? La démarche d'expérimentation est-elle une stratégie explicite pour répondre au diagnostic du contrat d'objectif ou du projet académique ?

Même si l'expérimentation dans le secteur éducatif n'est pas récente, sa pratique peut encore être poussée plus avant, compte tenu du potentiel qu'elle représente pour le système éducatif du point de vue du changement des pratiques, du changement systémique, de sa capacité à renseigner sur la façon de mettre en œuvre les politiques d'éducation et de contribuer à une amélioration des résultats. De ce fait, elle n'est pas l'affaire de spécialistes, mais s'affirme comme un instrument d'action publique qui doit être au cœur des enjeux éducatifs et au cœur de l'action des enseignants, car l'expérimentation est d'autant plus pertinente qu'elle ne se cantonne pas aux marges, et à l'action des enseignants, par le potentiel d'expression de la créativité et de développement professionnel qu'elle porte en elle.

Pour finir, comme technique ou instrument d'action publique, l'expérimentation n'a pas valeur de décision. Elle ne saurait par exemple se substituer au débat public, selon une approche qui pourrait être qualifiée de « positiviste ». Cependant, l'expérimentation renseigne de la façon la plus factuelle et objective possible sur un processus éducatif. En cela, elle constitue une donnée précieuse pour la décision. ■

NOTE

12. Virginie Massieu, *L'expérimentation comme exemple de transposition d'une méthode scientifique hors de son cadre d'origine*, 2011.

BIBLIOGRAPHIE

- [1] **Alexander, R. J.** (2003), *Culture and pedagogy international comparisons in primary education*. Oxford : Blackwell.
- [2] **Archer, M. S.** (1979), *Social origins of educational systems*. London : Sage.
- [3] **Elias, N.** (1991), *Qu'est-ce que la sociologie?* (Y. Hoffmann, Trans.). La Tour-d'Aigues : Ed. de l'Aube.
- [4] **Elmore, R. F.** (1976), *Follow Through : decisionmaking in a large-scale experiment*. Thèse de doctorat, Harvard University Graduate School of Education, Cambridge.
- [5] **Elmore, R. F.** (1980), Backward mapping: implementation research and policy decisions. *Political science quarterly*, 94(4), 601-616.
- [6] **Elmore, R. F.** (2004), *School reform from the inside out: policy, practice, and performance*. Cambridge : Harvard Education Press.
- [7] **Green, A.** (1990), *Education and state formation: the rise of education systems in England, France, and the USA*. New York : St. Martin's Press.
- [8] **Hattie, J.** (2008), *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*: Routledge.
- [9] **Lascoumes, P., & Le Galès, P.** (2007), *Sociologie de l'action publique*, A.Colin, Paris.
- [10] **Lipsky, M.** (1980), *Street-level bureaucracy: dilemmas of the individual in public services*. New York : Russell Sage Foundation.
- [11] **Meyer, J. W., & Rowan, B.** (1977), Institutionalized organizations: formal structure as myth and ceremony. *The American journal of sociology*, 83(2), 340-363.
- [12] **Meyer, J. W., & Rowan, B.** (1978), The structure of educational organizations. In M. W. Meyer (Ed.), *Environments and organizations* (1^{re} ed., pp. 78-109), Jossey-Bass, San Francisco.
- [13] **Muller, P., & Surel, Y.** (1998). *L'analyse des politiques publiques*, Montchrestien, Paris.
- [14] **Parsons, D. W.** (1995). *Public policy: an introduction to the theory and practice of policy analysis*, Edward Elgar, Aldershot.
- [15] **Pressman, J. L., & Wildavsky, A. B.** (1984), *Implementation : how great expectations in Washington are dashed in Oakland: or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes*(3^e ed.), University of California Press, Berkeley.
- [16] **Scheerens, J.** (2000). *Improving school effectiveness*, Paris, Unesco International Institute for Educational Planning.
- [17] **Tyack, D. B., & Cuban, L.** (1995). *Tinkering toward utopia: a century of public school reform*, Harvard University Press, Cambridge.
- [18] **Conseil d'État** (1996). *Sur le principe d'égalité*, Paris, La Documentation française.

L'expérimentation *bottom up*, l'exemple ultra-marin de l'académie de la Réunion

Pascal Chabernaud

CARDIE, académie de la Réunion

Pour être encore plus innovantes, et surtout mutualisées et diffusées, les initiatives des équipes locales doivent être accompagnées par une politique académique. L'exemple de l'académie de la Réunion montre comment ce pilotage s'est structuré depuis la loi organique relative aux lois de finances et la loi d'orientation et de programme pour l'avenir de l'École du 23 avril 2005. La politique de contractualisation mise en œuvre dans ce cadre constitue une modalité structurante de repérage et d'essaimage des expérimentations. L'article essaie également de tirer des enseignements de ces années de fonctionnement et envisage des améliorations possibles.

Dans l'académie de la Réunion, le pilotage des démarches dites de projets et de suivi d'actions particulières (dont les actions innovantes) a été infléchi par le passage d'une culture de moyens à une culture de résultats, porté par la loi organique relative aux lois de finances (LOLF) ; cette loi demandait en effet que des objectifs soient définis, qu'un engagement sur des résultats soient pris, qu'une appréciation des résultats obtenus soit réalisée. L'article 36 de la loi d'orientation et de programme pour l'avenir de l'École de 2005, en instituant un cadre légal pour la contractualisation entre chaque établissement public local d'enseignement (EPL) et l'autorité académique, a permis d'installer l'innovation et l'expérimentation comme des démarches transversales et structurelles de ces

contrats¹. L'incidence de ce changement dans le pilotage académique est au cœur de cette contribution.

AVANT 2006 : UN PILOTAGE PAR LES MOYENS PEU EFFICIENT

Un pilotage par les moyens principalement concentré sur le second degré

Jusqu'en 2006, le pilotage académique de l'innovation était essentiellement un pilotage par les moyens. Dans le premier degré, des moyens financiers pouvaient être consentis à des actions inscrites dans le champ artistique et culturel ou dans celui des technologies numériques. Ces actions trouvaient leur place dans des projets d'école dont la visibilité était possible au niveau de la circonscription. C'est cette dernière qui disposait d'une enveloppe de moyens complémentaires pour son projet de circonscription ; elle avait en charge de répartir les moyens accordés. Dans le second degré, par contre, (article L. 421-4 du code de l'éducation) un nombre important d'heures supplémentaires étaient déléguées par l'académie, souvent accompa-

NOTE

1. L'article 36 prévoit ainsi : « Le contrat d'objectifs conclu avec l'autorité académique définit les objectifs à atteindre par l'établissement pour satisfaire aux orientations nationales et académiques et mentionne les indicateurs qui permettront d'apprécier la réalisation de ces objectifs. Le contrat d'objectifs constitue un outil de dialogue entre les autorités académiques et le chef d'établissement, et l'occasion pour celui-ci de mettre en exergue les caractéristiques propres à son établissement ».

gnées de crédits pour des actions diverses d'ordre pédagogique. Ces « moyens dédiés » accompagnaient la dotation horaire globale (DHG). Ils donnaient lieu à un appel à projets au niveau académique avec un pré-cadrage tel que « action pédagogique-innovation », « comité d'éducation à la santé et à la citoyenneté (CESC) », « mission générale d'insertion (MGI) », « Culture », « Tice », « ouverture à l'international », « relations avec le projet académique », etc. En 2006, ces moyens dédiés correspondaient à une enveloppe globale de 207 330 euros, de 48 053 heures supplémentaires effectives (HSE) et de 1 823 indemnités pour activités péri-éducatives pour le second degré. Cette même année, 2 165 actions ont été proposées et 524 financées.

De plus, dans le second degré, les actions innovantes pouvaient relever d'un certain mode de contractualisation dans la répartition de la DHG, sous la forme d'une « part contractuelle » sensée être consacrée par les EPLE à la mise en œuvre d'une action prioritaire de leur projet d'établissement, et cela sur quatre années. Cette priorité était identifiée à la suite d'un diagnostic.

En 2006 toujours, 67 % des établissements du second degré s'inscrivaient dans une telle contractualisation. Or, la part contractuelle pesait au niveau académique 1,7 % de la DHG, soit une proportion très faible. C'est peut-être ce qui explique que, malgré les précautions prises (contrat établi pour une durée de quatre années, diagnostic posé, analyse attendue, démarche réflexive, pratiques questionnées), cette approche de la contractualisation par un soutien particulier ne s'est pas révélée être un levier suffisant. Les enquêtes menées

à la fin du dispositif de contractualisation montraient ainsi que la priorité de départ était le plus souvent oubliée et rarement évaluée dans la durée.

Un mécanisme qui, de fait, concentre l'attention sur le « hors-la-classe »

Ce mode de fonctionnement, vu *a posteriori*, nous montre que, dans un souci de repérer les actions mises en œuvre et de les aider, l'académie déchargeait en fait les structures de leurs responsabilités : priorisation et choix des actions à mener en interne, objectifs à atteindre, etc. L'attention était focalisée sur les actions périphériques du projet d'établissement ou d'école, et non sur les pratiques enseignantes mobilisées dans la quotidienneté.

Le financement « supplémentaire » donnait une certaine lisibilité des projets envisagés par les structures mais au travers d'un miroir déformant. En effet, rien n'indiquait qu'un collègue qui ne s'inscrivait pas dans cette démarche ne mettait pas pour autant en œuvre un vrai projet d'établissement riche en implications personnelles ; de même, de nombreuses actions présentées n'étaient pas nécessairement le signe de l'engagement d'une communauté éducative.

Par ailleurs, disposer de la description de 2 165 actions était une richesse relativement sous-exploitée car il était rare d'avoir un suivi de l'utilisation des moyens attribués et tout aussi rare d'obtenir des informations en retour sur les bénéfices apportés aux élèves ainsi que sur les analyses faites dans un contexte d'auto-évaluation. Des analyses restaient possibles au travers des applications informatiques de type gestion de base de données ; elles étaient globales, comparatives, thématiques.

Ainsi, au niveau académique, il y avait :

- une vraie difficulté à prendre en compte les effets des actions retenues et des actions poursuivies localement sans soutien financier fléché. Au mieux, l'action est évaluée positivement ou négativement suivant qu'elle se soit tenue ou non ;

- un fonctionnement par domaines et enveloppes, dissociant plutôt que fédérant.

Pour les EPLE et de façon moindre pour les écoles, on pouvait faire quatre constats :

- une absence de formalisation des pratiques pédagogiques pour la plupart déployées en marge de la classe ;

- un manque de responsabilisation dans la définition des objectifs et une faible incitation à l'exercer ;

- une multiplication de micro-actions rarement mises en lien avec une dynamique plus globale ;

- une corrélation entre « supplémentaire » et « prioritaire ».

L'EXPÉRIMENTATION COMME ACCOMPAGNEMENT AU CHANGEMENT DU CADRE... ET DES CADRES

Ce dispositif de contractualisation par les moyens arrivant à son terme à la rentrée scolaire 2007-2008, se posait alors le défi de la nouvelle contractualisation à engager, définie par le concept de « contrat d'objectifs ».

Du nouveau cadre institutionnel à une politique académique assumée

En 2007, la contractualisation par les résultats, telle que posée dans la

loi de 2005 devait être généralisée à l'ensemble des lycées et collèges de l'académie en phase avec le nouveau Projet stratégique académique 2007-2011.

La troisième priorité de ce projet était de renforcer le pilotage académique, avec trois objectifs :

1) favoriser la démarche de contractualisation, conforter l'autonomie des EPLE et optimiser le fonctionnement des bassins d'éducation ;

2) développer une culture commune de l'encadrement ;

3) assurer une meilleure gestion des ressources humaines.

Les expérimentations sont à la croisée de ces trois objectifs et l'essai-mage de la pratique expérimentale est fonction de l'évolution de chacun d'eux.

Le pilotage académique des expérimentations et de la contractualisation est assuré par le conseiller académique en recherche, développement, innovation, expérimentations (CARDIE) et englobe l'ensemble des démarches expérimentales : celles typiquement décrites dans l'article 34, mais aussi d'autres, d'ordre national, académique ou de circonscription. Ces trois derniers cas sont bien représentés. Le pilotage se fait en continu car, en dehors du repérage au travers d'un appel à candidatures, il comprend aussi les démarches de sensibilisation, d'encouragement, d'écriture des demandes, de suivi des projets, de rencontre des enseignants et d'accompagnement des équipes à l'auto-évaluation et à la modélisation.

La démarche de l'académie consiste à mettre en place un appel à projets qui se déroule en deux phases en fonction du calendrier académique. Le premier retour, début décembre, permet de faire connaître à l'auto-évaluation des projets mis en

œuvre dans le cadre de l'autonomie des structures (habituellement non dérogoires). Le deuxième retour, fin mai, est destiné à autoriser des projets pouvant se révéler dérogoires en amont de la rentrée scolaire suivante. Ces derniers sont cependant relativement rares. Le comité de pilotage de la performance sous l'autorité de l'inspecteur d'académie adjoint au recteur assure la lecture des demandes accompagnées des remarques des corps d'inspection. Cette gestion administrative des projets s'inscrit généralement en aval d'un premier contact avec les équipes.

Les outils de la contractualisation et du pilotage partagé

La mise en œuvre au plan académique d'une démarche de contractualisation par les résultats a notamment permis de renforcer le pilotage académique du premier degré, en particulier dans le domaine des démarches innovantes. Ce pilotage académique s'est appuyé d'abord sur la mise en place de contrats de performance des circonscriptions. Ensuite, la généralisation de l'accompagnement éducatif aux écoles avec conjointement une responsabilisation des directeurs par la mise à disposition d'un outil informatique de mise en paiement décentralisée des heures a joué un rôle important.

Surtout, des outils numériques ont été conçus pour permettre un pilotage partagé dans le cadre de la démarche de contractualisation : il s'agit de l'environnement numérique de travail « *Cellbureau* », proposé au niveau académique, et son module dédié aux projets d'écoles. Les indicateurs proposés dans ce cadre ont évolué ;

l'application offre la possibilité de fixer des cibles et, plus récemment, de décrire les actions menées et de les partager. Ce module est en lien direct avec l'application ÉVASION qui permet de gérer en ligne toutes les étapes des évaluations nationales ou académiques dans le premier et le second degré, ce qui permet une construction dynamique de certains indicateurs du projet d'école.

En 2011, la contractualisation s'est généralisée aux écoles en phase avec le renouvellement des contrats d'objectifs des collèges et la mise en place du programme « écoles, collèges et lycées pour l'ambition, l'innovation et la réussite » (ÉCLAIR). Les nombreuses formations organisées à destination des directeurs d'école abordent le champ de l'expérimentation. L'opportunité induite par la mise en place du socle commun de connaissances et de compétences permet le développement d'expérimentations pour favoriser la fluidité des parcours de réussite et aller vers l'École du socle.

Le recours aux nouvelles technologies vise à rendre service à l'ensemble des acteurs concernés par le pilotage pédagogique partagé et il est important de questionner les utilisateurs, de les impliquer dans les possibilités offertes ou à offrir.

L'environnement « projet d'école » permet désormais de construire un contrat d'objectifs d'école. Il intègre un module « ExplorActions » proposant le partage de toutes les fiches actions déposées au niveau académique et leur appropriation (5 175 actuellement). Ce module est aussi un outil de consolidation et d'enquête car il permet de tester la cohérence des fiches vis-à-vis des indicateurs retenus et de questionner par mot-clé ou par thème

tout en produisant des analyses statistiques. Le dispositif « expérimentation pédagogique » est proposé depuis peu. Les expérimentations seront ainsi facilement identifiables au travers du module « ExplorActions » et le résultat est déjà mesurable. Le « rendre-compte » est favorisé par la possibilité de déposer des documents propres à une action.

Dans le second degré, tous les EPLE ont été accompagnés par le service statistique académique dans leur démarche d'écriture de leurs contrats d'objectifs, puis reçus au niveau académique pour un dialogue. Une formation s'est tenue au sein de tous les bassins d'éducation et de formation pour faire la promotion du droit à expérimenter auprès des chefs d'établissements, d'enseignants des conseils pédagogiques, des inspecteurs de l'Éducation nationale (IEN) de circonscription.

Depuis 2010, un espace propre au suivi des contrats d'objectifs des EPLE est opérationnel et en cours d'évolution.

La notion de pilotage partagé est importante car elle demande que chaque niveau de responsabilité, chaque entité, assume son propre cheminement et sa stratégie déclinée en objectifs, en prenant appui sur son analyse. De plus, il faut un engagement dans une démarche collaborative de partage de données afin que chacun puisse se retrouver comme « acteur-auteur ». Selon Michel Blondel (2011) « *Remonter de l'information prédictive est la contrepartie de l'autonomie accordée sur le terrain, dans une logique de réciprocité. Elle implique la mise en œuvre d'un véritable travail de poupées gigognes, qui, par agrégats successifs, permet de donner à chaque niveau de décision*

les informations utiles à la décision, et de transmettre plus d'avis sur les opportunités et risques anticipés que sur des chiffres. Au-delà du chiffre, c'est au commentaire que les tableaux de bord doivent laisser la place ! »

Une dynamique d'apprentissage institutionnel et organisationnel

Venant compléter le dispositif de contractualisation, les rencontres pédagogiques académiques, mises en place à compter de la rentrée 2008, ont vocation à permettre de mieux apprécier, à partir de l'analyse des résultats des élèves, les modalités d'action mises en œuvre pour réaliser les objectifs définis. Elles constituent également pour les établissements et les responsables académiques l'opportunité de repérer et de diffuser les stratégies de réussite. Organisée sur la base d'un protocole simple, la rencontre s'appuie sur une analyse thématique. Elle a lieu dans les établissements et les circonscriptions. Elle implique la direction et le conseil pédagogique dans les établissements, l'inspecteur et les directeurs d'écoles dans les circonscriptions, le comité exécutif dans les réseaux d'éducation prioritaire. Elle est structurée autour de trois axes : l'axe didactique, l'axe vie et parcours scolaire et l'axe administratif et partenarial. Six thèmes sont abordés : trois d'entre eux sont choisis par l'établissement, la circonscription ou le réseau d'éducation prioritaire.

Après trois années de mise en œuvre, le constat porté sur les rencontres pédagogiques académiques est très positif. Au rythme d'une quarantaine de rendez-vous par année

scolaire, les rencontres pédagogiques académiques ont permis aux différentes parties impliquées de s'enrichir mutuellement, de donner plus de sens à la contractualisation, de placer chacun dans ses missions. De plus, elles ont vraiment joué leur rôle de moment privilégié pour le repérage et l'essaimage des expérimentations. Le CARDIE y est systématiquement représenté, comme lors des phases de dialogue pour l'élaboration des contrats d'objectifs.

En 2010-2011, les premières rencontres avec des écoles se sont tenues. Ici encore, les équipes impliquées, et particulièrement les directeurs, ont fait montre de leur engagement et de leur capacité à innover en situation.

La généralisation des contrats d'objectifs aux écoles et le renouvellement de tous les contrats des collèges sont aussi une opportunité pour parler de l'établissement formateur en référence à la notion d'organisation apprenante : « *Organisation au sein de laquelle les agents améliorent en continu leur capacité d'atteindre les résultats qu'ils souhaitent, où des modes de pensée innovants et l'esprit d'ouverture sont encouragés, où les aspirations collectives peuvent s'épanouir, et où les agents sont en permanence en situation d'apprentissage dans le cadre de l'organisation tout entière.* »²

Si encourager des équipes à s'investir dans des expérimentations ou repérer des projets participe de cette valorisation, il reste encore à mieux questionner toutes les compétences engagées dans les actions.

NOTE

2. http://www.eipa.nl/files/File/CAF/Brochure2006/CAF_Brochure_2006_2008_A4_FRENCH.pdf

SE FORMER ET INNOVER : L'EXPÉRIMENTATION AU SERVICE D'UN DÉVELOPPEMENT PROFESSIONNEL DES PERSONNELS

Deux actions peuvent nous sensibiliser à la dimension complexe et sys-

témique des expérimentations, afin de relever les questions qu'elles posent à l'institution dans un changement continu et, à présent, rapide.

L'exemple du dispositif *méthode au Long Cours 6^e*

La première expérimentation présentée ici s'inscrit dans un champ

insuffisamment exploré au niveau académique, celui des disciplines (cf. encadré ci-dessous).

L'exemple du premier encadré nous renseigne sur plusieurs points en matière de conduite de l'expérimentation pédagogique :

– il n'y a pas de génération spontanée en expérimentation : autant dans sa genèse, longue, pour « fabriquer »

Méthode au Long Cours 6^e

En 1987, un jeune professeur de lettres modernes et linguiste de formation, Lionel Tinlot, est frappé par le décalage existant entre les grammaires scolaires et la linguistique enseignée à l'université. Tout se passe comme si cette dernière n'avait aucune incidence sur la première. Il entreprend donc d'introduire dans ses cours, en collège, quelques tests utilisés par les linguistes, en particulier le test de commutation.

En 1997, les nouveaux programmes instaurent l'enseignement par séquence pour travailler la grammaire à partir de textes narratifs, descriptifs, explicatifs ou argumentatifs. Lionel Tinlot décide de produire son propre discours, discours dans lequel il place les constructions syntaxiques permettant de conserver une approche rigoureuse et progressive de la phrase : ainsi naît *Le Long Cours de 6^e*, un discours narratif (un conte) débutant par des phrases simples, des temps simples, un vocabulaire simple, et finissant par des phrases complexes, des temps composés, un vocabulaire et une syntaxe enrichis.

La mise en place du socle commun de connaissances et de compétences en 2006 et son accompagnement par l'inspection de lettres pousse l'enseignant à élaborer un référentiel des compétences minimales au collège. La méthode plus formalisée est alors présentée et validée. Reste à savoir si cette méthode est utilisable par des enseignants non spécialistes en linguistique. C'est le début de l'expérimentation. Dans un contexte éducatif assez difficile, en concertation avec l'inspection des lettres, une procédure est définie : deux classes travaillent avec la méthode et deux classes de même niveau travaillent avec le manuel classique. La coordinatrice des Centres académiques de lecture et écriture suit l'expérimentation et élabore des tests : une courte dictée, une question de grammaire portant sur la reconnaissance de classes de mots, et une rédaction. Les résultats sont très significatifs : en une année les élèves utilisant la méthode obtiennent une moyenne de 12,8/20 ; ceux utilisant le manuel classique obtiennent 7,9/20. Les uns progressent de 6 points, les autres de moins de 2 points.

L'année suivante, grâce aux appuis conjugués des inspections du premier et du second degrés, l'expérimentation est officialisée par la validation du recteur. Les deux professeurs qui utilisaient le manuel classique s'associent au projet et utilisent dès lors *la méthode au Long Cours*. Des tests sont passés en début puis en fin d'année : on constate à nouveau une progression de 6 points.

En 2009-2010, tous les élèves de sixième du collège Leconte de Lisle bénéficient de la méthode. On constate en particulier que ces élèves de 6^e sont au même niveau que ceux de 3^e utilisant les manuels classiques. Les élèves ayant utilisé la méthode en 5^e et en 4^e sont au niveau des secondes et, mieux encore, à la question portant sur la reconnaissance des classes de mots, les 6^e utilisant le test de commutation font bien mieux que les terminales littéraires ! Enfin et surtout, le nombre d'élèves en difficulté d'écriture (orthographe et expression écrite) chute de 45 % à 6 % ! En 2010-2011, la méthode est adoptée par les trois collèges du centre de Saint-Louis : 20 classes de sixième l'expérimentent. Les résultats sont comparables à ceux obtenus les années précédentes.

Grâce à son support numérique, la méthode *Le Long Cours 6^e* continue à évoluer : après trois années d'expérimentation, elle s'enrichit d'un didacticiel vidéo et de références au livret de compétences. Des professeurs de collèges élaborent actuellement des prolongements en lecture et des liens avec le programme d'histoire des arts. En liaison avec les 6^e, des professeurs des écoles se proposent d'adapter la méthode à leur public.

Appuyée par des animations pédagogiques de circonscription, la méthode intéresse de nombreux professeurs des écoles, d'autant plus qu'ils peuvent constater les progrès réalisés en 6^e par leurs anciens élèves. Elle est réinvestie de la même façon au niveau de la formation des enseignants débutants de la formation continue au plan académique de formation (PAF) et dans les rencontres académiques, en lien avec le CARDIE.

Dans la circonscription de Saint-Louis de la Réunion, une adaptation de la méthode au cycle 3 est en début d'élaboration. La diffusion d'un CD-Rom allant dans le sens d'une liaison CM2/6^e construite autour de la méthode est en cours. Des professeurs de lettres recherchent et testent dans les trois collèges de Saint-Louis des prolongements de la méthode en lecture/littérature. Les professeurs sont généralement satisfaits de la méthode mais certains trouvent que les exercices sont très répétitifs : ils doivent être informatisés et cette option est en cours de réalisation.

Malgré son succès pédagogique, la méthode n'a pas trouvé de succès commercial : la politique de développement de la vente en ligne du centre régional de documentation pédagogique (CRDP) en 2009 n'a pas été suivie et a posé des problèmes techniques.

Au niveau national, compte tenu des résultats obtenus dans l'Académie de la Réunion, l'emploi de la commutation et la nomenclature mériteraient d'être plus largement testés et exploités.

la compétence de l'enseignant, que dans son développement officiel, le facteur temps est essentiel pour mesurer les effets ;

- elle naît de la convergence entre une initiative locale, fondée sur l'expertise réelle d'un acteur, qui plus est ici dans le champ disciplinaire et du changement de cadre institutionnel (à la fois dans la dimension didactique, la grammaire de texte, et dans sa dimension évaluative, l'approche par compétences générée par le socle commun) ;

- il s'agit d'une aventure collective, dans le sens où elle prend réellement corps quand la pratique se partage et se communique ; la pratique d'un seul devient alors le champ d'expériences apprenantes pour un corps de métier à l'échelle d'un établissement, d'un réseau d'écoles ;

- l'accompagnement institutionnel a été déterminant, tant dans la phase de repérage, de validation initiale, que d'accompagnement dans son développement ; il a été partagé en plusieurs composantes jouant des rôles distincts : inspections, CARDIE, mission locale, CRDP, formation ;

- ceux qui en parlent le mieux sont les élèves ; leurs témoignages et leurs parcours donnent une légitimité aux travaux que le seul discours ne trouverait pas forcément ;

- l'expérimentation rencontre des difficultés quand elle est « choséifiée » ; l'expérience au CRDP nous renseigne sur le fait qu'il s'agit d'abord de développement professionnel et de changements de pratiques, et non d'outils ou de méthode, avec ou sans mode d'emploi ;

- rien n'est dit sur la valorisation des compétences développées (sans parler d'une validation des acquis de l'expérience), pour les acteurs de terrain au titre du parcours professionnel. Le

développement professionnel des personnels, d'une équipe, dépendra aussi de ce facteur de ressources humaines, objectif du projet académique ;

- la diffusion par simple essaimage est un modèle au demeurant limité si elle n'est pas prise en charge et organisée dans le cadre de dispositifs de mutualisation organisés, les plus variés possible (formation, rencontres, animations, co-formation etc.) ;

- l'articulation entre niveau local, niveau académique et niveau national reste encore à approfondir pour diffuser cette « bonne pratique », avec les précautions relevées ci-dessus.

L'exemple du livret de compétences expérimental à Saint-Denis

Cette deuxième expérience, non encore achevée pour l'heure puisque l'expérimentation se terminera à la fin 2012, complète certains aspects :

- une initiative nationale et partenariale n'a que peu d'effet sans le portage local, ici du CARDIE, en mesure de produire une connaissance experte des réseaux et une analyse du potentiel de certains terrains plus « perméables » que d'autres. La seule transmission ne peut suffire à rencontrer son public ;

- l'expérimentation nationale prend du sens quand elle se construit et se négocie pas à pas avec les acteurs, et en premier lieu avec inspections, directeurs, puis collectifs d'enseignants. Elle requiert information certes, mais aussi contextualisation, analyse de la prescription et des gains pour chaque niveau, pour l'école, pour les élèves, comme pour les parents et les associations ;

- l'expérimentation nationale (*top-down*) prend corps si elle peut répondre à une problématique locale (ici, comment s'approprier la logique du socle, et s'inscrire dans une perspective plus large que le domaine scolaire *stricto sensu*) et si les acteurs de terrain peuvent avoir des marges substantielles pour la modifier afin de se l'approprier ;

- Le développement de ressources, d'objets partagés comme les différentes productions est essentiel pour faire vivre l'action ; on passe d'une logique d'exécution à une logique d'acteurs et de contributeurs ;

- l'expérimentation est une occasion où tous peuvent apprendre, on le voit ici avec les différents partenaires, les uns par rapport aux autres, les uns avec les autres.

CONCLUSION

L'expérimentation ainsi comprise au niveau local apparaît dans un contexte insulaire et forcément spécifique, comme un vecteur de modernisation des pratiques professionnelles et des organisations, à l'échelle d'une académie, en matière de pilotage et d'accompagnement du changement, mais aussi (et surtout) à l'échelle d'une école ou d'un établissement. Elle enclenche un processus dynamique de développement professionnel des personnels, en portant l'attention et les moyens sur les acquis des élèves, compris dans une approche beaucoup plus systémique que le système « traditionnel ». ■

Le livret de compétences expérimental (LCE) dans la circonscription de Saint-Denis 2

Proposée dans le cadre du livre vert « reconnaître la valeur de la jeunesse » et soutenue par le Fonds d'expérimentation pour la Jeunesse (FEJ), l'expérimentation porte sur le livret de compétences expérimental prévu par l'article 11 de la loi du 24 novembre 2009 relative à l'orientation et à la formation professionnelle tout au long de la vie. Dans l'académie de La Réunion le projet « LCE » concerne huit collèges dont quatre en Réseau ambition réussite (RAR devenus ÉCLAIR) et quatre en Réseau de réussite scolaire (RRS), un lycée professionnel et un lycée d'enseignement général et technologique. Trois circonscriptions sont impliquées, soit environ 2 500 jeunes élèves.

En une année d'expérimentation, les résultats sont particulièrement riches et à la hauteur d'un investissement humain conséquent en termes de coordination, d'animation, de direction. De nombreuses réunions préparatoires se sont tenues entre les différents niveaux de pilotage et un travail important de communication au travers de rencontres a été mené. Elles viennent en appui au socle commun de connaissances et de compétences, référentiel partagé par tous, enseignants, élèves, parents, et connu des partenaires, associations, communes, caisse d'allocations familiales.

Des productions sont réalisées comme un guide d'utilisation du LCE à l'usage des parents, un guide d'explicitation des compétences 6 et 7 du socle, une attestation-type, une charte graphique avec les pictogrammes des sept compétences du socle à mettre en relation avec des actions : information des parents d'élèves en réunion de rentrée, concours de pictogrammes, réunions de travail dans toutes les écoles en présence des parents et des associations, remise des portfolios et des guides d'utilisation aux enseignants. À l'initiative du recteur, une journée académique « parents » s'est tenue en novembre 2011 pour mieux faire connaître le LCE, mais aussi les compétences attendues du socle et initier les partenaires à l'écriture d'attestations à insérer dans le portfolio de l'enfant.

Voici le bilan fait par une des écoles :

Points positifs

- Les parents sont motivés. Ils reconnaissent l'intérêt de LCE pour leur enfant.
- Grâce au concours de pictogrammes, les élèves se sont appropriés le projet.
- L'organisation du portfolio en index est positive.
- Les parents ont donné leur accord pour faire le suivi des attestations dans le cadre associatif.
- Les associations périscolaires, telles que l'ARCV et Jeunesse 2000, sont favorables à l'action.
- Les liens développés avec le collège assurent une continuité pédagogique, permettant de mieux connaître le profil de l'élève.
- Les parents veulent mettre en place une amicale afin de permettre une meilleure égalité des chances (équité) et une meilleure appropriation du socle commun des connaissances et de compétences.
- Nombre d'attestations distribuées dans le cadre associatif : 66
- Nombre d'attestations distribuées dans le cadre familial : 132

Points négatifs

- Beaucoup d'enfants ne sont pas dans des structures associatives.
- Les parents ont des difficultés à remplir les attestations.
- Les associations ont aussi des difficultés pour remplir les attestations (problème de disponibilité du personnel, surcharge de travail, pas d'intérêt direct).

La dernière réunion de concertation a permis de prendre la mesure de l'attachement porté à ce projet par ses participants, et de l'effet de propagation qu'il porte.

Ainsi, deux équipes de direction de collèges non impliquées initialement étaient présentes et ont fait part de leur intérêt et de leur volonté de poursuivre la démarche au niveau sixième. Il est envisagé de toucher également les associations qui interviennent en dehors du projet éducatif global porté par la commune au besoin en impliquant la direction régionale de la jeunesse, des sports et de la cohésion sociale (DRJSCS) ou, comme proposé par un directeur, directement les parents.

BIBLIOGRAPHIE

Blondel Michel (2011), *Du pré carré à la compétence collective*. éditions ESSEC Business school/L'Aube Villes et Territoires, 244 pages.

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 (Aptéé).

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.

Évaluations et méthodologie d'évaluation

Les méthodes d'expérimentation en question

Les expérimentations randomisées, également appelées expériences contrôlées, sont des méthodes d'évaluation rigoureuses, mais leur mise en œuvre soulève des problèmes qui peuvent biaiser les estimations que l'on en déduit. Ces biais sont le plus souvent inhérents à la démarche expérimentale. Leur nature, et les éventuelles corrections qu'il est possible de leur apporter, sont ici passées en revue. L'accent est également mis sur la question des plans et protocoles d'expérience, trop souvent ignorée par les praticiens de ces méthodes statistiques appliquées aux sciences sociales¹.

Denis Fougère

CNRS, Centre de recherche en économie et statistique (CREST, Paris),

Laboratoire interdisciplinaire d'évaluation des politiques publiques (LIEPP, Sciences po, Paris), IZA (Bonn) et CEPR (Londres)

Parmi les méthodes utilisées pour évaluer des politiques sociales ou éducatives, les expérimentations randomisées font aujourd'hui l'objet d'un intérêt tout particulier. Dans le champ de l'éducation, elles peuvent être utilisées pour évaluer les effets d'une innovation pédagogique, d'une modification des rythmes scolaires, de l'introduction de cours de soutien, etc., et cela, avant même que de telles interventions soient ou non généralisées. Les plaidoyers en faveur des expérimentations mettent souvent l'accent sur les biais statistiques inhérents aux évaluations conduites à l'aide de données d'observation non expérimentales. L'évaluation par expérimentation randomisée essaie de s'affranchir de ces biais de sélection en proposant d'affecter de manière aléatoire (par tirage au sort) les établissements scolaires, classes ou élèves échantillonnés à un groupe de traitement qui bénéficiera du dispositif ou programme éducatif devant être évalué, ou bien à un groupe de contrôle qui n'en bénéficiera pas.

Cet article ne se propose pas de synthétiser les résultats des expérimentations randomisées conduites en milieu scolaire, à l'étranger ou en France. Son objectif est plus modeste.

Il s'agit ici de dresser la liste des principales difficultés auxquelles les praticiens de ces méthodes doivent faire face. Pour la plupart, ces difficultés ne sont pas insurmontables, pour autant qu'elles soient bien identifiées. Certaines sont toutefois très sérieuses et entraînent des biais statistiques significatifs dont l'ampleur ne peut être facilement évaluée.

RÉTICENCES ET OPPOSITIONS

Les réticences et oppositions à la mise en œuvre d'expérimentations randomisées reposent en général sur plusieurs types d'arguments. Le premier type a trait aux coûts élevés des expérimentations. Ces coûts sont de natures diverses. Les expérimentations sont certes coûteuses en termes de budget et de temps, mais l'on doit reconnaître que, dans la collecte des données non expérimentales d'observation qui se présentent sous la forme de fichiers longitudinaux, ces deux types de coûts sont souvent à peine moins importants. Les coûts

NOTE

1. Je remercie Cédric AFSA pour sa lecture attentive et ses remarques constructives.

spécifiques aux expérimentations sont des coûts de nature politique : persuader les chefs d'établissement, les enseignants, les parents d'élèves, des avantages et nécessités d'une expérimentation entraîne certes des dépenses, mais fait également courir un risque de perte de réputation ou de crédibilité aux concepteurs de l'expérience et aux groupes de pression qui les soutiennent. On le sait, les opposants au principe des expérimentations sociales invoquent plus souvent des arguments éthiques que budgétaires pour faire obstacle à leur mise en œuvre. Ils condamnent le déni de traitement imposé aux personnes (ici, aux élèves) du groupe de contrôle et le tort substantiel que ce déni peut leur faire subir. Si le programme s'avère être significativement bénéfique, refuser ce traitement à certains à l'issue d'un tirage aléatoire et de ce fait rationner une ressource publique est contraire à l'éthique. Mais, comme le fait remarquer Burtless [1], l'argument est le même dans les expérimentations cliniques et, bien que les enjeux individuels y soient souvent bien plus importants, le recours à l'expérimentation par randomisation y est plus facilement admis. Devant décider de la justesse du recours à l'expérimentation randomisée, il nous faut donc répondre à une et une seule question : est-il *a priori* plus juste, ou plus bénéfique, de généraliser, éventuellement en le rendant obligatoire, un programme qui risque d'être préjudiciable, de ne pas rendre accessible à tout un chacun un programme qui peut s'avérer profitable, ou bien de mettre en place, préalablement à toute décision de généralisation ou d'abandon définitif du programme, une procédure d'évaluation par expérimentation avec l'aide et le consentement éclairé d'un

nombre limité d'individus dont certains acceptent de se voir refuser, à l'issue d'un tirage au sort et à l'exclusion de toute autre considération (de revenu, de sexe, d'origine sociale ou nationale), le droit d'accès au programme durant la période limitée de l'expérience? Cette dernière solution ne peut être envisagée que sous certaines conditions très précises, admises par les acteurs concernés à l'issue d'un débat préalable. Ces conditions, qui tout à la fois permettent de recourir à l'expérimentation et déterminent son protocole, sont au minimum au nombre de trois :

1. les participants à l'expérience doivent être pleinement informés des avantages comme des risques que leur participation peut leur procurer ou leur faire courir (principe de consentement éclairé);
2. dans les cas où l'on anticipe que le programme peut provoquer des risques ou des torts, il faut prévoir de compenser les élèves qui seront affectés au groupe de traitement; dans le cas où l'on anticipe que le programme peut être bénéfique, il faut prévoir de compenser ceux qui seront affectés au groupe de contrôle (principe de compensation);
3. enfin, une expérimentation n'est envisageable que dans les cas où l'on ignore *a priori* si le programme est bénéfique ou non; s'il s'avère préjudiciable, il est préférable d'observer ses effets sur un nombre limité d'élèves consentants que sur un très grand nombre d'élèves contraints (principe de précaution).

Un des enseignements les plus importants des expériences randomisées qui ont été conduites dans le domaine social, économique ou éducatif, tout d'abord aux Etats-Unis puis plus récemment en Europe, est que

leur réussite, c'est-à-dire l'assurance de la fiabilité de leurs conclusions, dépend de la participation étroite des personnels chargés de leur mise en œuvre, à la définition de leurs objectifs et de leurs modalités pratiques. Obtenir cet agrément est, pour l'administration centrale ou locale qui souhaite mettre en place une expérience, chose d'autant plus facile que le plan (*le design*) de l'expérience a été préalablement élaboré avec les acteurs chargés de la réalisation de l'opération et que ce plan est perçu comme respectant les objectifs et contraintes de l'expérience.

Mais, dans ce cas favorable, les résultats d'une expérimentation randomisée sont susceptibles d'être soumis à un certain nombre de biais inhérents, que nous allons maintenant énumérer, et dont il faut avoir conscience.

LE PRINCIPE DES EXPÉRIMENTATIONS RANDOMISÉES

Traditionnellement, les études statistiques qui tentent d'identifier et d'estimer les effets de politiques éducatives (par exemple, la création des zones d'éducation prioritaire) exploitent des données provenant d'enquêtes non expérimentales recueillies auprès d'établissements comparables. Pour mener l'évaluation, ces études utilisent généralement une comparaison des caractéristiques moyennes (par exemple, le taux de redoublement, les notes aux épreuves de mathématiques ou de français, etc.) de deux types d'établissements, ceux ayant bénéficié de la mesure examinée (en ce cas, la labellisation EP) et ceux n'en ayant pas bénéficié. La différence des moyennes de résultat

tats dans les deux groupes est en fait la combinaison de deux effets, l'effet spécifique du dispositif que l'on cherche à mesurer, mais aussi celui qui provient du fait que les deux groupes d'établissements peuvent se comporter différemment face au dispositif considéré. En effet, ce dispositif peut ne pas intéresser de la même façon tous les établissements. Certains peuvent demander avec force à bénéficier du label et des avantages qu'il procure, d'autres au contraire peuvent être réticents, estimant que la stigmatisation créée par un label de ce type est supérieure aux avantages afférents. En ce cas, la différence de résultats entre les deux groupes d'établissements reflète en partie ces comportements d'autosélection.

Pour qu'une expérimentation randomisée élimine le biais de sélection (ou d'autosélection), deux conditions doivent être vérifiées :

1. la probabilité de candidature au programme considéré ne doit pas être affectée par le fait que l'accès à ce programme est conditionné par le tirage au sort ;

2. les comportements ne doivent pas être modifiés par la participation à l'expérimentation.

Dans une expérimentation randomisée, le plan d'expérience le plus fréquent est celui où les établissements éligibles au dispositif sont d'abord informés des buts et principes de l'expérimentation, ceux acceptant d'y participer étant ensuite affectés aux groupes de traitement et de contrôle par tirage au sort. Le fait que la décision de participation à l'expérience précède le tirage au sort garantit que les groupes de contrôle et de traitement constituent des sous-échantillons aléatoires de la population participante. En ce cas, l'effet moyen du

programme pour les établissements qui souhaitent participer au dispositif peut être estimé par la différence entre la moyenne des résultats dans le groupe de traitement et la même moyenne dans le groupe de contrôle.

Lorsque le tirage au sort précède le choix des établissements de participer ou non au programme proposé, l'effet moyen de ce programme est plus difficilement identifiable. Le fait que certains établissements refusent de participer au programme dans le cadre de l'expérimentation ne permet pas d'estimer la valeur moyenne de l'effet du programme, une fois celui-ci généralisé. Manski [10] a néanmoins démontré qu'il est possible de borner la valeur de cet effet moyen, et qu'il est également possible de borner l'effet moyen d'une mesure généralisée mais facultative à partir des résultats d'une expérimentation randomisée.

LE BIAIS DE RANDOMISATION

Les expérimentations avec assignation aléatoire sont susceptibles d'être soumises à ce que les spécialistes de ces techniques appellent le biais de « randomisation ». Ce biais apparaît lorsque le groupe de traitement, constitué à l'issue de la procédure d'affectation aléatoire des participants, diffère du groupe de personnes auxquelles l'intervention (par exemple, la politique éducative) pourrait être appliquée, une fois l'expérience menée et jugée concluante. Ce biais a été maintes fois observé dans des essais cliniques, mais aussi dans de nombreuses expérimentations sociales. Il est en effet souvent plus difficile de persuader des personnes de participer à une expérience randomisée que de

les inciter à participer à une observation non expérimentale. Cette difficulté peut, par exemple, apparaître lorsqu'il s'agit de demander à des chefs d'établissement scolaire, des enseignants, des élèves ou des parents d'élèves de ces établissements, préalablement choisis de manière aléatoire, de participer à une expérimentation randomisée, et qu'un certain nombre d'entre eux refusent, ou bien, pire encore, acceptent, puis cessent de participer à l'expérience avant que celle-ci parvienne à son terme.

Par exemple, il est à craindre que les établissements qui refusent de participer à l'évaluation expérimentale d'une nouvelle politique éducative soient précisément ceux qui pensent en tirer le moins d'avantages. En ce cas, la différence des moyennes de résultat entre les deux groupes de traitement et de contrôle ne mesure pas vraiment ce que l'on souhaite mesurer, à savoir, l'effet moyen de la politique éducative, mais l'effet moyen de cette politique au sein des établissements, des classes, etc., qui ont accepté de participer à l'expérience. L'écart entre ces deux quantités peut être assez substantiel, comme l'ont montré des expériences conduites dans différents domaines.

LES EFFETS DE TYPE « HAWTHORNE »

Les effets de type « Hawthorne » sont une variante du biais de randomisation. Dans le cadre expérimental, ce type d'effet correspond au cas où le comportement des sujets étudiés est modifié par le fait d'avoir été ou non retenu pour participer à l'expérience, et non par le dispositif ou le programme lui-même. En effet, certains peuvent éprouver un sentiment de gratitude

ou au contraire d'injustice selon qu'ils ont été ou non choisis pour participer à une expérience susceptible d'améliorer leur situation matérielle. Le terme provient des expériences menées dans l'établissement d'Hawthorne de la Western Electric Company, à Chicago entre 1924 et 1933. La première de ces expériences montra que des variations de la luminosité sur le lieu de travail avaient pour effet une amélioration de la productivité et de la satisfaction des salariés, indépendamment du fait que la lumière fut plus ou moins forte. En présence d'effets de type « Hawthorne », l'estimation de l'impact du programme risque, là encore, d'être biaisée, le sens du biais (positif ou négatif) étant ici difficile à prévoir. Seule une observation très étroite des élèves, professeurs, chefs d'établissement, etc., permet alors de comprendre comment, et dans quelle mesure, leur comportement peut être éventuellement modifié par leur participation à l'expérimentation.

LA FAIBLESSE DES EFFECTIFS

Le biais de randomisation peut également résulter de la faiblesse des effectifs participant à l'expérience, les résultats étant en ce cas très imprécis. Par exemple, de trop petits échantillons conduisent trop souvent à rejeter l'hypothèse d'un effet positif (ou négatif) de l'intervention ou du programme expérimental, non parce que la différence des moyennes de résultat observés dans les deux groupes est effectivement positive (ou négative) mais parce que son écart-type est trop large pour aboutir à une conclusion suffisamment fiable. Étayant leurs analyses par des exemples concrets,

Heckman [4], Heckman et Smith [6], et Manski [9] ont ainsi pu montrer que des expériences conduites à l'aide d'échantillons trop faibles peuvent conduire à des résultats sensiblement différents de celles réalisées à l'aide de groupes de beaucoup plus grande taille. Les deux difficultés peuvent se cumuler dès lors que le consentement des participants pressentis peut dépendre de la taille des échantillons constitués ; pour eux, il peut être plus facile de participer à une expérience qui impliquera un grand nombre d'élèves ou d'établissements, les écarts individuels à la moyenne passant alors plus vraisemblablement inaperçus et les éventuels effets de stigmatisation étant de ce fait atténués. Mais cet avantage a une contrepartie : une expérimentation de grande dimension est tout à la fois plus coûteuse et plus difficile à mettre en œuvre.

LE BIAIS D'ATTRITION

Ce biais résulte du fait que certains participants, membres du groupe de traitement ou du groupe de contrôle, décident de quitter l'expérience avant la fin de celle-ci. Ce phénomène se produit plus fréquemment dans les expériences qui se déroulent sur une échelle de temps assez longue, par exemple sur une ou plusieurs années. L'importance de ce biais, inhérent aux expérimentations qui portent sur des sujets humains, a été reconstruite très tôt. Hausman et Wise [3] faisaient remarquer que le problème de l'attrition provient *de facto* de l'introduction du facteur temporel dans les expérimentations sociales ou cliniques et que ce problème n'apparaît pas dans les expérimentations généralement conduites en physique, en chimie ou en

biologie. On a pu, par exemple, observer un tel phénomène dans le cadre du programme STAR, qui a permis de tester, aux États-Unis, les effets d'une réduction de la taille des classes dans le cadre d'un programme expérimental conduit sur une très grande échelle [2]. Le biais d'attrition est difficilement réductible. Heckman, Smith et Taber [7] ont précisé la condition sous laquelle il est possible d'identifier l'effet moyen du programme sur le résultat des membres du groupe de traitement qui participent pleinement à l'expérimentation, c'est-à-dire jusqu'à son terme (cette quantité n'est pas exactement l'effet moyen du programme sur ceux qui en ont bénéficié, cette dernière quantité ne pouvant être identifiée en présence d'attrition). Cette condition stipule que le résultat moyen des membres du groupe de traitement qui abandonnent l'expérimentation avant qu'elle ne prenne fin doit être égal au résultat moyen des membres du groupe de contrôle qui auraient abandonné de la même façon l'expérimentation s'ils avaient été placés dans le groupe de traitement. Heckman, Smith et Taber [7] proposent des tests statistiques permettant de vérifier la validité de cette condition.

LE BIAIS DE SUBSTITUTION

Ce biais peut apparaître dès lors que certains membres du groupe de contrôle, n'ayant pas accès au programme éducatif durant le temps de l'expérience, cherchent à se procurer des substituts à ce programme avant la fin de l'expérience. Pensons à un programme éducatif qui accroît le nombre d'heures de mathématiques. Il est possible que les parents des élèves des classes faisant partie du

groupe de contrôle recourent alors à des cours privés, ou bien que les professeurs en charge de ces classes, hostiles au principe de l'expérimentation, décident unilatéralement d'intensifier le nombre et le contenu de l'enseignement de mathématiques. Il est clair qu'il est difficile de leur interdire d'agir ainsi, leur argument étant de ne pas défavoriser les élèves qui n'ont pas accès à ce programme qu'ils supposent bénéfique. Là encore, réduire ce biais, qui tend à sous-estimer les effets du programme ou de la politique éducative que l'on souhaite évaluer, est chose difficile. Heckman [5] a toutefois précisé les conditions sous lesquelles il reste possible d'identifier l'effet moyen du dispositif pour ceux qui en bénéficient dans le cas où le groupe de contrôle a accès à un substitut. La première condition stipule que, pour chaque participant, membre du groupe de traitement ou du groupe de contrôle, l'effet du dispositif devant être évalué est le même que celui de son substitut. La seconde condition est que l'effet du dispositif évalué doit être le même pour chaque participant. Enfin, il faut que l'effet moyen du dispositif pour les membres du groupe de traitement qui participent à l'expérimentation jusqu'à son terme soit égal à l'effet moyen du substitut au dispositif pour les membres du groupe de contrôle qui ont accès à ce substitut. On le voit, ces conditions sont extrêmement strictes et n'ont que peu de chances d'être vérifiées en pratique. En conséquence, la seule solution pragmatique, quoique difficile à mettre en œuvre, est de limiter, autant que faire se peut, les possibilités d'accès à des substituts au sein du groupe de contrôle.

MIEUX DÉFINIR LES PROTOCOLES ET PLANS D'EXPÉRIENCE

Dans de trop nombreux cas, les chercheurs conduisant des expérimentations randomisées en sciences sociales négligent de définir le protocole statistique qui assurera l'efficacité maximale des résultats de l'expérimentation, et ce, avant même que celle-ci ne débute sur le terrain. Ce souci est depuis longtemps au cœur de la démarche des statisticiens conduisant des expérimentations dans le cadre clinique ou dans d'autres disciplines (agronomie, physique, chimie, etc.), mais il semble avoir été jusqu'à aujourd'hui largement absent des préoccupations des chercheurs en sciences sociales ou en sciences de l'éducation. De ce fait, on a souvent pu constater que les estimations déduites des expérimentations sociales étaient statistiquement peu significatives et/ou peu robustes.

Le protocole expérimental doit être avant tout calibré par un calcul de puissance². Préalablement au lancement de l'expérimentation, les chercheurs doivent s'efforcer d'énoncer clairement la question à laquelle ils souhaitent répondre et spécifier l'hypothèse nulle correspondante (par exemple, qu'il n'existe pas de différence de résultats entre les deux groupes à l'issue de la mise en œuvre du programme expérimental dans le groupe de traitement). Ils doivent également faire des hypothèses concernant l'espérance mathématique de la variable d'intérêt dans le groupe de traitement (par exemple, le taux de redoublement est en moyenne plus faible dans ce groupe). Ces hypothèses leur permettent ensuite de calculer, compte-tenu du plan d'expérience, la taille minimale des échantillons permettant de détecter avec une certaine

puissance (souvent fixée à 80 %) un écart d'une certaine valeur entre les deux groupes. Plus l'écart postulé est élevé, plus l'hypothèse nulle est facile à rejeter, et plus la taille de l'échantillon est faible. Ce genre d'exercice permet de savoir *a priori* si le budget de l'expérience permet ou non de détecter un écart d'une certaine ampleur. Lorsque les chercheurs rédigent leur rapport, joindre le calcul de puissance permet d'informer le lecteur sur les hypothèses faites *ex ante* par les chercheurs.

Dans une étude récente, List, Sadoff et Wagner [8] énoncent quatre règles permettant d'améliorer l'efficacité statistique des protocoles expérimentaux :

1. lorsque la variable de résultat est continue (par exemple, une note à un examen ou à une évaluation) et lorsque la variance de la moyenne de cette variable est supposée être la même dans les deux groupes (de traitement et de contrôle), il faudrait affecter les sujets (ici les élèves) de

NOTE

2. La puissance statistique est la probabilité que l'hypothèse nulle (ici, l'absence de différence de résultats entre les deux groupes) soit rejetée, et que l'expérimentation puisse donc permettre de repérer l'association réellement existante entre le programme éducatif évalué et la variable de résultat considérée (par exemple, le taux de redoublement ou la moyenne des notes en mathématiques). La puissance est déterminée par différents facteurs, parmi lesquels la fréquence de la variable de résultat considérée, le protocole de l'expérience et la taille de l'échantillon. Lors de la mise en place de l'expérience, les chercheurs doivent opter pour une certaine puissance en fonction de laquelle la taille de l'échantillon est ensuite déterminée. Une puissance statistique de 80 % est généralement considérée comme le minimum exigible. Ce qui signifie qu'il y a 80 % de chance que l'expérimentation puisse mettre en évidence l'effet recherché.

manière égale dans les deux groupes ;

2. lorsque les variances ne sont pas identiques, le rapport des effectifs des deux groupes devrait être égal au rapport des écarts-types de la moyenne des résultats dans les deux groupes ;

3. lorsque le coût d'échantillonnage varie d'une cellule expérimentale à l'autre (une cellule pouvant être ici un établissement, une classe, un groupe d'élèves homogène), le rapport des tailles d'échantillon devrait être inversement proportionnel à la racine carrée des coûts relatifs ;

4. lorsque l'unité que l'on tire aléatoirement (par exemple, un établissement) est différente de l'unité d'analyse (par exemple, une classe), il faudrait tenir compte du coefficient de corrélation entre classes d'un même établissement.

LA QUALITÉ DES ÉVALUATEURS

Cet aspect, souvent ignoré, n'est pas le moindre risque encouru par une expérimentation randomisée. La qualité des agences ou opérateurs en charge de la mise en œuvre de l'expérimentation est par essence variable. Ces opérateurs sont en général choisis à l'issue d'une procédure d'appel d'offres, pilotée par un comité d'usagers, de praticiens et de personnes qualifiées. Rien ne garantit

que la proposition effectivement choisie émane d'un groupe d'experts de qualité suffisante. Somme toute, le nombre de statisticiens de valeur consacrant leur temps et leur énergie à la pratique des expérimentations est assez limité. Il est donc possible que les équipes répondant aux appels d'offres soient parfois de qualité moyenne et ne contiennent que rarement des chercheurs présentant des garanties de qualité suffisantes. Plus la pratique des expérimentations, et notamment des expérimentations à grande échelle, se diffusera, et plus le problème de qualité des évaluateurs se posera. Certains répondront qu'il suffit de former de plus en plus de jeunes chercheurs de valeur à cette méthode d'évaluation. Mais l'expérimentation n'est pas la seule technique statistique nécessaire à notre compréhension des phénomènes économiques, politiques et sociaux. Elle n'est pas non plus la plus exigeante, bien que les difficultés statistiques qu'elle soulève soient réelles, et trop souvent ignorées par les économistes et sociologues qui y consacrent tout ou partie de leur carrière.

CONCLUSION

Les expérimentations randomisées sont aujourd'hui un dispositif d'évaluation de plus en plus utilisé

et de mieux en mieux maîtrisé. Les gains que l'on peut en attendre dans un domaine aussi important que celui des politiques éducatives sont substantiels. Pour autant, leur mise en œuvre peut encore être améliorée. Parmi les progrès les plus notables devant être enregistrés, figure certainement la meilleure connaissance des problèmes méthodologiques qui leur sont inhérents. Cet article en a dressé la liste. Il appartient maintenant aux chercheurs d'œuvrer à la résolution de ces problèmes et de proposer des protocoles statistiques qui permettent d'obtenir des résultats fiables, et surtout des interprétations analytiques allant au-delà du simple constat « cela fonctionne/ cela ne fonctionne pas ». Pour le dire autrement, pour ceux qui veulent persuader du bien-fondé de la démarche expérimentale en sciences sociales, et tout particulièrement dans le domaine de l'éducation, l'impératif est aujourd'hui, selon moi, d'abandonner l'effervescence praticienne pour atteindre la maturité scientifique. À défaut de cette transition, beaucoup d'expérimentations randomisées risquent de livrer des résultats peu convaincants et insuffisamment articulés. ■

BIBLIOGRAPHIE

- [1] **Burtless G., 1995**, The Case for Randomized Field Trials in Economic and Policy Research, *Journal of Economic Perspectives*, vol. 9, n° 2, p. 63-84.
- [2] **Ding W., and Lehrer S. F., 2010**, Estimating Treatment Effects from Contaminated Multiperiod Education Experiments: The Dynamic Impacts of Class Size Reductions. *The Review of Economics and Statistics*, vol. 92, n° 1, p. 31-42.
- [3] **Hausman J., and Wise D., 1979**, Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment. *Econometrica*, Vol. 47, n° 2, p. 455-473.
- [4] **Heckman J., 1992**, Randomization and Social Policy Evaluation, dans *Evaluating Welfare and Training Programs*, édité par C. Manski et I. Garkinkel. Cambridge, Massachusetts, Harvard University Press, p. 201-230.
- [5] **Heckman J., 1997**, Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator. *The Journal of Human Resources*, vol. 32, n° 3, p.441-461.
- [6] **Heckman J. and Smith J., 1995**, Assessing the Case for Social Experiments, *Journal of Economic Perspectives*. vol. 9, n° 2, p. 85-110.
- [7] **Heckman J., Smith J., and C. Taber, 1998**, Accounting for Dropouts in Evaluation of Social Programs. *The Review of Economics and Statistics*, vol. 80, n° 1, p. 1-14.
- [8] **List J., Sadoff S. and Wagner M., 2010**, *So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design*, National Bureau of Economic Research, Working Paper 15701, Cambridge, Massachusetts.
- [9] **Manski C., 1995**. *Identification Problems in the Social Sciences*. Harvard University Press, 172 pages.
- [10] **Manski C., 1997**, The Mixing Problem in Programme Evaluation. *The Review of Economic Studies*, vol. 64, n° 4, p. 537-554.

Petit précis méthodologique d'évaluation qualitative

Sophie Divay

Centre associé d'Ile-de-France au Centre d'études et de recherches sur les qualifications (Céreq), université Paris 1 Panthéon-Sorbonne, CNRS, Paris

Les projets du Fonds d'expérimentation pour la jeunesse renouvellent et questionnent les pratiques d'évaluation, au coeur de ce dispositif d'expérimentation. Dans le cas des approches qualitatives, la proximité avec les porteurs de projets, leurs attentes, peuvent fragiliser la nécessaire objectivité de l'évaluateur. Des réponses méthodologiques existent pour lui permettre de dessiner les contours d'une posture juste et rigoureuse.

En France, la démarche dite « d'expérimentation sociale » s'est récemment développée dans le champ de l'action publique. Elle a notamment été largement plébiscitée dans le cadre de la politique en faveur des jeunes, menée par Martin Hirsch, Haut commissaire à la jeunesse, qui a ouvert en avril 2009, une première session à projets du fonds d'expérimentation pour la jeunesse.

Créé par la loi du 1^{er} décembre 2008 généralisant le RSA, ce fonds était initialement prévu pour financer des projets favorisant l'insertion des jeunes de moins de 26 ans. Son champ d'application s'est par la suite étendu à l'ensemble des questions relatives à la jeunesse.

Le Centre associé au Céreq¹ d'Ile-de-France s'est positionné en tant qu'évaluateur d'une expérimentation portée par un rectorat s'inscrivant dans l'un des programmes de l'appel d'offres intitulé : « Prévenir le décrochage scolaire et organiser les collaborations pour suivre les jeunes décrocheurs ».

Cette réponse commune à ce rectorat et au centre associé au Céreq d'Ile-de-France a été retenue par le Haut commissariat à la jeunesse (HCJ), et les opérations concrètes de terrain ont démarré en septembre 2009. Avec un

recul de plus de deux ans, il est maintenant possible de tirer certains enseignements de cette mission d'évaluation qui présente une particularité, celle de reposer sur une approche qualitative, alors que la plupart des expérimentations financées par le HCJ devaient déployer des démarches quantitatives, et ce, conformément aux prescriptions méthodologiques des commanditaires. Il s'agira ici de rendre compte d'une expérience *a priori* marginale, mais qui, de fait, se révèle efficace car adaptée aux conditions réelles de réalisation du dispositif d'expérimentation.

LES OBJECTIFS DU PORTEUR DE L'EXPÉRIMENTATION

Sous l'impulsion de son recteur, l'académie s'est engagée en 2009 dans la mise en place d'une expérimentation sociale financée par le HCJ impliquant vingt lycées situés dans tous les départements de son territoire administratif. Le projet met l'accent sur la volonté du rectorat d'inciter les établissements choisis

NOTE

1. Centre d'études et de recherches sur les qualifications

à mettre en route ou à développer l'activité d'un dispositif prévu de longue date dans les textes : les Cellules d'action contre l'absentéisme² (C2A). La mission officielle de ces structures relève de la prévention des sorties sans qualification des élèves dont les absences et les signes de démotivation doivent donner lieu à une prise en charge rapide. Ces C2A sont composées de membres de la direction et de l'équipe éducative (proviseur, proviseur-adjoint, conseiller principal d'éducation [CPE], conseiller d'orientation-psychologue [COP], assistante sociale, infirmière scolaire, professeurs, etc.). Ce groupe de travail est censé se réunir régulièrement afin d'établir un diagnostic partagé sur les situations des élèves repérés. Ces derniers peuvent se voir offrir un accompagnement, souvent individualisé, pour les aider à construire un projet personnel et à s'investir dans une voie répondant davantage à leurs attentes. Un éventail de solutions est susceptible d'être proposé à l'élève, allant du tutorat à la participation à un atelier théâtre en passant par des visites d'entreprises. Pour aller vite, la finalité est de trouver le moyen de « raccrocher les décrocheurs ».

L'expérimentation comprend différentes actions réparties sur trois années scolaires (de septembre 2009 à juin 2012) : un appui méthodologique *in situ* réalisé par des consultants auprès des membres des C2A, une formation inscrite au PAF (plan académique de formation) portant sur les problématiques de l'adolescence,

une formation à l'entretien assurée par les consultants pour aider les enseignants dans leur fonction de tuteur, des séminaires dits inter-C2A réunissant certains membres de tous les lycées lors de journées thématiques, et des réunions de travail réservées aux chefs d'établissement.

L'application de ce programme devrait, à terme, entraîner une plus grande mobilisation des membres des C2A, ainsi qu'une prise en charge plus efficace des élèves « décrocheurs », et donc une baisse des taux d'absentéisme et d'abandon en cours d'études. Afin de mesurer les effets de l'expérimentation sur l'activité des C2A, le rectorat comptait sur une application informatique appelée « Suivi des élèves »² que les lycées devaient utiliser, entre autres pour enregistrer les élèves suivis par leur C2A.

DE LA MÉTHODE IDÉALE AUX NÉCESSAIRES AMÉNAGEMENTS

Les expérimentations lancées dans le cadre du HCJ s'appuient sur un cadre méthodologique dit innovant parce que encore assez peu répandu en France dans le domaine des politiques publiques. La principale approche recommandée est celle de la « randomisation » ou de « l'assignation aléatoire » ; elle suppose l'application de méthodes économétriques sophistiquées et s'inscrit, comme le souligne Thomas Le Bianic et Antoine Vion [4], dans une logique du pilotage en amont des politiques publiques. Les conditions de réalisation de l'expérimentation sociale du rectorat sont rapidement apparues incompatibles avec l'application d'une telle démarche d'évaluation pour plusieurs raisons.

Premièrement, les actions programmées portaient sur le fonctionnement de structures collectives, les C2A, dont il s'agissait d'observer les évolutions des dynamiques relationnelles et organisationnelles. Si de tels changements sont objectivables, ils sont difficilement quantifiables.

Deuxièmement, les variations des taux d'absentéisme et d'abandon ne sont pas de bons indicateurs de l'efficacité de l'activité des C2A. D'une part, les établissements présentaient des situations très contrastées ; pour beaucoup, la C2A restait à construire, même si une structure portant ce nom existait. Les résultats obtenus au cours d'un suivi sur trois années ne pouvaient donc pas être significatifs. D'autre part, même si les C2A avaient été d'emblée fonctionnelles, la fiabilité des données quantitatives sur l'absentéisme et le décrochage, produites par les établissements, pose question. En effet, les pratiques d'enregistrement observées montrent une très grande diversité, notamment en ce qui concerne les critères appliqués, et ce, tant d'un établissement à l'autre, qu'au sein d'un même établissement d'une année sur l'autre.

Troisièmement, l'outil informatique « Suivi des élèves » n'a pas fourni les éléments de suivi escomptés puisque son utilisation a connu un démarrage très lent dans les lycées qui ont été confrontés à des difficultés techniques et qui ont perçu cette application comme une source supplémentaire de (sur)charge de travail. Par ailleurs, comment aurait-on pu interpréter des variations du nombre d'élèves accompagnés en C2A, sachant par exemple que l'opérateur consultant encourage les équipes à mieux cibler les élèves en difficulté, entraînant de ce fait, peut-être une baisse du nombre

NOTE

2. Nom d'emprunt permettant le respect de l'anonymat de l'académie concernée par l'expérimentation.

d'élèves suivis. Comment alors interpréter de tels éventuels résultats en s'en tenant à des données chiffrées ?

L'option méthodologique retenue a été celle de l'approche qualitative qui, si elle ne se conforme pas au modèle méthodologique dominant, ne trahit pas pour autant les orientations du HCJ puisque le *Guide méthodologique pour l'évaluation des expérimentations sociales*, publié en 2009, contient [2, p. 17] la préconisation suivante : « L'approche qualitative se justifie aussi comme démarche spécifique appropriée à certains types de questionnements qui ne peuvent pas, ou difficilement, faire l'objet de mesures quantitatives. Face à l'évaluation objectivée par la mesure, elle introduit un jugement de type inductif et compréhensif. » Cette consigne nous a autorisés à aller au-delà des aménagements déjà proposés par des économistes comme Philippe Lemistre [3] qui estime, pour sa part, que « les enquêtes de terrain qualitatives en amont, menées parallèlement, doivent accompagner le test expérimental ».

QUELLE POSTURE ADOPTER EN TANT QU'ÉVALUATEUR ?

Afin de s'acquitter au mieux de ses fonctions, l'évaluateur doit tenter d'adopter une attitude aussi neutre que possible, même si l'application de méthodes qualitatives le met moins à l'abri qu'une démarche quantitative. Il est, en effet, directement confronté aux personnes concernées par l'expérimentation à travers les entretiens ou les observations qu'il mène. La nécessaire prise de distance vis-à-vis du terrain suppose un travail réflexif constant qui l'amène à s'interroger *a posteriori* sur les biais

qu'il a pu introduire afin d'éviter de répéter les mêmes erreurs, sachant que sa simple présence introduit des perturbations qu'il ne peut pas de toute façon neutraliser [5]. Il aura à interpréter les résultats obtenus en tenant compte des biais identifiés. Un des plus fréquents est sans doute celui des attentes de rôle que les interlocuteurs rencontrés ont à son égard. Le titre d'évaluateur est lourd à porter : il évoque souvent une démarche d'audit, de contrôle, d'inspection. Ainsi, certains des proviseurs rencontrés dans le cadre de notre expérimentation se sont-ils efforcés de mettre en valeur, de façon parfois assez appuyée, le bon fonctionnement de leur établissement ou de leur C2A. Ces efforts ont été d'autant plus déployés que des tensions, voire conflits, existaient entre la direction et les personnels. Charge alors à l'évaluateur de rappeler ce qu'il vient évaluer : en l'occurrence, non pas le bon ou mauvais fonctionnement de la C2A, mais les actions dispensées dans le cadre de l'expérimentation.

Dans un tel contexte, il est essentiel de s'en tenir à des questions simples, telles que : « Qu'avez-vous pensé de l'accompagnement proposé ? », et de les répéter à l'identique sur tous les terrains. Sont par ailleurs à bannir tous jugements de valeurs portés sur les acteurs dans le milieu considéré (lycée d'accueil, autres lycées, services des inspections académiques et du rectorat, etc.). Les discours et positionnements des acteurs seront appréhendés selon une approche compréhensive, cherchant à dégager le sens qu'ils donnent à leurs actions, c'est-à-dire les logiques qu'ils poursuivent et leur cohérence interne.

Enfin, neutralité et approche compréhensive assureront d'autant mieux

un relevé objectif des informations que l'évaluateur s'abstiendra de délivrer tout conseil ou toute préconisation avant la fin du dispositif, résultats qu'il livrera en garantissant l'anonymat des personnes et des structures. Rien ne l'empêche, en revanche, de rendre compte de l'avancement de ses investigations en présentant, par exemple lors de comités de pilotage, le calendrier des opérations réalisées et programmées, ainsi que les méthodes appliquées pour collecter et analyser les données.

OBJETS ET TECHNIQUES D'ÉVALUATION

Afin de construire sa démarche, l'évaluateur a tout intérêt à définir *a priori*, au moins dans les grandes lignes, l'objet de son évaluation. S'il est mandaté pour évaluer les effets de l'expérimentation mise en œuvre, il ne saurait s'en tenir à déterminer si les objectifs du porteur ont été ou non atteints. Il lui incombe non seulement de répondre aux questions du porteur, mais également de poser des questions à ce même porteur, notamment sur la cohérence de son projet de départ. Émergent, en effet, au cours du travail de terrain, des questionnements émanant des personnes concernées qui peuvent aller jusqu'à remettre en cause en grande partie la pertinence du dispositif. Ainsi, nombre des membres des C2A interviewés se disent satisfaits de l'accompagnement *in situ* effectué par les opérateurs consultants. Toutefois, ils pointent de nombreux facteurs endogènes et exogènes à l'institution qui entravent l'efficacité de leur action de prévention : que faire face à la diminution récente des places de stages offertes par les entreprises,

source de démotivation des élèves et cause d'absentéisme? Comment pallier les conséquences déstabilisantes du fort *turn over* du personnel de direction existant au sein des établissements? Ou encore, comment maintenir la mobilisation des enseignants, membres d'une C2A, touchés par la suppression de postes dans le cadre des fermetures de classes?

La prise en compte des questions de départ, propres au porteur, ainsi que des questions émergentes, provenant des destinataires de l'expérimentation, suppose une approche empirique, mais aussi l'application d'un raisonnement inductif partant du recueil de faits particuliers, dont l'analyse débouche sur l'énoncé de principes généraux. Cette perspective constructiviste favorise la prise en compte de constats contre-intuitifs susceptibles de passer inaperçus faute d'une attention suffisante portée à la réalité des faits. Il faut, par exemple, interroger les professionnels des lycées pour mieux saisir les différentes facettes de l'absentéisme : quel profane aurait en effet imaginé que des élèves absentéistes passent une grande partie de leur temps dans les murs de leur lycée, en compagnie d'autres élèves, sans aller en cours?

Cette démarche empirique, inductive et compréhensive, est mise en œuvre à travers l'utilisation de techniques dites classiques de la méthode qualitative. L'entretien semi-directif, enregistré, totalement ou partiellement retranscrit et fondé sur un guide d'entretien souple, représente un des instruments les plus utilisés. La passation des entretiens est de préférence individuelle; faute de temps, elle peut être collective, mais il est alors préférable de rassembler des pairs au cours de la même interview (un groupe com-

posé de professionnels appartenant à la même catégorie) et d'éviter autant que faire se peut la réunion de personnels en interdépendance hiérarchique. Les entretiens téléphoniques sont également envisageables; ils peuvent intervenir après une première interview menée en passation directe, et seulement afin de recueillir un complément d'information précis. S'ajoutent à ces techniques celles de l'observation non participante et de la constitution de monographies, description circonstanciée des composantes d'un milieu donné.

CONTEXTUALISATION DES DONNÉES

L'évaluateur ne saurait engager ses investigations sans, au préalable, procéder à une analyse configurationnelle du milieu dans lequel l'expérimentation se déroule. Faute d'une compréhension des interactions, interdépendances, réseaux, rapports de force, luttes qui structurent les relations entre différents acteurs d'une même sphère, ainsi que les relations entre les différentes sphères, il risque, au mieux, de faire preuve d'une grande naïveté dans l'interprétation de ses données et, au pire, d'être instrumentalisé ou manipulé par les personnes rencontrées selon les intérêts qu'elles défendent. Le « doute méthodologique » doit habiter toutes les démarches de l'évaluateur jusqu'au terme de son travail même si, au fur et à mesure, il saisira de mieux en mieux les composantes implicites et informelles de l'espace social localisé où il intervient. Pour ce faire, il lui faut organiser des entretiens individuels avec les acteurs-clés de l'expérimentation, souvent placés au plus haut de la hiérarchie de l'institution. Le but est

de les questionner sur leurs intentions de départ, leurs attentes par rapport au dispositif, leurs contraintes, leurs inquiétudes, mais aussi sur la nature des rapports, hiérarchiques ou non, entretenus avec les autres acteurs-clés. Beaucoup d'informations seront recueillies indirectement, c'est-à-dire à travers des commentaires ou propos accusateurs [8] que les uns formulent sur les autres. L'ensemble de ces informations n'a pour finalité que d'éclairer la réflexion de l'évaluateur et restera confidentiel.

L'ANALYSE DES DONNÉES COLLECTÉES

Que faire du matériau recueilli? Et tout d'abord, quel crédit accorder aux discours des personnes interviewées? Les risques d'imprécision, d'erreur (dans les dates, les noms, les lieux, etc.) existent, sans parler des fausses informations qui peuvent être volontairement délivrées. Un premier remède consiste à distinguer les différentes catégories de données. Le contenu d'un entretien est composé de faits objectifs, de description de pratiques et de points de vue subjectifs. Les deux premières composantes sont assez facilement vérifiables par le recoupement des entretiens menés dont le nombre ne doit pas être négligeable. Plus de 70 entretiens ont ainsi été menés auprès des dix premiers lycées ayant été accompagnés. Enfin, l'effet, dit de saturation, est un indice de la fiabilité des informations recueillies : il arrive, en effet, un moment où les questions posées suscitent des réponses semblables, et où l'apparition des réponses atypiques s'épuise. Ces réponses atypiques, parce que très originales ou porteuses de contradiction ou d'inco-

hérence par rapport aux réponses les plus fréquentes, sont très instructives car elles apportent un supplément d'information. Leur apparente contradiction s'explique bien souvent par la position singulière qu'occupe un interviewé par rapport aux membres de sa catégorie. Par exemple, les enseignants participent de manière générale assez peu aux C2A ; or, au sein d'un établissement, nous avons rencontré une professeure, Mme D., très fortement impliquée dans la C2A qui essayait, sans grand succès, de susciter la participation de ses homologues. Cette énigme a vite été élucidée, lorsque Mme D., professeure de ST2S (série sciences et technologies de la santé et du social), a expliqué qu'elle avait exercé la profession d'assistante sociale pendant treize ans, avant d'entrer à l'Éducation nationale.

Il n'est pas envisageable de présenter ici l'ensemble des méthodes d'analyse du matériau recueilli ; deux d'entre elles seront retenues car elles s'avèrent particulièrement opérationnelles. Il s'agit de la démarche comparative et de la construction de typologies.

Les comparaisons de données de même catégorie sont à établir dans le temps et dans l'espace. Les comparaisons synchroniques permettent en l'occurrence de mettre en regard les différents degrés d'activité des C2A (activité estimée en fonction de plusieurs indicateurs) des lycées considérés au cours d'une même période. Les comparaisons diachroniques offrent la possibilité d'évaluer les évolutions d'activité des C2A à un an d'intervalle.

Par ailleurs, malgré la non-calculabilité de certaines dimensions comme « les effets structurants de l'accompagnement sur le fonctionnement des C2A », une typologie de ces effets a pu être constituée. Comme

Dominique Schnapper [7] l'indique, ce traitement des données passe tout d'abord par un classement « par tas » du matériau empirique. Cette mise en ordre s'opère par regroupements thématiques progressifs. Les « tas » constitués sont à nouveau réorganisés au sein de *types* qui ne relèvent non pas d'une catégorisation des performances d'équipes (active, dynamique, passive, etc.), mais d'un ordonnancement de différents types d'actions perpétrées par le consultant sur des publics cibles. C'est ainsi qu'ont pu être identifiés plusieurs types d'effets « rationalisateurs » produits par le dispositif expérimental sur le fonctionnement des C2A (cf. tableau ci-dessous).

CONCLUSION

Pour finir, il convient d'insister sur la nature des résultats délivrés par l'évaluateur à l'issue de l'expérimentation. Les préconisations ne prendront pas la forme de conclusions définitives et d'orientations unidirectionnelles, mais mettront en évidence un éventail de causes dont la combinaison est susceptible, dans un contexte donné, d'améliorer l'efficacité du traitement des problèmes sociaux ciblés. On évitera ainsi de

sacrifier au travers méthodologique qu'Olivier de Sardan nomme « une réduction à un facteur unique ». Cette erreur d'interprétation tend à « réduire n facteurs, empiriquement observables et pouvant tous jouer un rôle dans l'« explication » d'une situation sociale locale, à *un seul* d'entre eux. (...) La complexité de la vie sociale et des interactions rend finalement rares les contextes où un facteur et un seul, peut être invoqué. Et pourtant, la « monofactorisation », autrement dit le fait de ne proposer à un processus social donné qu'une seule « cause » ou « origine » là où l'expérience de la recherche comme l'histoire des sciences sociales devraient inciter à plaider la multifactorialité, est loin d'avoir disparu du panorama des sciences sociales. » [6].

Quant à la généralisation des dispositifs testés, l'évaluateur se prononcera avec prudence, tant la réussite d'un transfert de « bonnes pratiques » d'un complexe espace-temps à un autre est peu assurée. Il est plus réaliste d'identifier des modes de transmission suffisamment souples permettant à de nouveaux acteurs de s'approprier ces « bonnes pratiques » et de les adapter aux spécificités de leur milieu. ■

Typologie des effets « rationalisateurs » de l'accompagnement sur les C2A

Type de rationalisation des activités	Effets opératoires produits
Communicationnelle	Meilleure diffusion et partage de l'information, effet cohésif sur l'équipe, etc.
Organisationnelle	Structuration spatiale et temporelle du travail, calendrier, espace réservé aux C2A, etc.
Opérationnelle	Mise en place de nouvelles actions, redéfinition d'opérations, procédures, etc.
Fonctionnelle	Définition ou redéfinition des rôles et fonctions respectifs des membres des C2A
Instrumentale	Construction ou reconfiguration d'outils, fiches, formulaires, dossiers informatisés, etc.
Directionnelle	Ciblage des publics prioritaires, définition des décrocheurs et des absentéistes, etc.

BIBLIOGRAPHIE

- [1] **Divay Sophie**, 2011, L'évaluation qualitative : objectiver sans compter, *Bref*, n° 286.
- [2] **Haut Commissariat à la Jeunesse**, 2009, *Guide Méthodologique pour l'évaluation des expérimentations sociales*, Paris.
- [3] **Lemistre Philippe**, 2010, L'évaluation en sciences sociales est-elle un mythe? Le cas des politiques éducatives, in Thomas Couppié, Dominique Epiphane *et alii*, *Évaluation et données longitudinales : quelles relations ? XVII^{es} journées d'étude sur les données longitudinales dans l'analyse du marché du travail*. Céreq, Relief n° 30, mai 2010, pp.111-121, Marseille.
- [4] **Le Bianic Thomas, Vion Antoine**, 2008, Introduction générale, in Le Bianic Thomas, Vion Antoine, *Action publique et légitimités professionnelles*. LGDJ, Lextenso éditions, pp. 11-35, Paris.
- [5] **Mauger Gérard**, 1991, Enquêter en milieu populaire, *Genèses*, n° 6, pp. 125-143.
- [6] **Olivier de Sardan Jean-Pierre**, 2008, *La rigueur du qualitatif. Les contraintes empiriques de l'interprétation socio-anthropologique*. Bruylant-Acadamia, pp. 272-273, Louvain-La-Neuve.
- [7] **Schnapper Dominique**, 1999, *La compréhension sociologique. Démarche de l'analyse typologique*, PUF, Paris.
- [8] **Schwartz Olivier**, 1993, L'empirisme irréductible, Postface à Anderson Niels, Le Hobo. *Sociologie du sans-abri*, Nathan, Paris.

Typologie des élèves de Poitou-Charentes âgés de 11, 13 et 15 ans.

Valorisation régionale de l'enquête internationale HBSC.

L'enquête HBSC Poitiers a été menée en 2007 dans les écoles élémentaires, collèges et lycées de cette académie auprès de 2 237 élèves âgés de 11, 13 et 15 ans. Le rectorat de Poitiers avait pour objectif principal d'orienter la politique de prévention et d'éducation pour la santé à partir d'indicateurs fiables et actualisés sur les comportements de santé des jeunes et d'éviter des actions basées sur les représentations des adultes. Au-delà des analyses descriptives, les méthodes statistiques utilisées ont permis de constituer une typologie d'élèves réparties en cinq groupes homogènes. Les individus ainsi regroupés partagent des caractéristiques, comportements ou ressentis communs même s'ils ne sont pas identiques.

Stéphane Robin

Chargé d'études à l'observatoire régional de la santé Poitou-Charentes

Marie-Thérèse Roux

Médecin scolaire conseillère auprès du recteur de l'académie de Poitiers, rectorat de Poitiers

Julie Debarre

Chargée d'études, statisticienne à l'observatoire régional de la santé Poitou-Charentes

Pr Pierre Ingrand

Professeur de santé publique et bio-statistique, directeur de l'Institut universitaire de santé publique de l'université de Poitiers, président de l'observatoire régional de la santé Poitou-Charentes

Pr Daniel Marcelli

Professeur de psychiatrie de l'enfant et de l'adolescent à la faculté de médecine de Poitiers, chef du service de pédopsychiatrie au CHU de Poitiers

Dr Emmanuelle Godeau

Docteur en médecine, adjointe au médecin conseiller du recteur de l'académie de Toulouse, investigatrice principale de l'enquête HBSC en France, université de Toulouse

L'enquête *Health Behaviour in School-aged Children* (HBSC) est une enquête internationale, conduite tous les 4 ans depuis 1982 sous l'égide du bureau régional européen de l'Organisation mondiale de la santé (OMS). Elle vise à connaître l'état de santé et le bien-être des élèves âgés de 11 ans, 13 ans et 15 ans, les comportements de santé et leurs déterminants, ainsi que les différents environnements dans lesquels ils évoluent (famille, école, amis, relations amoureuses, etc.). Reconduite périodiquement, cette enquête permet de suivre les évolutions des comportements dans le temps et de comparer les données nationales avec l'ensemble des pays européens, et plus largement

du monde occidental. Pour cela, un protocole de recherche commun s'est imposé à l'ensemble des pays participants. En France, c'est le service médical du rectorat de Toulouse¹ et l'association pour le développement d'HBSC France qui assurent et coordonnent la mise en place des enquêtes selon un protocole de recherche commun, en collaboration avec l'INSERM U1027. Cette équipe a également assuré la coordination

NOTE

1. Les docteurs Félix Navarro, médecin conseiller du recteur, et Emmanuelle Godeau, adjointe au médecin conseiller du recteur de l'académie de Toulouse assurent la coordination de l'enquête HBSC en France depuis 1998. Le Dr Godeau en est l'investigatrice principale depuis 2002.

des enquêtes régionales qui ont eu lieu en 2007 dans les académies de Toulouse, Strasbourg, Nancy-Metz et Poitiers. Les procédures utilisées pour ces échantillons régionaux sont issues de celles appliquées à l'échantillon national HBSC 2006, elles-mêmes reposant sur les recommandations internationales. Dans ces quatre académies, l'équipe HBSC France a également apporté une aide méthodologique pour le nettoyage des bases de données, la construction de certains indicateurs synthétiques ; elle a aussi accompagné les équipes régionales dans leur travail de traitement et d'analyse. Par ailleurs, dans le cadre d'un projet financé par l'Institut de recherche en santé publique (IReSP) dont l'objectif était d'échanger sur les méthodes employées, sur l'interprétation des indicateurs ainsi que sur la mise en place des enquêtes HBSC futures, plusieurs séminaires méthodologiques ont été organisés localement, faisant interagir les différents partenaires des quatre académies et des experts scientifiques selon les thèmes des séminaires².

OBJECTIFS DE L'ÉTUDE

Les analyses ont porté plus spécifiquement sur la satisfaction scolaire et le cursus scolaire, l'objectif étant double : il s'agissait d'une part, de connaître quels étaient les élèves qui avaient des rapports compliqués avec l'institution scolaire et des difficultés scolaires, d'en connaître plus précisément les caractéristiques, les

différents environnements et comportements ; d'autre part, également « d'alimenter » les professionnels de l'Éducation nationale, notamment ceux en lien avec les comités d'éducation à la santé et à la citoyenneté, afin qu'ils puissent mettre en place des actions de prévention ou d'éducation en se basant sur des informations fiables et actualisées.

Un groupe de travail académique a été constitué, composé du médecin conseillère technique auprès du recteur, de l'infirmière conseillère technique, de l'assistante sociale conseillère technique, du proviseur vie scolaire chargé de mission, des médecins conseillers techniques responsables départementales, du statisticien du rectorat, de la coordinatrice des comités d'éducation à la santé et à la citoyenneté, du professeur de santé publique de l'institut universitaire de santé publique de Poitiers, d'un chargé d'études de l'observatoire régional de la santé (ORS) de Poitou-Charentes, et d'un interne de santé publique, stagiaire à l'ORS à cette période.

En mettant en place cette étude, le rectorat de Poitiers avait pour objectif principal d'infléchir la politique de prévention à partir de l'analyse des comportements de santé des jeunes, et non en s'appuyant sur les représentations des adultes. L'objectif pour le rectorat était donc d'orienter la politique de prévention et d'éducation pour la santé à partir d'indicateurs fiables et actualisés sur les comportements de santé des jeunes et d'éviter des actions basées sur les représentations des adultes.

Pour atteindre cet objectif, trois axes de travail ont été développés :

- l'identification des groupes d'élèves à cibler en priorité dans le cadre d'une politique de prévention ;

- le renforcement des compétences des enseignants à la problématique de l'adolescence afin de renforcer leur confiance et leur légitimité à intervenir dans ce domaine ;

- l'accroissement de la cohérence des adultes à proximité des jeunes, au sein des équipes éducatives, et en lien avec les parents.

MATÉRIEL ET MÉTHODES

Échantillonnage

Champ de l'enquête

Le champ de l'enquête est constitué d'élèves âgés de 11 ans, 13 ans et 15 ans, scolarisés dans un établissement sous tutelle du ministère de l'Éducation nationale, dans les secteurs public et privé sous contrat, dans l'académie de Poitiers au moment de l'enquête (janvier à mars 2007).

Constitution de l'échantillon

La méthode d'échantillonnage retenue est celle du sondage par grappe. Dans le premier degré, les grappes sont des écoles ; dans le second degré, les grappes sont des divisions. Les unités primaires (écoles dans le premier degré, divisions dans le second degré) ont été échantillonnées par sondage aléatoire simple stratifié.

Questionnaire

L'outil de recueil utilisé est un questionnaire autoadministré anonyme, rempli en classe sous la responsabilité d'un enquêteur extérieur à la classe. Il repose sur un tronc commun permettant les comparaisons internationales et sur un module optionnel de questions complémentaires. Chaque reconduction de l'enquête est l'occasion pour

NOTE

2. Appel à projet 2008 « Territoires et santé » de l'IReSP, financements DGS, MiRe-DREES, InVS.

les chercheurs d'HBSC de réexaminer le questionnaire et, le cas échéant, de supprimer des questions, de les améliorer ou d'en remplacer. Les nouvelles questions sont ensuite testées puis validées. Les thèmes abordés concernent l'environnement familial, les relations avec les pairs, le vécu scolaire, l'état de santé ressenti, les comportements sexuels, les violences, les activités quotidiennes, les activités physiques et sportives, la nutrition et les consommations de produits psychoactifs.

La possibilité de refuser que leur enfant participe est donnée aux parents par une lettre d'information (consentement passif) et aux élèves en début de passation. L'anonymat et la confidentialité sont garantis à toutes les étapes de la procédure. Ce protocole de recherche a fait l'objet d'une déclaration à la CNIL.

Méthodologie statistique

Afin d'atteindre les objectifs fixés, des méthodes d'analyses statistiques ont été définies. L'objectif final était de définir des profils d'élèves, c'est-à-dire des groupes homogènes d'individus, partageant des caractéristiques, des comportements et des environnements communs. Pour cela, différentes étapes ont été nécessaires :

– L'étape 1 a consisté à mener des analyses bi-variées par sexe, par classe d'âge et par niveau scolaire. Les tests statistiques employés sont le test du Chi2 lorsque les effectifs le permettaient pour les comparaisons de pourcentages (variables qualitatives) et une analyse de variance pour les comparaisons de moyennes (variable quantitative). Les résultats ainsi obtenus permettent de décrire plus précisément l'échantillon en fonction des caractéristiques, des différents environnements et des

comportements des élèves. C'est une étape descriptive indispensable à la meilleure connaissance de la population étudiée et de la base de données.

– L'étape 2 comporte en fait deux sous-étapes intermédiaires : une analyse factorielle des correspondances (AFC) et une classification ascendante hiérarchique.

En considérant que les élèves constituent les lignes d'un tableau et que les variables incluses dans le modèle sont les colonnes de ce même tableau (âge, sexe, niveau scolaire, environnement familial, environnement scolaire, comportements, etc.), l'AFC permet de repérer les lignes (les élèves) qui s'associent de la même manière aux différentes colonnes du tableau (les variables), c'est-à-dire, d'une part celles qui sont dépendantes statistiquement de cette colonne et d'autre part, celles qui sont indépendantes. En d'autres termes, l'AFC permet de réaliser une typologie dans laquelle on cherche les lignes dont la répartition s'écarte le plus de celle de l'ensemble de la population, celles qui se ressemblent entre elles et celles qui s'opposent, par rapport aux colonnes.

L'AFC permet donc de savoir en quoi les individus se ressemblent ou se différencient les uns des autres, en considérant qu'un individu est décrit totalement par les variables actives incluses dans le modèle. Les individus se ressemblent lorsque les valeurs pour chacune de ces variables sont proches. Les liaisons entre les variables sont étudiées à partir d'un nombre restreint de « variables synthétiques » obtenues par l'AFC, qui résument l'ensemble des informations de plusieurs variables liées entre elles. De plus, la classification va permettre de constituer une

typologie, c'est-à-dire de définir des groupes homogènes d'individus. Enfin, elle permet de montrer, pour chaque groupe constitué, en quoi les individus qui le constituent se ressemblent, autrement dit quelles sont les caractéristiques communes de ces individus.

Les variables prises en compte dans l'analyse sont des variables uniquement qualitatives, et leur choix a été réalisé de telle sorte que les différents champs du questionnaire soient représentés. Dans la première partie de l'analyse correspondant à l'analyse factorielle des correspondances multiples, des variables actives ont été définies, en particulier afin de prendre en compte le problème des réponses manquantes. Les axes factoriels sont ainsi caractérisés. La classification hiérarchique ascendante permet de trouver un optimum du nombre de groupes et, ensuite, elle permet d'interpréter les données à partir des variables qui ont servi à classer les individus.

L'analyse a porté sur soixante variables, et ces variables avaient plusieurs modalités, ce qui a représenté plus de 200 modalités au total. En prenant en compte ce grand nombre de modalités, l'analyse factorielle de correspondance multiple met en évidence que trois axes factoriels permettent de décrire l'ensemble des données. La deuxième étape permet de définir à partir de la classification hiérarchique ascendante le nombre de groupes qu'il faut retenir. Il n'existe pas de critère unique pour choisir un nombre de groupes pertinent mais des aides reposant sur des indicateurs statistiques. L'idée de cette démarche est de constituer des groupes qui soient les plus homogènes possibles à l'intérieur des

Description de quelques indicateurs

FAS (Family Affluence Scale)

L'indicateur est construit à partir de quatre questions :

- Combien d'ordinateurs possède ta famille ?
- Est-ce que ta famille a une voiture ?
- Est-ce que tu as une chambre pour toi tout(e) seul(e) ?
- Durant les 12 derniers mois, combien de fois as-tu voyagé avec ta famille pour partir en vacances ?

L'échelle ainsi constituée permet de classer les familles selon trois niveaux.

Échelle de Cantril

L'échelle de Cantril permet à l'élève de se situer sur une échelle allant de 0 à 10, 0 étant la pire vie possible et 10 la meilleure vie possible. À partir de 6, l'élève est plutôt satisfait de sa vie.

Syndrome de plainte

L'indicateur est construit à partir de huit questions : Durant les six derniers mois, tous les combien as-tu ...

- ... eu mal au ventre ?
- ... eu mal à la tête ?
- ... eu mal au dos ?
- ... eu des difficultés à t'endormir ?
- ... eu des étourdissements ?
- ... été déprimé ?
- ... été irritable ou de mauvaise humeur ?
- ... été nerveux ?

Un syndrome de plaintes est repéré lorsque deux symptômes de cette liste sont déclarés à une fréquence au moins égale à 2 dans la semaine.

Usage régulier

Avoir consommé au moins 10 fois un produit (alcool ou cannabis) dans les 30 derniers jours.

Soutien entre élèves

Indicateur construit à partir de trois items :

- les élèves de ma classe ont du plaisir à être ensemble ;
- la plupart des élèves de ma classe sont gentils et prêts à m'aider ;
- les autres élèves m'acceptent comme je suis.

Classement des répondants en trois catégories (soutien fort, moyen, faible)

groupes mais aussi les plus différents possible entre les groupes. Selon les objectifs définis, soit le choix d'un grand nombre de groupes est privilégié pour une analyse fine, soit au contraire un petit nombre de groupes pour une analyse la plus synthétique possible. C'est cette deuxième option qui a été retenue.

LES RÉSULTATS

Statistiques descriptives

La population

Sur les 2 809 élèves sélectionnés, 7,8 % n'ont pas participé à l'enquête suite à un refus des parents (218 élèves), 2,2 % ont refusé de

participer de leur propre initiative (63 élèves), et 246 étaient absents (8,8 %). Ainsi, finalement 2 282 élèves ont rempli le questionnaire mais 38, mal remplis ont été exclus lors de la saisie et 7 autres supprimés lors du nettoyage de la base. L'analyse des données a donc concerné 2 237 élèves, dont 695 âgés de 11 ans, 796 âgés de 13 ans et 745 âgés de 15 ans (dans un cas, l'âge n'est pas renseigné). Près des trois quarts de l'échantillon sont scolarisés au collège, 6 % en classe de CM2 et un cinquième au lycée.

La structure familiale

Les familles dites nucléaires, dans lesquelles vivent ensemble le père et la mère avec leurs enfants, sont très majoritaires. Près des trois quarts des élèves ont déclaré vivre dans ce type de famille. Plus d'un élève sur dix vit dans une famille monoparentale et la même proportion vit dans une famille recomposée.

La situation financière

Peu d'élèves pensent que leur famille n'est pas à l'aise financièrement (moins de 10 %) tandis que près de 15 % pensent qu'elle est très à l'aise. Ces derniers sont moins nombreux parmi les plus âgés.

L'indicateur FAS (*cf. encadré*) qui permet également d'estimer de façon plus indirecte l'aisance financière confirme également que moins d'un élève sur dix (8 %) vivrait dans une famille peu aisée.

Les relations familiales

La communication avec les différents membres de la famille est facile pour une grande partie de l'échantillon avec des différences selon le sexe, les filles la trouvant plus souvent compliquée que les garçons. Des différences sont

également observées selon le niveau scolaire et l'âge : les élèves des lycées semblent avoir le plus de difficultés de communication avec les membres de leur famille. Ce sont les relations avec le père qui sont les plus difficiles, notamment pour les filles qui sont une majorité dans ce cas. Les relations avec les beaux-parents sont encore plus problématiques. Les filles ont plus de difficultés avec leur beau-père tandis que les garçons ont des relations un peu plus problématiques avec leur belle-mère.

La satisfaction personnelle

La satisfaction de l'élève vis-à-vis de sa vie est estimée par le biais de l'échelle de Cantril. Moins d'un élève sur cinq a une satisfaction amoindrie (score < 6/10), cette proportion étant plus importante chez les filles. Le ressenti est moins favorable chez les plus âgés.

La santé ressentie

Lorsqu'on leur pose directement la question sur leur état de santé perçu, presque 9 élèves sur dix s'estiment en bonne, voire en excellente santé, les garçons un peu plus fréquemment que les filles, et les plus jeunes plus fréquemment que les plus âgés. Une série de questions a permis d'identifier les élèves souffrant d'un syndrome de plaintes (*cf. encadré*). Plus d'un tiers des élèves est concerné par des plaintes somatiques fréquentes, les filles plus fréquemment que les garçons.

La satisfaction scolaire

Au moins un cinquième des élèves déclare aimer beaucoup l'école et 46 % l'aimer un peu. Plus d'un élève sur dix ne l'aiment pas du tout. L'appréciation scolaire décroît avec l'âge : alors qu'à 11 ans, 38 % l'aiment beaucoup et à

peine 6 % ne l'aiment pas du tout, ils ne sont qu'un sur dix à l'aimer beaucoup à 15 ans et presque 14 % à ne pas l'aimer du tout. Les filles répondent plus fréquemment que les garçons aimer beaucoup l'école (25,2 % contre 16,9 %) et deux fois moins souvent ne pas l'aimer du tout (7,1 % contre 15,5 %).

Le jugement est plus favorable dans l'élémentaire qu'au collège, puis qu'au lycée, avec une dépréciation plus importante entre l'élémentaire et le collège.

- *Les relations entre élèves*

Les relations entre élèves sont étudiées à travers une série de trois questions : le plaisir d'être ensemble, l'entraide, l'acceptation par les autres élèves. Un peu plus d'un élève sur dix bénéficierait d'un faible soutien de la part des autres élèves, plus de la moitié d'un soutien moyen et un peu moins d'un tiers d'un soutien fort.

- *Les résultats scolaires*

Un peu plus d'un élève sur dix pense que ses résultats sont mauvais par rapport à ses camarades. La plus grande partie (43 %) pense qu'ils sont moyens, un peu plus d'un tiers qu'ils sont bons et un sur dix qu'ils sont très bons, avec une dégradation de l'appréciation chez les plus âgés.

Ces premiers résultats partiels permettent de décrire l'échantillon obtenu mais sont peu opérationnels par rapport aux objectifs fixés. Ils constituent malgré tout une étape indispensable à une meilleure compréhension et connaissance de la population étudiée. La deuxième étape permet, par des méthodes statistiques adaptées, (*cf. méthodologie statistique*) d'obtenir des groupes rassemblant des élèves dont les caractéristiques, ressentis et comportements

se ressemblent, sans qu'ils soient tous identiques.

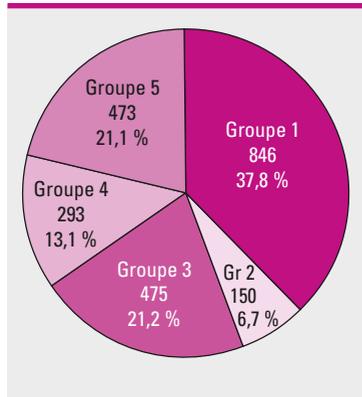
Les profils d'élèves

Une mise en garde est nécessaire avant la lecture de ces résultats. Comme indiqué dans la méthodologie, il s'agit de regrouper des individus selon les informations disponibles en considérant que ces dernières permettent de décrire au mieux l'individu. Or, certaines informations sont manquantes car les domaines abordés dans l'enquête sont limités. De plus, il s'agit de données déclaratives exprimant le ressenti de l'élève et non pas de données objectives. Enfin, les individus sont rassemblés dans des groupes qui sont décrits par rapport à un écart observé ou non par rapport à la moyenne observée dans l'échantillon. Les individus d'un groupe ne sont pas identiques, ils se ressemblent sur certains éléments mais peuvent s'éloigner sur d'autres. Il s'agit donc de ne pas céder à la facilité de la stigmatisation qui constitue un risque important de ces méthodes d'analyse.

Cinq groupes d'élèves ont été obtenus. Le groupe 1 est le plus important avec plus d'un tiers des élèves, puis deux groupes sont constitués pour chacun d'un cinquième d'entre eux. Enfin, le groupe 4 représente plus de 10 % des élèves et le groupe 2, le plus petit, moins de 7 % (*cf. graphique 1*).

Pour chacun des groupes obtenus, nous avons invité un pédopsychiatre du centre hospitalier Henri Laborit de Poitiers à nous apporter un angle de lecture différent des résultats. Le professeur Daniel Marcelli a mené de nombreux travaux scientifiques sur la problématique des adolescents et sur les liens avec la scolarité. Il est également très impliqué sur la thématique

Graphique 1 – Répartition des élèves par groupes (effectifs, pourcentages)



de l'accueil et de l'accompagnement des adolescents au sein des établissements scolaires.

L'âge constitue un élément important dans la constitution des groupes : deux groupes dans lesquels les élèves âgés de 11 ans sont plus nombreux que dans l'ensemble (groupes 2 et 3), deux autres plutôt âgés de 13 ans (groupes 1 et 4) et un dernier dans lequel la proportion de ceux âgés de 15 ans est très importante. Un peu plus de la moitié des élèves sont situés dans des groupes qui cumulent les indicateurs favorables dans tous les domaines (groupes 1 et 3). Une petite minorité est en situation moyenne : ils vivent plus souvent dans une famille

monoparentale ou recomposée et s'ils déclarent être en bonne santé, cette affirmation est contredite par des indicateurs indirects (plaintes somatiques, satisfaction de la vie) qui montrent une situation plus mitigée (groupe 2). Les deux derniers groupes cumulent les situations défavorables : situation sociofamiliale, vécu scolaire, violences, indicateurs de santé et consommations de produits psychoactifs chez les plus âgés (groupes 4 et 5).

Le groupe 1 (graphique 2)

Ce groupe est composé majoritairement de filles et plus fréquemment d'élèves âgés de 13 ans. La proportion d'élèves vivant dans une famille nucléaire est plus importante dans ce groupe qu'en moyenne. La communication avec les membres de la famille est un peu plus fréquemment jugée comme étant facile, avec un contrôle et un intérêt parental important. Les élèves de ce groupe considèrent plus souvent que leur famille est moyennement à l'aise financièrement. De plus, les statuts économiques et sociaux du père comme de la mère sont un peu plus fréquemment moyens et élevés qu'en moyenne.

Ce groupe est caractérisé par un dia-

logue plus fréquemment difficile avec les amis du sexe opposé. Ces élèves déclarent moins souvent sortir le soir ou après l'école. Le soutien qu'ils perçoivent entre élèves est plus fréquemment moyen (74 % contre 57 %).

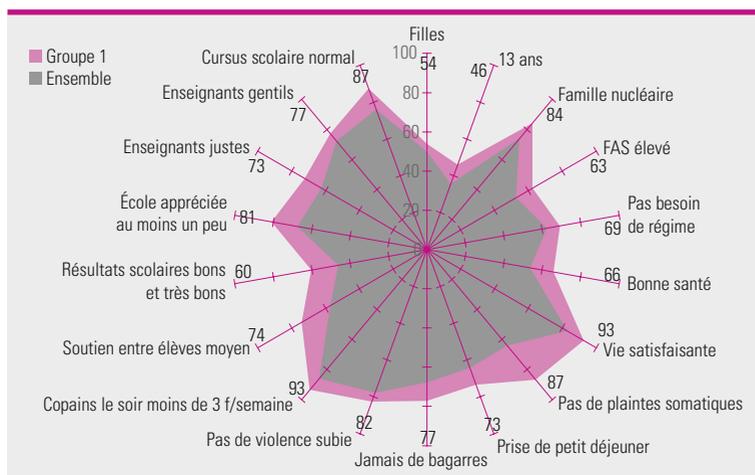
Ils estiment plus que les autres avoir des résultats scolaires très bons ou bons. Ils déclarent plus souvent aimer au moins un peu l'école mais c'est surtout la proportion des élèves qui n'aiment pas du tout l'école qui est moins importante. Les relations avec les enseignants sont plutôt favorables : ils sont plus souvent jugés justes, gentils, tandis que le travail scolaire est un peu moins fréquemment ressenti comme étant fatigant ou difficile, qu'au niveau global.

Sur l'ensemble des indicateurs de santé ressentie et de qualité de vie, les résultats montrent également une situation favorable pour ce groupe par rapport au niveau global. Ils déclarent plus fréquemment ne pas avoir besoin d'un régime alimentaire et se considérer du bon poids. Ils sont également plus nombreux à déclarer prendre un petit déjeuner tous les jours. Ils pratiquent plus fréquemment une activité sportive soutenue (4-6 heures par semaine).

Les niveaux de consommation de produits psychoactifs sont inférieurs à ce qui est observé au niveau global (tabac, expérimentation de cannabis, ivresses), sauf pour l'expérimentation de l'alcool qui est comparable au niveau global.

Ces élèves sont plus nombreux à n'avoir jamais été victimes de violences à l'école et ils sont moins fréquemment les auteurs de violences ou de brimades.

Graphique 2 – Distribution des élèves du groupe 1 pour quelques indicateurs (%)



Le regard du professeur Marcelli³

« Cet ensemble relativement homogène de jeunes adolescents autour

de 13 ans, qui s'appuient sur leurs parents, ont de bonnes relations avec leurs proches, ont un bon investissement dans leur scolarité. Ils ne consomment pas trop de produits et ont des habitudes de vie favorables. Ils veulent avoir de bonnes relations avec les pairs mais ne savent pas comment s'y prendre, en particulier avec ceux du sexe opposé. Ces jeunes hésitent encore à sortir du cocon familial, bien qu'ils y soient à l'aise. On voit bien ici l'écart entre les indicateurs de bien-être et ceux qui marquent la difficulté des relations avec les pairs, ce qui témoigne de la poussée de l'adolescence. Ce sont typiquement les questions que se posent des 13-14 ans. »

Le groupe 2 (graphique 3)

Dans ce groupe, les élèves âgés de 11 ans sont les mieux représentés (44 % contre 31 % au niveau global). Les familles monoparentales (un quart) et recomposées (un cinquième) sont plus représentées qu'au niveau global. La communication dans la famille est plus fréquemment difficile,

avec un contrôle et un intérêt parental comparables à la moyenne observée. Les élèves de ce groupe jugent plus souvent que leur famille est plutôt située à un niveau économique moyen ou bas selon l'indicateur FAS.

Le dialogue avec les amis du sexe opposé est très difficile pour plus de huit élèves sur dix ce qui est largement au-dessus de la moyenne (moins de 20 %) ; c'est également le cas avec le meilleur ami. Dans ce groupe, plus d'un tiers déclarent avoir moins de trois amis du même sexe et les sorties le soir ou après l'école sont moins fréquentes qu'au niveau global.

À l'école, les résultats obtenus sont plus fréquemment jugés comme étant moyens. Sur les autres indicateurs de scolarité, la situation est comparable au niveau global mais un pourcentage plus important d'entre eux trouvent les enseignants gentils. Un élève sur dix déclare subir des brimades plusieurs fois par semaine ce qui est plus fréquent qu'au niveau global. Ils sont plus nombreux à rester devant la télévision ou sur Internet moins de deux heures par jour. Leurs comportements alimentaires ne diffèrent pas de ce qui est observé au

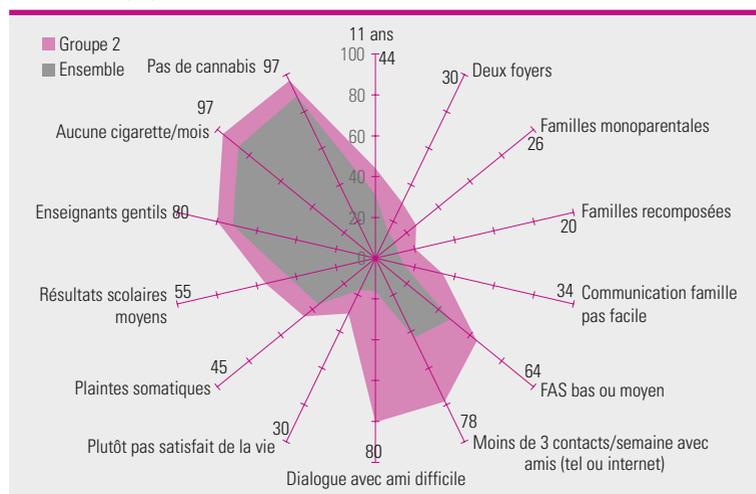
niveau global. Les indicateurs de santé indirects sont défavorables (satisfaction de la vie, plaintes somatiques) mais leur jugement sur leur état de santé est comparable à ce qui est observé au niveau global. Les niveaux de consommation de produits psychoactifs sont inférieurs à ce qui est observé au niveau global, pour ce qui concerne le tabac et l'expérimentation du cannabis et de l'ivresse.

En revanche, l'expérimentation de l'alcool est comparable à ce qui est observé au niveau global.

Le regard du professeur Marcelli

« Ce groupe est potentiellement inquiétant. Dès l'âge de 11 ans, on note un début de consommations d'alcool. On sait que l'alcool est le premier produit consommé, avant le tabac et le haschisch mais on sait aussi que plus cette consommation est précoce, plus elle annonce les suivantes. Les familles sont moins « contenantes » que les autres, une fragilité au niveau émotionnel est à noter. Les indicateurs de santé sont moins bons. Ces jeunes, à l'orée de l'adolescence, ont eu dans l'enfance des événements de vie négatifs et vécus des fragilités environnementales et psychologiques. On est en droit de penser que, lorsqu'ils auront 15 ans, ils feront partie du groupe inquiétant (cf. groupe 5). Ce petit groupe arrive à l'adolescence avec un bagage qui le handicape. »

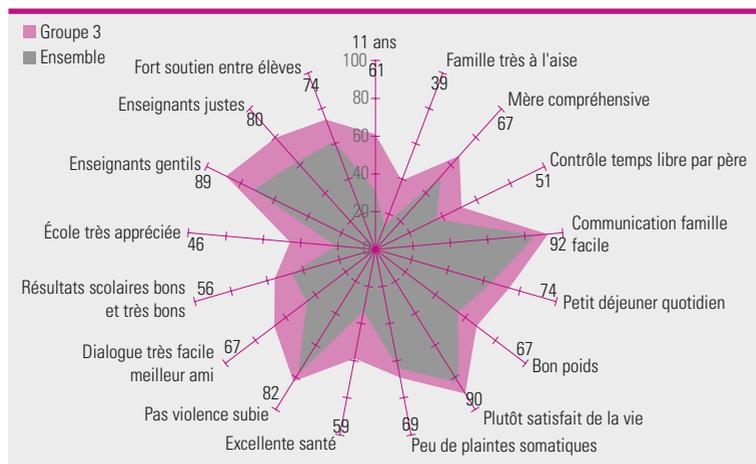
Graphique 3 – Distribution des élèves du groupe 2 pour quelques indicateurs (%)



NOTE

3. Le professeur Daniel Marcelli est pédo-psychiatre, chef du service de psychiatrie infanto-juvénile du CH Henri Laborit de Poitiers.

Graphique 4 – Distribution des élèves du groupe 3 pour quelques indicateurs (%)



Le groupe 3 (graphique 4)

Les élèves âgés de 11 ans sont également fortement représentés dans ce groupe.

Ils jugent plus souvent que leur famille est très à l'aise financièrement, tandis que les autres indicateurs décrivant la situation socioéconomique ne diffèrent pas de la moyenne. La communication dans la famille est jugée facile pour plus de neuf élèves sur dix, et le contrôle et l'intérêt des parents, de la mère comme du père, sont plus forts qu'au niveau global.

Ces élèves ressentent un fort soutien

de leurs camarades de classe pour les trois quarts d'entre eux, et le dialogue avec le meilleur ami et avec les amis du sexe opposé est plus fréquemment très facile qu'au niveau global. Ils sont moins souvent confrontés aux brimades et aux violences.

Les résultats scolaires sont plus fréquemment « très bons » que pour l'ensemble de l'échantillon, et ils sont plus souvent à l'aise à l'école, vis-à-vis des indicateurs de l'organisation scolaire et des relations avec les enseignants mais également par rapport au travail scolaire (stress, fatigue liée au travail scolaire).

Près de la moitié d'entre eux déclare aimer beaucoup l'école (contre un cinquième au niveau global).

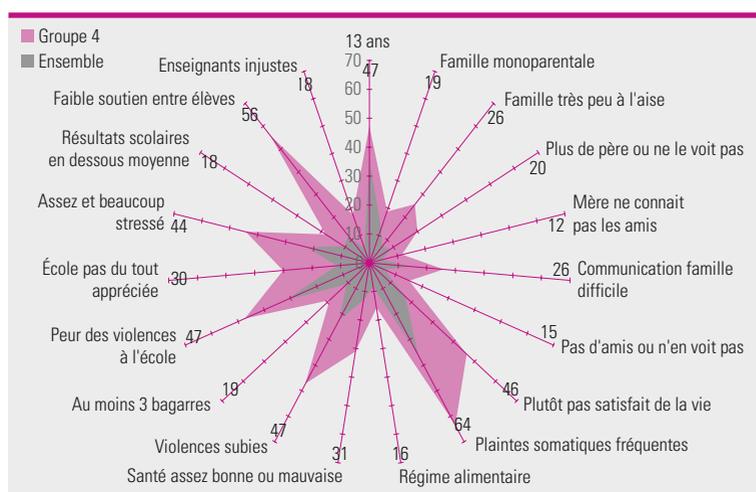
Dans ce groupe, l'activité physique est plus souvent satisfaisante et les élèves passent moins de temps devant la télévision ou sur Internet. Ils se perçoivent plus souvent au bon poids qu'au niveau global et déclarent plus fréquemment ne pas avoir besoin de suivre un régime. La proportion d'élèves prenant quotidiennement un petit déjeuner est plus importante qu'en moyenne.

Les indicateurs de santé et de qualité de vie sont plus favorables qu'au niveau global. Par exemple, la moitié d'entre eux s'estiment en excellente santé. En termes de consommation de produits psychoactifs, les niveaux observés sont moindres qu'au niveau global.

Le regard du professeur Marcelli

« Pour des enseignants, c'est le groupe d'enfants idéal mais ce ne sont pas encore des adolescents. Ce sont des grands enfants avec un environnement porteur et qui vont bien. »

Graphique 5 – Distribution des élèves du groupe 4 pour quelques indicateurs (%)



Le groupe 4 (graphique 5)

Ce groupe est constitué à près de la moitié d'élèves de 13 ans.

Ils vivent moins fréquemment au sein d'une famille nucléaire qu'au niveau global et un peu plus fréquemment dans une famille monoparentale. Il s'agit plus fréquemment de familles en difficultés financières (plus du quart), vivant en situation économique défavorable ou moyenne (indicateur FAS). La communication dans la famille est ressentie comme difficile pour un

quart d'entre eux. Plus fréquemment qu'au niveau global les pères ne savent rien des amis ou des activités libres.

Les relations avec les pairs ne sont pas satisfaisantes : près d'un tiers déclare ne pas aimer être avec les autres élèves, considérant que les élèves ne sont pas gentils pour plus de la moitié d'entre eux. Ils rapportent plus souvent ne pas se sentir acceptés comme ils sont. Ils participent plus souvent à des bagarres et en sont plus souvent les victimes. Ils sont plus nombreux à subir des brimades répétées et en sont également plus fréquemment les auteurs.

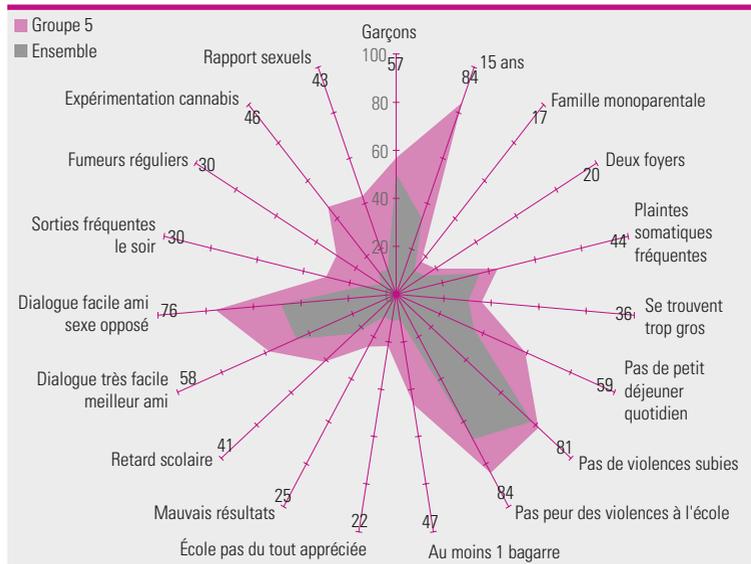
Leur relation à l'école est plus compliquée qu'en moyenne : Ils considèrent plus souvent que les enseignants ne sont pas justes et les résultats scolaires sont beaucoup plus souvent en dessous de la moyenne qu'au niveau global. Ils déclarent trois fois plus souvent ne pas aimer du tout l'école. Le travail est plus fréquemment jugé difficile, fatigant et générateur de stress.

Dans ce groupe, ils sont presque deux fois plus nombreux qu'au niveau global à être non sportifs. Ils consomment plus fréquemment des sodas ou des sucreries tous les jours et sont moins nombreux à prendre un petit déjeuner le matin. La moitié d'entre eux se trouvent trop gros. Un quart d'entre eux avouent ne pas faire de régime mais en avoir besoin tandis que 16 % en font un au moment de l'enquête. Les indicateurs de santé mettent en évidence une situation dégradée, que ce soit sur l'état de santé global ressenti, la qualité de vie ou les plaintes somatiques.

Le regard du professeur Marcelli

« Ce sont les adolescents de 5^e qui sont dans un environnement pas ou

Graphique 6 – Distribution des élèves du groupe 5 pour quelques indicateurs (%)



peu soutenant, avec des situations difficiles, des contextes environnementaux tout aussi difficiles. Ils ressentent la société comme hostile et négative, pour des raisons justifiées ou pas. On peut craindre que ce groupe évolue à 15 ans vers des polyconsommations. »

Le groupe 5 (graphique 6)

Un cinquième de l'échantillon des élèves est dans ce dernier groupe. Il est constitué à près de 85 % d'élèves âgés de 15 ans avec une proportion plus importante de garçons (57 %). Ces élèves vivent moins fréquemment au sein d'une famille nucléaire qu'au niveau global et plus fréquemment dans une famille monoparentale. Il s'agit plus fréquemment de familles moyennement à l'aise financièrement.

Le dialogue avec les amis est considéré comme très facile, que ce soit pour les amis du même sexe ou non, et ils les rencontrent souvent après l'école ou le soir.

Les résultats scolaires sont déclarés beaucoup plus fréquemment en

dessous de la moyenne qu'au niveau global ou moyen, et ils sont plus souvent en retard scolaire. Ils déclarent plus que les autres ne pas aimer du tout l'école ou pas trop.

Ils n'ont pas peur des violences et participent plus fréquemment à des bagarres. Ils sont également plus souvent brimeurs qu'au niveau global et sont plus nombreux à déclarer ne jamais subir de violences ou de brimades.

La proportion d'élèves qui ne prennent pas quotidiennement un petit déjeuner est beaucoup plus importante qu'au niveau global. Ils sont également plus nombreux à consommer des sodas ou des sucreries tous les jours, et à se considérer trop gros. Un quart d'entre eux ne font pas de régime mais estiment qu'ils en ont besoin. Ils pratiquent plus souvent le sport de manière intensive.

Les indicateurs de santé sont plutôt dégradés : ils sont plus nombreux à se sentir en assez bonne santé et ont plus souvent des plaintes somatiques fréquentes. Cependant, leur niveau de satisfaction concernant la vie est

comparable à la moyenne. Enfin, les niveaux de consommation et d'expérimentation sont beaucoup plus importants qu'au niveau global, que ce soit pour l'alcool, le tabac ou le cannabis.

Le regard du professeur Marcelli

Dans ce cinquième groupe, on est confronté à des adolescents à risque. La bascule est assez brutale entre 13 et 15 ans : la famille n'est plus « contenant ». C'est avec les pairs que ces jeunes passent la majorité de leur temps. Bien sûr, comme l'école est un lieu contenant, ça ne leur plaît pas, pas plus que la famille : « je n'aime pas être entre 4 murs ». Un jeune qui va bien aime bien se bouger mais pas trop. On relèvera le cumul des facteurs à risque : plaintes somatiques, mauvaise hygiène de vie, bagarres, difficultés scolaires, fréquence des sorties, rapports sexuels, etc. Ce cumul montre la grande difficulté dans laquelle ces jeunes évoluent.

Synthèse des résultats

Cette typologie d'élèves a été réélaborée en intégrant l'ensemble des élèves quel que soit leur âge et cet élément a un poids important dans la constitution des groupes : un groupe est fortement constitué d'élèves âgés de 11 ans, un autre d'élèves de 13 ans, et enfin un dernier d'élèves de 15 ans, ce qui traduit indirectement la dynamique de l'adolescence et l'évolution rapide des comportements et des attitudes à cet âge. En revanche, il est à noter que le genre semble jouer un rôle moins important pour discriminer ces groupes, bien que le groupe 1 soit plus

particulièrement marqué par une plus forte proportion de filles et le groupe 5 un peu plus par les garçons.

Les groupes 1 et 3, qui représentent environ 60 % de la population étudiée, rassemblent les élèves qui vont bien. Les groupes 2 et 4 regroupent les élèves plutôt âgés de 11 et 13 ans qui sont à risque de déviances (20 %). Ces derniers auraient besoin d'un accompagnement par un adulte de confiance qui les rassure, les comprend et les considère avec bienveillance, ce que l'expert appellerait un tuteur de résilience. Pour ces jeunes, le problème est qu'ils n'ont pas de « contenant familial » et que cette absence est poursuivie à l'école où l'adulte ne joue pas non plus ce rôle, d'où une rupture plus ou moins importante avec l'encadrement et les enseignants.

DISCUSSION

Les cinq groupes obtenus se différencient les uns des autres mais pas sur la totalité des dimensions en même temps. Cependant, chacun d'entre eux partage des éléments propres qui le caractérisent.

Il faut prendre en compte également les limites propres à ce type d'approche.

Le premier point est son aspect exploratoire : ces groupes sont constitués à partir de données déclaratives à un moment donné et sont amenées à varier en fonction de l'histoire de l'élève.

Le deuxième point est que ces données restent parcellaires même si elles balayent un grand nombre de domaines. Or, les analyses partent du présupposé que l'individu est décrit dans sa globalité par les données disponibles recueillies par le biais de l'autoquestionnaire. De même, l'ensemble des

élèves n'a pas répondu à la totalité des questions ; or, dans ce type d'analyse, les non-réponses sont traitées comme si elles ne jouaient aucun rôle dans l'interprétation des résultats (c'est-à-dire qu'on considère que les élèves qui ne répondent pas, ne sont pas différents de ceux qui ont répondu).

Enfin, et c'est un point très important, les groupes rassemblent des individus qui se ressemblent mais qui ne sont pas identiques. Les élèves d'un groupe sont ainsi rassemblés par rapport à un ensemble de points communs mais d'autres éléments peuvent varier. Dans la description qui est faite de chaque groupe, il est mis en évidence des caractéristiques qui sont plus ou moins présentes dans le groupe par rapport à ce qui est observé en moyenne sur l'ensemble de l'échantillon, mais qui ne sont pas partagées par la totalité des élèves de ce groupe. Le danger est donc la généralisation abusive et la stigmatisation.

Une fois ces limites prises en compte, d'autres points constituent des avantages de poids par rapport aux objectifs fixés.

Premièrement, cette démarche apporte une vision synthétique des données recueillies. En effet, la difficulté face à une grande masse d'informations est de leur donner un sens, une grille de lecture fidèle, un moyen de résumer l'information et de mettre en évidence des ponts entre les différents éléments qui la composent. Cette méthode le permet.

Le deuxième point qui nous a conduit à utiliser cette méthode est la possibilité, par ce biais, d'identifier des groupes pour lesquels sont repérées des difficultés afin d'engager des actions adaptées.

Enfin, les résultats de cette typologie conduisent à se poser certaines

questions : comme dit plus haut, les individus formant un groupe se ressemblent mais ne sont pas identiques. Certains partagent nombre de caractéristiques mais d'autres différent. L'un des objectifs de l'étude étant également d'appréhender plus précisément la satisfaction scolaire, il apparaît dans certains groupes des proportions d'élèves n'aimant pas du tout l'école très importantes par rapport à la moyenne. Or, dans ces groupes, certains élèves, même s'ils sont minoritaires, déclarent se sentir bien à l'école. Peut-on expliquer qu'un individu qui partage un ensemble de caractéristiques, de

comportements ait un ressenti différent ? Est-ce que cela constitue une limite liée à l'individu, à son goût propre qui échapperait alors à un certain déterminisme ? On touche ici les limites des analyses quantitatives et on perçoit tout l'intérêt des analyses qualitatives qui viendraient les compléter.

I CONCLUSION

À partir de données issues d'une enquête autodéclarative, il est possible, par le biais de méthodes statistiques adaptées, d'obtenir une

grille de lecture rendant compte d'un ensemble de données abordant des domaines diversifiés. À partir de données de qualité, assurées dans cette étude par un protocole d'enquête éprouvé, il a été possible de relier un ensemble de caractéristiques mais également de comportements et, ainsi, de décrire au mieux les élèves dans leur globalité en les regroupant en fonction de leurs ressemblances et différences. Grâce au repérage de groupes cibles, les éléments de compréhension ainsi fournis ont une incidence opérationnelle dans le cadre de la mise en place de programmes de prévention adaptés. ■

BIBLIOGRAPHIE

Currie C., Nic Gabhainn S., Godeau E., Roberts C., Smith R. & Currie D. (2008) (Eds.) *Inequalities in young people's health: HBSC international report from the 2005/2006 Survey (Health Policy for Children and Adolescents, No. 5)*. WHO Regional Office for Europe, Copenhagen.

Godeau E., Arnaud C., Navarro F., (2008) *La santé des élèves de 11 à 15 ans en France/2006*. Études santé, Inpes, 271 p.

Robin S., Debarre J., (2008) *HBSC Poitiers 2007, Comportements de santé des élèves de 11, 13 et 15 ans de l'académie de Poitiers*, ORS Poitou-Charentes, 4 p.

Robin S., Debarre J., (2010) *HBSC Poitiers 2007, Typologie des élèves âgés de 11, 13 et 15 ans de l'académie de Poitiers*, ORS Poitou-Charentes, 4 p.

Robin S., Debarre J., (2010) *HBSC Poitiers 2007, Recherche thématique : les facteurs associés aux usages de l'alcool*, ORS Poitou-Charentes, 4 p.

Sites internet :

www.ors-poitou-charentes.org

www.hbsc.org

L'autoévaluation pour accompagner l'expérimentation et le projet d'établissement ?

Quelques pistes à partir d'expériences anglo-saxonnes

Romuald Normand

Laboratoire Triangle UMR 5206
Institut français de l'éducation
École normale supérieure de Lyon

La démarche d'autoévaluation de l'expérimentation et du projet d'établissement est peu développée en France. Pourtant, comme le montrent des expériences réalisées dans d'autres pays, elle peut constituer un levier de la dynamique de changement de l'établissement et d'amélioration de ses résultats. Après avoir décrit le contexte de mise en œuvre de l'autoévaluation en Écosse, Angleterre et Nouvelle-Zélande, cet article propose quelques pistes susceptibles d'aider le chef d'établissement à piloter le conseil pédagogique et à jeter les bases d'une nouvelle approche de l'évaluation de son établissement partagée avec ses équipes.

L'évaluation de l'établissement public local d'enseignement (EPL) en France, contrairement à d'autres pays, est encore assez peu développée.

Même si les établissements scolaires établissent aujourd'hui des contrats d'objectifs avec les autorités académiques, et que le projet d'établissement s'appuie sur un diagnostic, le chef d'établissement dispose d'assez peu de ressources et d'outils pour sensibiliser ses équipes pédagogiques à une culture de l'évaluation. Certes, des instruments existent, comme les indicateurs de pilotage des établissements secondaires (IPES) ou les évaluations nationales en français et en mathématiques, mais ils relèvent d'une approche de l'évaluation externe, et servent davantage au pilotage qu'à une réflexion sur la dynamique de changement de l'organisation pédagogique et des pratiques d'enseignement [1]. Pour remédier à cet inconvénient, une démarche d'autoévaluation, comme celle pratiquée dans d'autres pays, apparaît une solution intéressante pour améliorer la coordination d'activités souvent éclatées entre l'administra-

tif, l'éducatif et le pédagogique, mais aussi pour susciter une réflexion d'ensemble sur la définition et la mise en œuvre du projet d'établissement. L'autoévaluation peut permettre d'ouvrir « la boîte noire » de l'établissement scolaire en donnant à voir les logiques d'action et les valeurs conditionnant un véritable changement améliorant la réussite des élèves. Pour s'en convaincre, il suffit de prendre quelques exemples empruntés à la situation de pays déjà avancés en matière d'autoévaluation [2] : l'Écosse, l'Angleterre et la Nouvelle-Zélande. Même si, en France, ces démarches apparaissent relativement difficiles à mettre en œuvre à l'échelle de l'établissement, nous verrons qu'elles donnent à réfléchir à une meilleure articulation entre l'évaluation, le conseil pédagogique et les expérimentations mises en œuvre dans le cadre de l'article 34 de la loi d'orientation et de programme pour l'avenir de l'École du 23 avril 2005. L'autoévaluation de l'expérimentation peut constituer un levier du projet d'établissement et créer une dynamique collective, en suscitant une réflexion et un travail en équipes.

I L'ÉCOSSE

Comment savoir si notre établissement est efficace ?

Contrairement à une idée répandue, l'Écosse présente bien des similarités avec la France dans le domaine de l'éducation. Celles-ci reposent sur un fort sentiment d'identité nationale formé progressivement au travers des conflits avec l'Angleterre, le proche voisin et l'ennemi héréditaire. La société écossaise est fortement engagée en faveur de l'égalité des chances et d'un enseignement public démocratisant. Les enseignants sont attachés aux savoirs, la définition de ces derniers s'inscrivant dans la tradition et « l'héritage des Lumières ». La profession enseignante jouit d'une bonne reconnaissance auprès du public. Elle dispose d'une relative autonomie tout en combinant une adhésion assez forte à la vie syndicale et aux conceptions du service public. Les enseignants sont hostiles, comme leurs concitoyens, à l'introduction d'une logique marchande dans l'éducation et à l'ingérence du monde économique dans la question éducative. Avant la France, les responsables de l'éducation écossais ont organisé un grand débat national sur l'avenir de l'école et ils ont progressivement mis en place un socle de connaissances et de compétences à acquérir par les élèves. Une obligation de résultats s'est imposée au système éducatif qui vise l'amélioration de l'efficacité. Première différence avec la France, les résultats des établissements et les avis de l'inspection sont publics. Seconde différence, un accord assez large a été obtenu entre le gouvernement et les organisations syndicales pour

transformer le statut des enseignants. Ces derniers doivent 35 heures de service dans l'établissement moyennant un maximum de 22,5 heures dans la classe. Ils doivent satisfaire à une obligation de formation professionnelle continue. En contrepartie, ils ont obtenu une augmentation de salaire de 21 % en 3 ans.

Le ministère a fixé des objectifs au système éducatif pour faire entrer l'école écossaise dans le XXI^e siècle. Il propose aux acteurs de la communauté éducative un « voyage vers l'excellence » qui est illustré dans les documents officiels par un navigateur GPS. Les objectifs du système éducatif, qui visent à diversifier les modes d'éducation des élèves, sont organisés selon des principes d'excellence. Les enseignants doivent partager des valeurs avec les jeunes (culture de la réussite, mais aussi promotion du bien-être et du respect mutuel). Le travail doit être pensé comme un collectif, ouvert aux partenaires de l'école et orienté vers la résolution de problèmes communs. La tâche du management est de promouvoir le *leadership*, c'est-à-dire un nouveau partage des rôles et des responsabilités dans l'établissement ainsi qu'une vision commune.

L'organisation officielle des programmes scolaires est centrée sur le développement des compétences des élèves en numératie et littératie, en même temps que leur bien-être et leur santé. Ces programmes reposent sur des valeurs (sagesse, compassion, justice, intégrité) et doivent, non seulement contribuer à améliorer les résultats des élèves, mais aussi favoriser une diversité d'expériences dans l'établissement scolaire. L'accent est mis sur l'interdisciplinarité et la communauté de vie entre élèves et ensei-

gnants mais aussi sur les dispositifs d'accompagnement et de soutien personnalisé dans les apprentissages. Dans la réflexion des autorités écossaises, les apprentissages comptent tout autant que l'enseignement lui-même. C'est donc l'élève et les apprentissages qui sont au cœur du dispositif. Cela a des conséquences sur l'évaluation qui joue un rôle essentiel. L'évaluation est au service des apprentissages (*assessment for learning*) [3].

Comment l'Écosse en est-elle arrivée à ces conceptions particulières de l'évaluation, en considérant qu'elles étaient essentielles à la résolution des difficultés des élèves ? C'est la conséquence d'une réflexion conduite par les responsables de l'éducation à la suite de réformes menées au cours des années 1990. Dans le contexte de l'obligation de résultats, le ministère de l'Éducation avait introduit des évaluations nationales, établi des recommandations pour leurs usages, développé un système d'information et une batterie d'indicateurs de la performance et tenté d'introduire une culture de l'évaluation. Mais il a rencontré un certain nombre de difficultés. Les enseignants et les parents d'élèves résistaient à la passation des tests. Les données recueillies dans l'établissement étaient peu fiables et de mauvaise qualité. L'encadrement avait une compréhension limitée des enjeux de l'évaluation. Sur le terrain, les équipes disposaient d'un faible soutien et elles n'étaient guère accompagnées. Dans l'esprit des enseignants et des chefs d'établissement, l'évaluation était strictement séparée des programmes scolaires. En 1999, un rapport de l'inspection écossaise concluait que le système d'évaluation était fragmenté et peu efficace. Après une consultation

publique conduite en 2000, des changements furent réclamés pour améliorer l'évaluation. L'inspection écossaise engagea une réflexion nouvelle afin de changer ses pratiques et promouvoir un cadre d'autoévaluation de l'établissement scolaire.

Depuis, les fonctions de l'inspection ont bien changé. Au lieu d'un contrôle et d'une surveillance étroite, les inspecteurs accompagnent la démarche d'autoévaluation de l'établissement scolaire. Les procédures d'autoévaluation sont jugés prioritaires et sont placées au centre des procédures d'audit [4]. Seule une petite fraction d'établissements fait l'objet d'une inspection annuelle (en cas de mauvais résultats ou de problèmes identifiés). Cinq grands principes ont été retenus pour les visites d'établissement :

- **Écoute** : le recueil des avis des élèves fait partie intégrante du processus d'inspection, de même que la discussion ouverte avec les enseignants.

- **Consultation** : les autorités locales et les parents sont invités à donner leur avis avant la visite de l'établissement.

- **Participation** : à la fin de l'inspection, les responsables de l'établissement et l'équipe pédagogique sont invités à compléter un questionnaire et à exprimer leur avis.

- **Discussion** : le rapport d'inspection est discuté avec le chef d'établissement et le responsable de l'autorité locale. Ces derniers peuvent utiliser une procédure d'appel auprès du ministère en cas de désaccord.

- **Conseil** : les recommandations du rapport débouchent sur un calendrier qui doit être accepté par l'établissement et l'autorité locale, ainsi qu'un suivi des changements opérés dans les 12-18 mois.

À partir du cadre d'autoévaluation de l'inspection écossaise, les membres de l'établissement sont invités à se poser un ensemble de questions sur leur situation, leurs forces et leurs faiblesses, et les moyens d'y remédier. Enseignants, responsables de l'établissement, parents d'élèves, ont à leur disposition un ensemble d'indicateurs qu'ils peuvent utiliser pour se positionner. Chaque indicateur est accompagné d'exemples détaillés de bonnes et de mauvaises pratiques, sur la base d'inspections déjà réalisées dans d'autres établissements. En tout, il existe 30 indicateurs relatifs à des domaines d'activités variés comme les résultats des élèves, la contribution de l'établissement à la communauté locale, le travail en équipes, la mise en œuvre de l'évaluation des apprentissages, le dialogue avec les familles, la formation dans l'établissement, le management et le *leadership*. Pour renseigner ces indicateurs, les membres de l'établissement sont invités à mobiliser plusieurs types d'informations : des données statistiques, mais aussi celles produites à la suite d'entretiens ou d'observations dans l'établissement. La démarche d'autoévaluation s'opère selon une planification et une progression sur l'ensemble de l'année scolaire. Ensuite, l'inspection écossaise confronte son propre audit aux résultats de l'autoévaluation menée par les membres de l'établissement. L'établissement est situé sur une échelle de qualité allant de 1 à 6 pour les différents domaines évalués par les indicateurs. Les informations doivent être détaillées et précisées pour chaque domaine. Cette démarche d'autoévaluation, planifiée sur l'année, est inséparable du projet d'établissement et doit entretenir la dynamique de changement de l'organisation pédagogique.

L'ANGLETERRE

Les établissements parlent par eux-mêmes

Comme L'Écosse, l'Angleterre a rencontré des problèmes à la fin des années 1990 dans la mise en œuvre de l'évaluation externe de son système éducatif. Des critiques ont été formulées à l'encontre de la politique de « stigmatisation » des établissements scolaires catégorisés comme « défaillants » dans les quartiers défavorisés. Les études sur les visites d'établissement menées par l'inspection ont montré des effets pervers sur des données faussées, des recommandations mal exécutées et des inspections pauvrement argumentées.

Une baisse de la performance a pu être observée dans certains établissements qui passaient tout leur temps à préparer les inspections dans un climat de stress. Certains d'entre eux cherchaient à se débarrasser de leurs mauvais élèves ou à renforcer leurs procédures de sélection, ce qui augmentait les abandons de scolarité et les sorties sans qualification. En 1999, le parlement a reconnu l'importance du stress des enseignants et a demandé que l'inspection rétablisse la confiance dans les établissements et cherche à promouvoir des démarches d'autoévaluation. Dans le même temps, le *National Union of Teachers* (principal syndicat des enseignants) a fait réaliser une étude pour promouvoir l'autoévaluation des établissements scolaires (« les établissements parlent par eux-mêmes ») et il a poussé au boycott des évaluations nationales [5]. Le Haut conseil de l'éducation a demandé plus de responsabilités dans le rendre compte en recommandant de renforcer le poids

de l'établissement dans le pilotage et la dynamique de changement, de mettre l'accent sur une obligation de résultats et d'alléger la pression externe.

En 2004, l'idée d'une nouvelle relation entre le gouvernement anglais et les établissements scolaires était formulée par le secrétaire d'État à l'éducation David Miliband [6]. Il affirmait que le temps était venu pour le système éducatif de mettre en place une « obligation de résultats intelligente » s'appuyant sur une démarche d'autoévaluation de l'établissement scolaire, une inspection resserrée et centrée sur la dynamique de changement et l'amélioration des pratiques pédagogiques, la mise en œuvre d'un système d'information améliorant la communication entre l'ensemble des partenaires du système éducatif. L'objectif était d'accorder plus de libertés aux établissements dans la définition de leurs priorités stratégiques, d'alléger les contraintes bureaucratiques, de renforcer les liens avec les parents d'élèves, d'utiliser des technologies de pointe dans la collecte des données et la communication entre l'échelon local et l'échelon national, et d'alléger les inspections pour en réduire les coûts relativement élevés.

Le gouvernement demandait donc aux inspecteurs d'adopter un cadre plus approprié à la compréhension des forces et des faiblesses de l'établissement. Les inspections, plus courtes, devaient s'appuyer sur des interactions avec l'équipe de direction et prendre comme point de départ la démarche d'autoévaluation conduite par l'établissement. Alors que ce dernier était alerté plus tardivement de la visite de l'inspection, afin de réduire le stress des équipes pédago-

giques, des critères communs étaient adoptés. Les inspecteurs devaient davantage s'appuyer sur des preuves objectives et se mettre au service de la dynamique de changement de l'établissement en respectant des normes et un code de conduite pour que les personnes soient traitées avec justice, courtoisie, dans un dialogue constructif et avec un minimum de stress.

La nouvelle relation avec les établissements scolaires s'appuie désormais sur un système d'information élargi comprenant différents éléments [7]. Un profil de chaque établissement fait l'objet d'une diffusion assez large. Il contient des informations sur la réussite des élèves et leur progrès, au regard des *benchmarks* relatifs aux établissements scolaires de caractéristiques et de performance similaires. Chaque unité doit préciser comment elle s'occupe des élèves, présenter les analyses et les recommandations de l'inspection, démontrer la manière dont elle fournit une offre adaptée au contexte local, et afficher son plan stratégique. Ce document entend contribuer au renforcement du dialogue entre l'établissement et les parents d'élèves. Une nouvelle stratégie de communication est établie entre les autorités académiques et l'établissement scolaire : les informations sont fournies au coup par coup, en fonction des besoins, afin d'éviter d'alourdir le travail du chef d'établissement par trop de données inutiles. Ce dernier est informé régulièrement de leur mise à jour et il peut télécharger de nombreuses ressources selon l'ordre de ses priorités. En janvier, les établissements soumettent leur PLASC (*Pupil Annual School Census*) : des données comprenant, pour chaque élève, son état-civil, sa date d'admission dans

l'établissement, son genre, son appartenance ethnique, sa langue maternelle, son code postal de résidence et son identifiant. En contrepartie, les équipes pédagogiques reçoivent leur « paquet d'automne » dénommé PANDA (*Performance AND Assessment data*). Ce système d'information très riche fait apparaître les performances des établissements relatives à des *benchmarks* : il permet à l'établissement d'utiliser l'information dès la rentrée scolaire. Le système d'information PANDA fournit un ensemble de données pour aider les établissements à améliorer leur performance. Les équipes pédagogiques peuvent utiliser les données de valeur ajoutée pour comparer les progrès individuels des élèves et des groupes d'élèves entre les différents cycles de scolarité. Il existe donc une information détaillée des résultats de chaque élève aux tests et aux examens, à l'échelle d'un groupe-classe, et les enseignants peuvent connaître chaque situation individuelle et son évolution de manière très précise. Ces tests nationaux sont publiés régulièrement par le ministère anglais de l'éducation. Ils sont disponibles à l'échelle de l'établissement mais servent aussi au pilotage national.

L'inspection a retenu huit grands principes pour la démarche d'autoévaluation [8, 9]. L'obligation de résultats « intelligente » doit s'appuyer sur le regard propre de l'établissement et sur la façon dont il définit des priorités pour s'améliorer. L'autoévaluation est ancrée dans les pratiques quotidiennes et doit questionner l'apprentissage des élèves, leur réussite et leur développement. En s'appuyant sur des preuves objectives, elle permet de situer l'établissement par rapport à des *benchmarks*. Elle implique

l'équipe pédagogique mais aussi les élèves, les parents et les membres du conseil d'administration.

Préalablement à chaque visite d'inspection, l'établissement doit remplir un formulaire d'autoévaluation montrant ses forces et ses faiblesses selon quatre critères : extraordinaire, bon, satisfaisant, inadapté. Pour cela, il doit sélectionner les données disponibles les plus pertinentes pour chaque domaine évalué (réussite des élèves, orientation, qualité de l'offre, partage des responsabilités et management). Toute la communauté éducative doit porter un jugement en utilisant le vocabulaire approprié et en apportant la preuve que l'établissement est performant. Elle doit préciser les moyens envisagés pour améliorer l'existant. Le cadre d'évaluation comprend aussi des données chiffrées sur l'établissement et des informations montrant qu'il respecte les instructions officielles, notamment en termes d'égalité des chances et de lutte contre les discriminations.

La mise en œuvre de cette démarche d'autoévaluation est accompagnée de la création d'une nouvelle fonction : le partenaire de la dynamique de changement (*School Improvement Partner*). Ce rôle est souvent assumé par un chef d'établissement d'une autre circonscription qui vient aider le chef d'établissement à se poser les bonnes questions et à définir les meilleures orientations stratégiques. Il est considéré comme un « ami critique » : ami, puisque contrairement à l'inspection, il n'est pas là pour établir des recommandations, mais aussi critique parce qu'il dispose d'un recul suffisant, lui permettant de contribuer efficacement à la dynamique de changement. Ses modalités d'intervention sont

strictement définies : respect de la confidentialité et de l'autonomie de l'établissement, soutien professionnel dans l'amélioration des pratiques pédagogiques, évaluation fondée sur des preuves. Comme pour les visites d'inspection, son intervention est inversement proportionnelle au succès de l'établissement en termes de performance et de réussite des élèves. « L'ami critique » doit aider l'établissement à se situer dans un cycle continu d'évaluation, d'action et de planification. Il contribue à la formation continue des enseignants. Il rencontre régulièrement le chef d'établissement pour l'aider à se poser les bonnes questions : comment l'établissement est-il performant ? Quels sont les facteurs-clés ? Quelles sont les priorités et les cibles pour s'améliorer ? Comment l'établissement les met en œuvre ?

LA NOUVELLE-ZÉLANDE

Des enseignants enquêteurs sur leurs propres pratiques

Le programme scolaire Néo-Zélandais se définit par un engagement au service de la communauté, incluant la protection des droits de la communauté Maori, le respect de la diversité culturelle et l'intégration sociale, un niveau élevé d'attentes et d'exigences vis-à-vis du système éducatif, une cohérence orientée vers l'avenir, et l'apprendre pour apprendre. Des compétences-clés définissent les capacités à acquérir par les élèves pour vivre et apprendre tout au long de la vie : penser ; utiliser la langue, les symboles et les textes ; se prendre en charge ; développer la relation aux autres ; participer et

contribuer à la communauté. Le programme scolaire est décomposé en domaines d'apprentissage : anglais, lettres et arts, langues étrangères, mathématiques et statistiques, sciences, sciences sociales, technologie. En Nouvelle-Zélande, il n'existe pas de système d'évaluations nationales comme en Angleterre. Celles-ci ne concernent qu'un petit échantillon d'élèves mais elles sont conçues et administrées avec l'aide des enseignants. Le corps d'inspection, réduit à une petite agence, ne pratique pas de visite systématique des établissements scolaires. Ces derniers sont très autonomes mais ils peuvent participer à des programmes nationaux élaborés par le ministère de l'éducation néo-zélandais. Toutefois, comme en Angleterre, la Nouvelle-Zélande a mis en œuvre une stratégie pour la littératie et la numératie afin d'améliorer les résultats des élèves, particulièrement ceux des Maori qui connaissent un échec scolaire important. L'obligation de résultats et le système d'évaluation néo-zélandais sont principalement centrés sur l'apprentissage des compétences en littératie et numératie.

C'est dans la relation entre l'enseignement et les apprentissages des élèves que le système éducatif néo-zélandais se révèle particulièrement innovant. Selon les instructions officielles, l'enseignant est considéré comme un enquêteur sur sa propre pratique. Avec ses collègues, il doit utiliser les résultats de l'évaluation des élèves pour redéfinir non seulement les conditions d'apprentissage, mais aussi les stratégies pédagogiques permettant de remédier à leurs difficultés. Les apprentissages et les pratiques d'enseignement sont donc interrogés en permanence selon

un cycle continu afin d'améliorer le dispositif au service des progrès des élèves. Cette évaluation est considérée par les responsables de l'éducation comme la base du développement professionnel des enseignants [10]. Ceux-ci doivent avoir à l'esprit la question suivante : qu'avons-nous besoin d'apprendre, comme professionnels, pour accompagner l'apprentissage de nos élèves ? Cette démarche est au centre de la dynamique de changement de l'organisation pédagogique. Elle est considérée comme le moyen le plus efficace d'améliorer l'apprentissage professionnel des enseignants. Les évaluations et le développement professionnel sont donc étroitement liés au projet d'établissement selon un cycle périodique d'enquêtes à plusieurs. Tous les membres de la communauté locale de l'établissement sont associés à la démarche d'auto-évaluation : direction, enseignants, élèves, parents d'élèves, partenaires de l'établissement.

Dans la classe, l'évaluation est également au service des apprentissages. Les enseignants doivent mettre en œuvre une évaluation compréhensive des compétences des élèves et de leurs marges de progression. Cette forme d'évaluation diffère de la note et du test de compétences. L'enseignant doit accompagner le développement cognitif des élèves en les aidant à comprendre et à situer leurs progrès dans les différents domaines de compétences. L'évaluation pour les apprentissages signifie que les élèves sont impliqués dans le processus et qu'ils bénéficieraient régulièrement d'un retour d'information (*feedback*) de la part de l'enseignant. Les élèves sont donc invités à développer l'autoévaluation

et l'évaluation entre pairs, tandis que les enseignants partagent avec eux leurs objectifs et leurs critères. Pour mettre en œuvre cette évaluation pour les apprentissages, les enseignants utilisent un logiciel spécifique afin de produire des jugements fiables et pertinents sur les connaissances et les compétences des élèves. Ils peuvent suivre leurs progrès pour chaque domaine du programme scolaire, en précisant les objectifs à atteindre, et en repérant facilement les forces et les faiblesses de l'élève, de la classe, ou de l'établissement. Pour chaque compétence évaluée, l'enseignant dispose d'un exemple de bonnes pratiques (des situations d'apprentissage ayant fait leur preuve) auquel il peut accéder grâce à une banque de données sur internet. Ces « modèles de bonnes pratiques » servent à identifier les caractéristiques-clés de l'apprentissage et de la réussite à une étape particulière du développement de l'élève. Ils aident les élèves, les enseignants et les parents à connaître les niveaux de réussite et les prochaines étapes de l'apprentissage. Ils guident les enseignants dans leur interprétation des programmes scolaires quand ils ont à produire des jugements sur le travail réalisé par chaque élève.

LA FRANCE

L'autoévaluation comme support de l'expérimentation pédagogique ?

L'évaluation de l'EPL en France se résume souvent à l'élaboration d'un diagnostic des forces et faiblesses du projet d'établissement, lequel est ensuite présenté et discuté avec

les équipes pédagogiques. Bien qu'il permette de positionner l'établissement au regard d'un certain nombre d'indicateurs concernant la réussite des élèves aux examens, leur orientation, les problèmes de discipline et de violence, il demeure assez général quant au fonctionnement pédagogique de l'organisation scolaire et ne permet pas vraiment de mobiliser les enseignants dans une dynamique de changement. Malgré tout, le droit à l'expérimentation pédagogique offre des possibilités de travail collectif autour d'un projet ou d'une innovation, tout en suscitant un questionnement sur l'environnement interne de l'établissement, ses possibilités de diversification dans la prise en charge des élèves, les reformulations possibles des pratiques d'enseignement et des conditions d'apprentissage. Un tel dispositif d'expérimentation peut constituer une ressource pour le conseil pédagogique en l'amenant progressivement à un questionnement plus général sur les possibilités d'élargissement ou de transfert de certaines pratiques de l'expérimentation à l'ensemble du projet d'établissement. Dès lors, une démarche d'autoévaluation de l'expérimentation, bien maîtrisée et bien ciblée, peut créer une dynamique de changement et susciter une réflexivité nouvelle de la part des enseignants les plus innovants, laquelle bénéficiera ensuite à l'ensemble de la communauté éducative. En fait, les enseignants sont déjà des évaluateurs de leurs propres pratiques. Mais l'objectif est, pour les enseignants engagés dans l'expérimentation, de mieux ancrer cette autoévaluation informelle dans une démarche fondée sur une réflexion commune et sur des faits tangibles et partagés.

Il s'agit, pour le chef d'établissement et son équipe, d'établir un cadre de référence visant à promouvoir des discussions utiles entre les enseignants en les aidant à s'approprier une culture de l'enquête. Une première image de l'expérimentation est obtenue, en associant, si besoin, les élèves et leurs parents. Des domaines prioritaires sont identifiés qui nécessitent ensuite une enquête plus approfondie. La tâche du groupe engagé dans l'autoévaluation de l'expérimentation est de rechercher un consensus tout en respectant les désaccords. Il s'agit d'étayer les discussions par des faits ou des éléments de preuve, en faisant attention à ce que les jugements s'appuient sur une information valide et vérifiable, et non sur des préjugés. Le domaine évalué peut être choisi pour des raisons différentes. Dans un cas, parce que les opinions sur l'expérimentation sont divergentes, dans l'autre parce qu'il n'y a pas de données ou qu'il s'agit d'une faiblesse réelle de l'établissement ou encore, d'un axe de progrès à développer. L'autoévaluation de l'expérimentation n'est pas une fin en soi mais le début d'un processus continu et cyclique, orienté vers l'amélioration du dispositif et des pratiques pédagogiques, mais aussi vers le progrès des élèves et leur réussite.

Une fois les domaines d'enquête choisis, l'étape suivante consiste à utiliser les outils appropriés pour collecter des données. Néanmoins, l'expérience montre que les outils ne doivent pas être choisis de manière aléatoire, mais en tenant compte de ce qu'ils permettent d'évaluer, et des limitations propres à leur usage. Il est utile de se poser les bonnes questions avant de privilégier telle ou telle méthode d'analyse. Un outil pour quoi faire? Qui utilisera les résultats ?

Quels peuvent être les effets non désirés ? Est-il possible réellement d'utiliser cet outil ? Est-ce que cela peut être mis facilement en pratique ? Est-ce qu'il y a le temps pour cela ? L'outil permet-il d'évaluer en surface ou en profondeur ? Quelles sont les données attendues ? Qui sera impliqué dans la démarche ? Combien de temps cela prendra pour avoir un retour ? Ces outils doivent être économiques (simples et faciles d'usage), informatifs (apporter une information nouvelle), formatifs (aider au choix d'options ou de stratégies), adaptables (capables d'être utilisés de manière flexible en fonction des besoins), conviviaux (susciter un engouement ou un enthousiasme). La démarche d'autoévaluation de l'expérimentation a plus de chances d'aboutir si les outils sont utilisés pour approfondir et comprendre certains aspects, en rendant explicite ce qui était implicite ou simplement pris comme une évidence non questionnée.

Les outils à disposition sont nombreux pour collecter des données et remplir le cadre d'autoévaluation. Ils peuvent prendre la forme d'un questionnaire, d'entretiens ou d'observations conduites au sein de l'établissement. Les données peuvent provenir du diagnostic porté par le chef d'établissement ou d'une enquête diligentée par lui sur un domaine spécifique. Des discussions par groupe sur une thématique précise peuvent être organisées. Un profil de l'établissement peut être élaboré de même qu'un arbre des priorités. L'objectif de l'autoévaluation est d'aider les enseignants à se poser les bonnes questions et à anticiper les effets attendus en termes de changement de leurs pratiques pédagogiques et d'amélioration de la réussite des élèves. Elle doit nourrir la réflexion

des enseignants engagés dans l'expérimentation en favorisant un retour régulier des informations, leur diffusion et leur partage au sein du groupe. Elle doit servir de mémoire en capitalisant progressivement les réussites mais aussi les échecs puis en débouchant sur une synthèse finale qui sera utile au conseil pédagogique. Il est toutefois impossible de couvrir tous les domaines évalués par des indicateurs. Les résultats quantifiables ne doivent pas être surreprésentés au détriment d'autres informations tout aussi utiles. Si des indicateurs sont retenus, ils doivent être largement compréhensibles et partagés par tous.

Il n'existe pas de démarche d'autoévaluation de l'expérimentation qui soit supérieure à une autre parce que plus objective, plus efficace, plus formalisée. C'est d'abord une enquête à plusieurs acteurs sur un environnement pédagogique qui change en permanence. Si le groupe part de la définition des statuts, des missions, ou de l'identité professionnelle des uns et des autres pour construire la démarche, il a toutes les chances de ne pas aboutir. Il doit d'abord accepter le dialogue, c'est-à-dire une rencontre entre des points de vue nécessairement divergents. Le chef d'établissement doit aussi favoriser l'émergence de pratiques innovantes de la part des enseignants engagés dans la démarche. C'est un processus de co-construction de connaissances et de compétences qui amène chacun à partager sa propre expertise avec celle des autres. La démarche d'autoévaluation de l'expérimentation est donc une activité sociale : elle peut être source de conflits, d'incompréhensions, de malentendus. Il est nécessaire d'accompagner les équipes, d'opérer des médiations

pour donner du sens à la démarche, et maintenir la dynamique.

Celle-ci peut être accompagnée par un « ami critique » : inspecteur, chef d'établissement, conseiller en développement, formateur, chercheur, etc. Celui-ci peut être défini comme une personne de confiance posant des questions « dérangement », proposant des données à étudier avec un autre regard et critiquant le travail déjà réalisé. « L'ami critique » n'est pas neutre, mais il agit pour soutenir l'expérimentation et pour la défendre, en travaillant avec l'encadrement et les enseignants pour les aider à améliorer leur projet avec la distance nécessaire. C'est une personne extérieure qui n'entretient aucune familiarité avec les membres de l'établissement. Il apporte son expertise à la mise en place de la démarche d'autoévaluation, en clarifiant les objectifs et en créant un climat favorable. Il aide les personnes à s'approprier un cadre de référence. Il donne son avis sur le choix des outils et leur usage. Il contribue à la collecte et à l'interprétation des données. S'il veut être efficace, « l'ami critique » doit posséder des qualités d'ouverture (intervenir sans *a priori*), d'écoute (entendre les besoins qui s'expriment dans l'expérimentation), de compréhension (bien comprendre le contexte de l'autoévaluation), de pertinence (donner des avis utiles), d'empathie et de communication (partager des idées et se faire accepter), de persuasion (faire prendre les décisions les plus favorables), et d'innovation (trouver des solutions inattendues). « L'ami critique » peut aussi épouser

différents rôles : conseiller scientifique (il propose des méthodes et un soutien efficaces), organisateur (il prépare les réunions et les synthèses), entraîneur (il imprime une direction en rassurant et encourageant les personnes), facilitateur (il gère les relations interpersonnelles), ou arbitre (il favorise la cohérence et le rapprochement des points de vue). « L'ami critique » doit avoir une attitude compréhensive à l'égard de l'expérimentation sans être trop directif ou chercher à imposer des solutions clés en main.

CONCLUSION

Une démarche d'autoévaluation de l'expérimentation, même accompagnée, rencontrera forcément des obstacles qu'il s'agit d'anticiper. Par peur du changement ou du dévoilement, des membres de la communauté éducative peuvent défendre un *statu quo* parce qu'ils ont intérêt à ce que rien ne change. D'autres s'enfermeront dans une approche formaliste nuisible à la créativité et à la dynamique de changement. Il existe aussi de bons arguments pour considérer que les problèmes et les difficultés sont d'abord de la responsabilité des autres, ce qui évite d'interroger et d'explicitier sa propre pratique. Certains individus cherchent à complexifier les dispositifs et les processus, en renforçant les contraintes ou en rajoutant des règles afin de rendre plus difficile la démarche entreprise.

Mais, si elle est activée, l'autoévaluation de l'expérimentation peut contribuer à changer la culture

professionnelle des enseignants et à susciter un travail d'équipe au bénéfice du progrès des élèves, à condition qu'elle se centre sur les conditions d'enseignement et d'apprentissage, et non sur des dispositifs ou des projets périphériques. Comme l'a montré H.S. Timperley [10], l'enquête des enseignants sur leur propre pratique est le premier vecteur de l'amélioration des résultats des élèves. La démarche d'autoévaluation de l'expérimentation peut constituer une propédeutique à une approche plus globale dans l'établissement. Elle peut alimenter les réflexions du conseil pédagogique et aider à la constitution, mais aussi à la révision des axes de progrès. Elle constitue un outil de dialogue entre l'équipe pédagogique et le chef d'établissement, entre l'établissement, l'inspection et les autorités académiques. Elle est gage de confiance dans la diffusion et l'appropriation d'une culture de l'évaluation. En ne se réduisant pas à une approche instrumentale et techniciste, la démarche d'autoévaluation reconnaît la diversité des principes de référence qui caractérisent l'organisation pédagogique. Elle contribue à construire l'identité de l'établissement scolaire tout en tenant compte de son environnement local. C'est aussi un exercice de démocratie délibérative ouvert à l'expression des droits des élèves et des parents d'élèves pour les questions scolaires. ■

BIBLIOGRAPHIE

- [1] **Derouet J.-L., Normand R., Bessy T.**, (2009), *Quelles politiques pour l'égalité ? Savoirs, gouvernances et obligation de résultats, Réflexions croisées franco-américaines*, Lyon, INRP.
- [2] **Normand R., Derouet J.-L. (coord.)**, (2011), *Évaluation, développement professionnel et organisation scolaire. Revue française de pédagogie*, n° 174. Dossier thématique.
- [3] **Black P., William D.**, (1998), *Inside the Black Box: Raising standards through classroom assessment*, King's college, Londres.
- [4] **HMIE**, (2002), *How good is our Schools?*, Her Majesty Inspectorate in Education, Edimbourg.
- [5] **MacBeath J.**, (1999), *Schools must speak for themselves. The case for school self-evaluation*, Routledge, Londres.
- [6] **MacBeath J.** (2006), *School Inspection and Self-evaluation: working with the New Relationship*, RoutledgeFalmer, Londres.
- [7] **DFES**, (2004), *A new relationships with Schools*, Department for Education and Skills, Londres
- [8] **OFSTED**, (2004), *The Common Inspection Framework for Inspecting Education and Training*, Office for Standards in Education, Londres.
- [9] **OFSTED**, (2005), *Self-Evaluation Form for secondary schools*, Office for Standards in Education, Londres.
- [10] **Timperley, H.S.**, (2008), *Teacher Professional Learning and Development*. International Academy of Education/ International Bureau of Education. Pays-Bas.

Expérimentations et résultats d'évaluations

Évaluation de l'effet du dispositif d'enseignement intégré de science et technologie (EIST)

Premiers résultats de l'analyse des progressions des élèves sur trois temps de mesure

Le dispositif expérimental d'enseignement intégré des sciences et technologies (EIST) a fait l'objet d'une évaluation conduite par la DEPP, qui a suivi une cohorte d'élèves scolarisés en 6^e à la rentrée 2008. L'analyse des progressions de ces élèves (trois temps de mesure jusqu'en fin de 5^e), ne laisse apparaître aucun effet significatif sur la performance des élèves ayant bénéficié de l'EIST par rapport aux élèves de l'échantillon témoin à ce stade de l'expérimentation. Cette absence d'effet concerne aussi bien les performances cognitives que les attitudes à l'égard des sciences. Les effets liés au biais de sélection, à l'attrition, ou encore au choix des instruments de mesure sont discutés. En outre, cet article aborde précisément la méthodologie suivie pour mesurer les progressions des élèves dans le temps, à travers la construction d'une échelle de performance commune aux trois temps de mesure. Au-delà de l'évaluation de l'EIST, un résultat frappant est apparu : la baisse générale du niveau d'intérêt et de motivation à l'égard des sciences, au cours des premières années de collège.

Marion Le Cam, Thierry Rocher

Bureau de l'évaluation des élèves
DEPP, MENJVA

Remerciements à

Pascal Bessonneau (DEPP) et **Ahmed Ould Ahmed Jiddou** (ENSAE)

pour la mise en œuvre des calculs psychométriques

Jérôme Goidin et **Ginette Bourny** (DEPP)

pour la conception et le suivi du dispositif.

QU'EST-CE QUE L'EIST ?

Depuis l'année scolaire 2006-2007, l'enseignement des sciences et de la technologie au collège fait l'objet d'une expérimentation portée par l'Académie des sciences qui préconise un rapprochement des enseignements de sciences de la vie et de la Terre (SVT), de sciences physiques et chimiques (SPC) et de technologie, lors des deux premières années du collège. Il s'agit de retarder le morcellement de ces enseignements et de promouvoir l'utilisation de la démarche d'investigation basée sur l'expérience et l'observation afin de favoriser le développement de la curiosité des élèves et de leur donner le goût pour les disciplines scientifiques et technologiques.

Les 38 collèges volontaires au départ pour expérimenter cette démarche, se sont engagés à assurer pendant quatre ans un enseignement intégré de science et technologie (EIST) pour un (ou plusieurs) groupe(s) d'élèves de niveau 6^e et/ou 5^e. Pour l'année 2008-2009, date de la mise

en place de l'évaluation du dispositif, 36 collèges ont participé à l'expérimentation pour le niveau 6^e, et 10 ont participé à l'expérimentation pour le niveau 5^e.

Concrètement, la situation la plus fréquente est la suivante : les élèves de deux classes sont répartis en trois groupes. Chaque groupe reçoit pendant 3 heures 30 par semaine un enseignement intégrant les trois disciplines, mais assuré par un seul professeur de l'équipe EIST (constituée le plus souvent d'un professeur de SVT, un de SPC et un de technologie). Les trois enseignants préparent de façon concertée leurs progressions et leurs cours à partir de thèmes transversaux qui permettent d'intégrer les différents programmes disciplinaires. Ces thèmes sont encadrés par un « guide pédagogique », élaboré conjointement par l'Académie des sciences et les Inspections générales concernées.

Avec le soutien de la DGESCO (direction générale de l'enseignement scolaire), certaines conditions ont été réunies pour faciliter la mise en œuvre de cette expérimentation :

1. la réintégration d'un enseignement de SPC en 6^e ;

2. l'augmentation du volume horaire de science et technologie en 6^e (3 heures 30 hebdomadaire au lieu de 3 heures) ;

3. la mise en œuvre de classes à effectifs réduits pour cet enseignement (trois groupes constitués à partir de deux classes) ;

4. et enfin, la prise en compte financière d'une heure de concertation par semaine pour faciliter le travail en commun des enseignants impliqués, nécessaire au bon déroulement du projet.

En première analyse, il apparaît difficile d'être en mesure de distinguer les effets propres à chacun de ces aspects sur les progrès des élèves. Par conséquent, l'évaluation ne concerne pas seulement l'aspect « pédagogique », c'est-à-dire l'enseignement intégré de sciences et technologie, mais également ces quatre aspects qui constituent des « facteurs confondus ».

LE DISPOSITIF D'ÉVALUATION

L'évaluation des effets de l'EIST a débuté lors de l'année scolaire 2008-2009. La DEPP a été contactée pendant l'été 2008 pour une mise en place de l'évaluation dès l'automne 2008. Ce calendrier très serré explique certaines caractéristiques du dispositif d'évaluation, que nous détaillons ci-après.

Constitution des groupes

Trois groupes d'élèves, de 6^e et de 5^e, ont été constitués :

- Groupe 1 (EIST) : les élèves qui reçoivent un enseignement EIST (collèges de France métropolitaine

uniquement, deux établissements situés outre-mer expérimentent l'EIST mais ne sont pas concernés par l'évaluation) ;

- Groupe 2 : un échantillon d'élèves qui reçoivent un enseignement « traditionnel » mais scolarisés dans des collèges qui pratiquent l'EIST dans d'autres classes ;

- Groupe 3 (témoin) : un échantillon d'élèves qui reçoivent un enseignement « traditionnel » dans des collèges non impliqués dans l'EIST.

Il est très important de signaler que les collèges engagés dans l'EIST sont des établissements volontaires. Certaines caractéristiques de ces collèges, et en particulier celles de leurs enseignants de sciences, sont susceptibles d'expliquer la participation à l'expérimentation. Certaines de ces caractéristiques, difficilement observables, comme le degré de motivation des enseignants, la cohésion des équipes pédagogiques, etc. sont potentiellement corrélées avec les progrès réalisés par les élèves en science. Un effet de sélection est donc à craindre. En d'autres termes, si un effet positif sur l'acquisition de compétences scientifiques est observé, cet effet sera-t-il attribuable au dispositif expérimental lui-même ou bien plus simplement à ces caractéristiques non observées des enseignants des collèges volontaires ?

Afin de pouvoir répondre à cette question, au moins de façon partielle, un deuxième groupe a été considéré, regroupant des élèves scolarisés dans les collèges engagés dans l'EIST mais dont la classe n'est pas concernée par l'EIST. La constitution de ce groupe a procédé par échantillonnage d'une classe entière, non concernée par l'EIST, au sein des collèges ayant des classes impliquées dans le dis-

positif. Si un effet positif est observé sur le groupe expérimental et également sur ce groupe particulier, il est fort probable qu'il soit dû à des caractéristiques des enseignants, qui sont en partie communes à ces deux groupes. Malheureusement, comme nous le verrons plus loin, les données recueillies pour ce groupe d'élèves se révèlent de faible qualité et rendent difficile une analyse approfondie en ce sens.

Le troisième groupe (groupe témoin) sert de référence. Il s'agit d'un échantillon représentatif des élèves des académies concernées, la population-cible étant celle des élèves scolarisés en 6^e et en 5^e dans les collèges publics et privés sous contrat en France métropolitaine (cf. annexe 1).

Suivi longitudinal

Les analyses présentées ici portent sur les élèves de 6^e qui ont été suivis et interrogés à trois moments de leur scolarité, une première fois en début d'année scolaire (novembre 2008), une deuxième fois en fin d'année (mai 2009), et une troisième fois en fin de 5^e (mai 2010)¹. Ces élèves sont suivis et évalués chaque année, jusqu'à leur arrivée en fin de collège : en 4^e (mai 2011) et en 3^e (mai 2012). L'objectif de ces évaluations est de déterminer s'il existe une différence de niveau des acquis en termes de

NOTE

1. Les élèves de 5^e à la rentrée 2008 ont également fait l'objet d'un suivi. Cependant, vu le faible nombre d'établissements proposant l'EIST en 5^e et le manque d'informations précises quant à leur participation concrète à l'EIST, nous centrons ici l'analyse sur les élèves en 6^e à la rentrée 2008. Des études ultérieures seront menées pour préciser, si possible, les effets obtenus sur l'échantillon des élèves de 5^e à la rentrée 2008.

connaissances, de compétences et d'attitudes à l'égard des sciences, entre les élèves qui ont bénéficié du dispositif EIST et les élèves qui ont reçu un enseignement « traditionnel » (échantillon témoin).

Il faut noter que les élèves ayant redoublé ou ayant changé d'établissement dans l'intervalle ne sont pas suivis, ce qui aurait nécessité la mise en place de moyens beaucoup plus importants que ceux alloués à cette évaluation. Une description des phénomènes d'attrition « structurelle » est présentée dans la suite de l'article.

L'ÉVALUATION DES CONNAISSANCES ET DES COMPÉTENCES EN SCIENCES ET TECHNOLOGIE

Conception sous contraintes

Le cadre de l'évaluation des performances cognitives est celui des programmes disciplinaires et du socle commun de connaissances et de compétences. Le groupe de concepteurs, constitué d'enseignants, a été configuré de manière à assurer la représentation de chacune des disciplines, ainsi que la présence d'enseignants appartenant à la fois à des collèges EIST et à des collèges non-EIST.

Le tableau 1 présente la répartition des items² selon les compétences visées, pour les trois sessions d'évaluation analysées ici (début de 6^e, fin de 6^e, fin de 5^e). En outre, la conception

NOTE

2. Le mot « items » est ici à entendre au sens de l'information minimale recueillie – la question – et non au sens des items du socle commun de connaissances et de compétences « Livret personnel de compétences ».

Tableau 1 – Répartition des items selon les compétences visées

Compétences	session 1	session 2	session 3
	début 6 ^e	fin 6 ^e	fin 5 ^e
Rechercher, extraire, organiser l'information	27	20	17
Réaliser, mesurer, appliquer des consignes	-	4	9
Pratiquer une démarche	17	11	14
Présenter la démarche et les résultats	-	2	4
Connaître la notion	10	25	11
Total	54	62	55

Lecture : l'épreuve de sciences de la première session comportait 54 items dont 27 relatifs à la compétence « rechercher, extraire, organiser l'information »

Source : MENJVA-DEPP

a veillé à respecter l'équilibre selon les différentes connaissances du socle commun : univers, matière, vivant, énergie, objets techniques, environnement et développement durable.

Comme nous l'avons évoqué plus haut, le calendrier très serré a contraint certains aspects du dispositif d'évaluation. Concernant la conception des épreuves cognitives, ces contraintes portent, d'une part sur les modalités d'expérimentation des instruments, d'autre part sur le format des épreuves.

Concernant l'expérimentation des instruments de mesure, rappelons qu'il s'agit d'une étape cruciale dans la démarche d'élaboration d'épreuves. Pour des dispositifs d'évaluation tels que CEDRE ou PISA par exemple, deux années sont nécessaires à la construction des outils d'évaluation avant leur utilisation finale : la première année, le travail consiste à construire les instruments et à les « cobayer » auprès de quelques classes ; la deuxième année, il s'agit d'effectuer l'expérimentation d'un grand nombre d'instruments auprès d'un échantillon représentatif, afin de sélectionner les plus pertinents, en se basant sur des critères statistiques spécifiques. Ce processus s'inscrit dans le rythme de l'année scolaire : les différentes versions doivent être passées à la même période. C'est pourquoi deux années

de travail sont nécessaires en amont de l'évaluation finale. Dans le cadre de l'EIST, ce calendrier a dû être sévèrement comprimé, ce qui explique certains choix.

Une conséquence importante concerne le format des items : les épreuves des deux premières sessions – début de 6^e et fin de 6^e – sont constituées de questions à choix multiples (QCM), pour partie issues de l'évaluation CEDRE CM2 2007 en sciences expérimentales. Les QCM ont plusieurs avantages : la capacité à couvrir un domaine, à travers l'emploi d'un grand nombre de questions, garantissant ainsi la représentativité de l'échantillon des questions ; la simplicité et l'objectivité de la correction ; la brièveté de la réponse (les taux de non-réponse sont très bas en comparaison de ceux que l'on peut observer sur certaines questions ouvertes qui nécessitent une implication dans la rédaction de la réponse).

En revanche, ce format d'items ne permet pas d'évaluer certains aspects importants, comme la production d'une argumentation par exemple. Ce déficit n'est pas gênant si l'objectif est d'évaluer le niveau initial des élèves, essentiellement en tant que variable de contrôle. En termes de variables mesurant le produit de l'expérimentation, le format de QCM peut néanmoins apparaître trop

restrictif. En particulier, l'EIST insiste sur la démarche d'investigation, sur l'expérience, qui sont des dimensions plus difficilement appréhendables sous ce format, et même plus généralement sous une forme d'évaluation papier-crayon.

C'est pourquoi les épreuves des années suivantes intègrent des questions ouvertes, qui ont pu faire l'objet d'une expérimentation préalable. En 2010, l'épreuve de 5^e comportait déjà 10 questions ouvertes (QO) parmi les 55 proposées. Dans les évaluations de 4^e et de 3^e, la mise en œuvre d'une démarche scientifique a été accentuée : de nombreuses questions ouvertes ont été introduites et permettront d'approfondir l'analyse des effets de l'EIST, qui, par hypothèse, devraient se concentrer sur ces dimensions.

Une même échelle pour les trois temps de mesure

Nous avons souhaité nous inscrire dans une perspective de développement des acquisitions en sciences et donc pouvoir suivre dans le temps les progressions individuelles des élèves sur une échelle unique de performances dans cette matière. Cette exigence nécessite de pouvoir relier les résultats d'une évaluation à une autre. Il n'était cependant pas possible de proposer exactement les mêmes épreuves aux élèves pour les trois temps de mesure : les programmes scolaires évoluent et il convient d'évaluer les performances des élèves au regard des attendus existant au moment où ils sont évalués.

C'est pourquoi nous avons développé des évaluations différentes selon les sessions mais, afin d'assurer la comparabilité, nous avons

introduit des items communs entre les épreuves de deux sessions successives : entre l'évaluation de début et de fin de 6^e ; entre l'évaluation de fin de 6^e et de fin de 5^e. Cette architecture en forme d'escalier permet de construire une échelle de scores commune à l'ensemble des temps de mesure et donc de suivre les progrès des élèves sur cette échelle.

L'ajustement des métriques (ou *equating*) consiste à positionner les élèves sur la même échelle, quelle que soit l'évaluation qu'ils ont passée. Ce travail implique le recours à des modélisations spécifiques, présentées dans l'encadré « Création d'une échelle commune de performance » (p. 84).

Analyse des items

Préalablement au calcul du score global, nous avons procédé à l'analyse interne des épreuves dans l'objectif de consolider les instruments de mesure. Il s'agit d'identifier et d'éliminer les items qui présentent un dysfonctionnement. Premièrement, il peut s'agir d'une mauvaise propriété liée à la mesure. Deuxièmement, la construction du score global étant issue d'une modélisation statistique, certains items présentent des défauts d'ajustement au modèle employé. Enfin, parmi les items repris d'une session à l'autre, certains laissent apparaître un fonctionnement différent selon le temps de mesure.

Plus spécifiquement, ce « toilettage » des épreuves est opéré selon trois critères :

- Faible pouvoir discriminant : la réussite à l'item est peu corrélée au score obtenu aux autres items. Cette corrélation item-test indique dans quelle mesure l'item s'inscrit dans la dimension générale. Elle indique

également la différence de performance constatée entre les individus qui réussissent l'item et ceux qui l'échouent ; pour un item très discriminant, les élèves qui réussissent cet item ont des scores aux autres items nettement supérieurs à ceux des élèves qui échouent l'item.

- Mauvais ajustement au modèle de réponse à l'item : un modèle de réponse à l'item à deux paramètres est appliqué pour chaque temps de mesure (*cf. encadré*). Les items s'ajustant mal à cette modélisation, c'est-à-dire les items dont le modèle prédit mal le comportement, ont été éliminés.

- Fonctionnement différentiel de l'item (FDI) : dans le but de produire une échelle commune aux trois moments de mesure, il est nécessaire que les items communs – repris d'une session à l'autre – adoptent un comportement similaire. En particulier, l'hypothèse est faite que la hiérarchie de difficulté de ces items est conservée d'une session à l'autre. Les items présentant un fonctionnement différentiel selon le moment de mesure sont éliminés.

Le tableau 2 synthétise le résultat de ce toilettage et dénombre les items identifiés pour chacune des caractéristiques présentées ci-dessus. Il apparaît que les défauts d'ajustement au modèle et les fonctionnements différentiels ne représentent qu'une petite partie des causes d'élimination des items. Le principal facteur d'élimination des items pour dysfonctionnement est celui de la mauvaise discrimination³, conséquence directe du resserrement des phases d'expérimentation.

Ces chiffres témoignent du caractère éminemment empirique du processus de mesure et invitent à prendre

Tableau 2 – « Toilettage » des épreuves

		N total	Mauvaise discrimination	Mauvais ajustement	FDI	N final	N communs
Session 1	Début 6 ^e	54	12	4		36	
					2		6
Session 2	Fin 6 ^e	62	26	3		31	
					0		8
Session 3	Fin 5 ^e	55	4	4		47	

Lecture : parmi les 54 items de l'épreuve de sciences de la première session, 12 ont été éliminés en raison d'un indice de discrimination (ou corrélation item-test) trop faible, 4 ont été éliminés car ils s'ajustaient mal au modèle de mesure (modèle de réponse à l'item) et 2 ont été éliminés car ils présentaient un fonctionnement différentiel (FDI) entre la première et la deuxième session. Au final, l'épreuve de début de 6^e comporte 36 items, dont 6 sont communs avec l'épreuve de fin de 6^e.

Source : MEN/JVA-DEPP

conscience de la temporalité nécessaire à l'aboutissement de ce processus. Toute mesure des compétences est un construit, dont il convient de vérifier empiriquement le fonctionnement au préalable.

Une analyse complémentaire concerne les FDI en fonction du fait de suivre ou non le dispositif EIST. Autrement dit, à score égal sur l'échelle commune, les élèves

qui suivent le dispositif EIST ont-ils des résultats différents des autres sur certains items ? Aucun fonctionnement différentiel de ce type n'a été observé pour les trois premiers temps de mesure. Mais si, réellement, l'EIST produit un effet sur la démarche d'investigation par exemple, il sera intéressant de reconduire ce type d'analyses sur les sessions ultérieures.

aux phénomènes environnementaux ; attrait pour les expériences ; engagement pour les sciences ; activités en relation avec les sciences en dehors de l'école (cf. annexe 2).

La construction des scores a suivi deux étapes : 1) validation des dimensions et 2) calcul des indices.

Dans un premier temps, la structure dimensionnelle est explorée et validée à l'aide d'une analyse en facteurs communs permettant d'identifier les facteurs sous-jacents mesurés par ces items. Afin de simplifier la structure des données, la méthode de rotation oblique des axes factoriels est utilisée ; autrement dit, les facteurs ne sont pas considérés comme indépendants. Finalement, un modèle satisfaisant en six facteurs est obtenu.

Dans un second temps, pour chacun des facteurs préalablement identifiés, une analyse en composantes principales (ACP) est réalisée sur les items qui le composent. L'ACP permet de synthétiser de manière optimale l'ensemble de ces items en une seule variable continue : il s'agit d'une moyenne pondérée des questions concernées. Les poids de chaque question sont calculés sur les données de la première session, puis sont repris à l'identique pour le calcul des scores des sessions suivantes.

NOTE

3. Il est frappant en effet de constater que la proportion d'items peu discriminants est très différente selon les sessions : environ 20 % pour la première session (début de 6^e), plus de 40 % pour la deuxième session (fin de 6^e) et seulement 7 % pour la troisième session (fin de 5^e). Ces différences s'expliquent par le fait que, selon les cas, les items ont fait l'objet ou non d'une expérimentation préalable. Ainsi, pour la deuxième session, aucun des items n'a pu être testé car ils ont été construits au cours de l'année scolaire 2008/2009 pour être directement passés en fin d'année scolaire. Pour les items de fin de 5^e, les items ont également été construits lors de l'année scolaire 2008/2009 et ils ont pu être testés sur un autre échantillon d'élèves de 5^e en mai 2009, afin de sélectionner une épreuve de qualité à proposer aux élèves du dispositif expérimental en mai 2010, arrivés en fin de 5^e. S'agissant des items de l'évaluation de début de 6^e de novembre 2008, une partie d'entre eux provenait de l'évaluation CEDRE sciences expérimentales ayant eu lieu en 2006 auprès d'un échantillon d'élèves de fin de CM2. Ceci explique la situation intermédiaire de cette épreuve en termes de nombre d'items éliminés.

L'ÉVALUATION DES ATTITUDES À L'ÉGARD DES SCIENCES

Lors de chaque session d'évaluation, les élèves ont également passé un questionnaire visant à mesurer leurs sentiments, leurs motivations et leurs attitudes vis-à-vis des sciences (variables conatives). Ce questionnaire est repris intégralement et à l'identique pour chacun des temps de mesure : contrairement aux épreuves cognitives, le questionnement de ces dimensions ne nécessite pas d'ajustement spécifique selon le niveau scolaire interrogé.

Au total, ce questionnaire comporte 26 questions de type Likert (« tout à fait d'accord », « d'accord », etc.) qui se répartissent selon six facteurs : sentiment d'efficacité en science ; les sciences dans l'avenir et dans le futur métier ; sensibilisation

Création d'une échelle commune de performance

Modèle de réponse à l'item

La construction d'une échelle de scores commune aux trois épreuves cognitives s'inscrit dans le cadre des modèles de réponse à l'item (MRI). Plus précisément, nous utilisons un MRI à deux paramètres (dit aussi « 2PL, two-parameter logistic model »).

Dans ce modèle, les élèves sont uniquement définis par leur niveau de compétence, qui correspond à leur position sur la dimension latente θ . Plus leur niveau de compétence est élevé, plus la probabilité de réussite est importante.

Les items, quant à eux, sont caractérisés par deux paramètres. D'une part, le niveau de difficulté traduit le degré de succès de l'item : lorsque le niveau de difficulté d'un item est élevé, la probabilité de le réussir est réduite. D'autre part, le paramètre de discrimination, traduit la sensibilité de l'item à une variation du niveau de compétence. Ce paramètre, aussi appelé « pente » de l'item, est positif : lorsque la discrimination est élevée, une variation du niveau de compétence entraîne une variation importante de la probabilité de répondre correctement.

De manière formelle, ce modèle exprime la probabilité qu'un élève i réussisse l'item j , sous la forme suivante :

$$P(Y_i^j = 1 / \theta_i, a_j, b_j) = \frac{\exp(D a_j (\theta_i - b_j))}{1 + \exp(D a_j (\theta_i - b_j))}$$

où Y_i^j est la réponse de l'élève i à l'item j (de valeur 1 si l'item est réussi, 0 sinon), θ_i est le niveau de compétence de l'individu i , D est une constante qui vaut 1,7, b_j est le paramètre de difficulté de l'item j et a_j est le paramètre de discrimination de l'item j .

Ces modèles sont très largement employés dans le domaine de l'évaluation des compétences, tant au niveau international (PISA, PIRLS) qu'au niveau national (CEDRE). Pour plus de précisions sur ces modèles, le lecteur pourra consulter, en français, Dicks *et al.* (1994).

Ajustement des métriques (equating)

Pour chaque session, les niveaux de compétence des élèves ont été estimés selon un MRI à deux paramètres. Cependant, les paramètres a_j , b_j et θ_j ne sont pas uniques : en appliquant une transformation linéaire donnée à ces paramètres, on peut obtenir une autre solution acceptable. Il faut donc imposer des contraintes identifiantes, par exemple, une moyenne nulle et un écart-type de 1 des θ_j , pour chaque session. Dans ce cas, la comparaison des niveaux de compétences est relative à chaque session : chacun des niveaux est ramené au niveau moyen de la session correspondante. Autrement dit, les trois jeux d'estimation des θ_j ne sont pas positionnés sur la même échelle. En s'appuyant sur les items communs, repris d'une épreuve à l'autre, il est possible de relier les niveaux de compétence selon une métrique commune aux trois sessions. De nombreuses méthodes ont été proposées pour réaliser cet ajustement des métriques (ou « equating ») à partir d'items communs. Nous avons retenu ici celle de Stocking et Lord (1983).

Les paramètres du modèle sont tout d'abord estimés indépendamment pour les trois sessions. Puis, nous avons cherché à déterminer la relation linéaire qui permette de transformer les niveaux de compétences de la deuxième session sur l'échelle de compétence de la première.

Remarquons tout d'abord que si on effectue le changement d'échelle $\theta^* = A\theta + B$, alors les nouveaux paramètres d'items $b^* = Ab + B$ et $a^* = a/A$ ne modifient pas la probabilité de réussite.

Il s'agit donc de proposer une procédure pour estimer A et B de manière à ce que les paramètres des items communs soient très proches, selon qu'on les estime lors de la première session ou lors de la deuxième. C'est le principe sous-jacent à l'equating dans le cadre des MRI : les paramètres d'items donnés sont indépendants des groupes d'élèves à partir desquels ils sont estimés. Autrement dit, ce sont les niveaux de compétence des élèves qui peuvent être différents d'un groupe à l'autre mais pas les paramètres des items. Plus précisément, soient (a_{j1}, b_{j1}) les paramètres de l'item commun j estimés lors de la première session, (a_{j2}, b_{j2}) les paramètres de l'item commun j estimés lors de la deuxième session et (a_{j2}^*, b_{j2}^*) les valeurs des paramètres estimés lors de la deuxième session et transformés sur l'échelle de la première. On cherche les « meilleurs » A et B tels que $b_{j2}^* = A b_{j2} + B$ et $a_{j2}^* = a_{j2} / A$, et tels que les b_{j2}^* (respectivement les a_{j2}^*) soient proches des b_{j1} (respectivement les a_{j1}). Pour cela, la procédure consiste à minimiser l'écart entre les deux scores « vrais », d'une estimation à l'autre (i.e. d'un jeu de paramètres à l'autre). Stocking et Lord (1983) proposent ainsi de minimiser F :

$$F = \frac{1}{n} \sum_{i=1}^n (\xi_i - \xi_i^*)^2$$

où ξ_i et ξ_i^* sont les scores « vrais » des n élèves, obtenus sur les J items communs : $\xi_i = \sum_{j=1}^J P_{j1}(\theta_i)$ et $\xi_i^* = \sum_{j=1}^J P_{j2}^*(\theta_i)$ avec P_{j1} et P_{j2}^*

les probabilités de réussite estimées respectivement avec les paramètres (a_{j1}, b_{j1}) et (a_{j2}^*, b_{j2}^*) .

Cela revient à dire qu'à niveau de compétence θ donné, les scores « vrais » obtenus aux items communs aux deux sessions devraient être sinon égaux, du moins très proches, quelle que soit la session considérée. Sur ce principe, A et B sont estimés, puis appliqués à la transformation des θ de la deuxième session sur l'échelle de compétence de la première.

La même démarche est employée pour relier les métriques des deuxième et troisième sessions. Au final, par transitivité, les niveaux de compétence sont tous mis sur la même échelle, quelle que soit la session.

RÉSULTATS

Participation

Le tableau 3 présente les chiffres de la participation des élèves aux évaluations, relativement à l'ensemble des élèves visés au départ. Globalement, 3 408 élèves ont participé aux trois sessions d'évaluation et 4 241 ont participé aux deux premières sessions, sur les 4 699 élèves attendus au départ. Les valeurs manquantes lors des deux premières sessions (458 élèves) s'expliquent, soit parce que les élèves ont été absents à l'une ou l'autre de ces sessions, soit parce que certains établissements n'ont pas fait passer une des sessions (deux établissements-témoins n'ont pas fait passer la première session).

La perte observée lors de la troisième session (soit 833 élèves) est due principalement (71 % des cas) aux élèves qui ont quitté l'établissement pour différentes raisons (déménagement, mesure de discipline, etc.), mais également aux élèves qui ont redoublé la 6^e (28 % des cas) ou qui ont sauté la classe de 5^e (1 % des cas) et qui ne sont donc pas concernés par l'évaluation

de fin de 5^e. À cette perte s'ajoutent des non-réponses : un établissement expérimentant l'EIST, et 5 établissements-témoins n'ont pas fait passer la session 3.

Le taux de participation aux trois sessions est finalement de 75,7 % des élèves du groupe 1 qui expérimentent l'EIST, de 70,9 % des élèves du groupe 3 – le groupe témoin – et seulement de 67,7 % des élèves du groupe 2.

Le groupe 2, constitué d'élèves qui reçoivent un enseignement traditionnel mais dans les collèges qui pratiquent l'EIST dans d'autres classes, a donc un effectif particulièrement faible (465 élèves ayant participé aux trois sessions). Ce groupe avait déjà un effectif plus faible au départ. Il devait concerner une classe de chaque établissement expérimentant l'EIST, mais une dizaine d'entre eux n'ont pas du tout évalué de classe suivant un enseignement traditionnel. C'est pourquoi ces élèves ont été écartés par la suite des analyses présentées.

Attrition

L'attrition concerne-t-elle les mêmes élèves selon les groupes ?

Le tableau 4 présente quelques

caractéristiques des élèves des groupes EIST et témoin selon deux cas : l'ensemble des élèves évalués au départ (N = 3 804) et ceux analysés ici, c'est-à-dire pour lesquels nous disposons de l'ensemble des résultats aux trois temps de mesure (N = 2 942).

L'observation des scores bruts obtenus aux épreuves de la session 1 montre que les élèves du groupe 1 (EIST) avaient au départ un niveau légèrement supérieur à celui des élèves du groupe 3 (témoin), avec un score moyen de 24,75 contre 24,27. Les deux échantillons présentent également des différences de structures, l'échantillon-témoin ayant été tiré de façon à être représentatif des 12 académies concernées par l'expérimentation, tandis que les établissements qui expérimentent l'EIST ont été sélectionnés sur la base du volontariat. Ainsi, le taux d'élèves ayant un retard scolaire à l'entrée en 6^e était un peu plus élevé pour le groupe-témoin avec 20,4 % d'élèves contre 19 % pour le groupe EIST. La répartition des établissements selon le secteur est également différente, le secteur privé étant sous-représenté dans l'échantillon EIST avec seulement 4 %

Tableau 3 – Participation des élèves

		EIST (groupe 1)		Groupe 2		Témoin (groupe 3)			Total			
		Étab. EIST	Élèves	Élèves	Étab. Non EIST	Élèves	Étab.	Élèves				
2008-2009	nb visé	35	2 035	100 %	687	100 %	81	1 977	100 %	116	4 699	100 %
Session 1	nb testé	35	1 960		635		79	1 845		114	4 440	
Session 2	nb testé	35	1 925		612		81	1 870		116	4 407	
Participation sessions 1 et 2		35	1 852	91,0 %	593	86,3 %	79	1 796	90,8 %	114	4 241	90,3 %
2009-2010	nb visé	35	1 798		605		81	1 750		116	4 153	
Session 3	nb testé	34	1 647		515		76	1 547		110	3 709	
Participation sessions 1, 2 et 3		34	1 541	75,7 %	465	67,7 %	74	1 402	70,9 %	108	3 408	72,5 %
2010-2011	nb visé	35	1 629		560		81	1 584		116	3 773	
Session 4	nb testé	35	1 435		486		78	1 361		113	3 282	

Source : MENJVA-DEPP

Lecture : le suivi de cohorte a concerné 2 035 élèves de 6^e suivant le dispositif expérimental EIST à la rentrée 2008. Parmi eux, 1 960 ont participé à l'épreuve de début 6^e, 1 925 ont participé à celle de fin de 6^e et 1 852 ont participé aux deux, ce qui représente 91 % des élèves visés au départ. En fin de 5^e (session 3), on comptabilise 1 541 élèves qui ont participé aux épreuves des trois sessions, soit 75,7 % des élèves visés au départ (pour rappel, les élèves ayant redoublé la 6^e n'ont pas été suivis).

Tableau 4 – Caractéristiques des groupes de répondants

3 804 élèves ayant passé la session 1				2 942 élèves ayant passé les trois sessions				
Scores bruts (session 1)				Scores bruts (session 1)				
	N	Moyenne	Écart-type	N	Moyenne	Écart-type		
EIST	1 959	24,75	17,40	1 540	25,26	15,30		
Témoin	1 845	24,27	13,30	1 402	24,79	12,90		
% d'élèves en retard				% d'élèves en retard				
	N			N				
EIST	1 959	19,0 %		1 540	16,7 %			
Témoin	1 845	20,4 %		1 402	18,6 %			
% selon le secteur				% selon le secteur				
	N	Public hors EP	Éducation prioritaire	Privé	N	Public hors EP	Éducation prioritaire	Privé
EIST	1 959	82,0 %	14,0 %	4,0 %	1 540	82,9 %	13,0 %	4,0 %
Témoin	1 845	65,4 %	15,7 %	18,9 %	1 402	67,9 %	14,2 %	17,9 %

Note : Les 1 540 élèves du dispositif EIST ayant participé aux trois sessions d'évaluation ont obtenu un score moyen de 25,26 à l'épreuve de science de la première session, alors que l'ensemble des élèves du dispositif EIST (1 959 élèves) évalués au départ lors de la première session ont obtenu un score moyen de 24,75. Cela signifie que les élèves « perdus » depuis la première session sont moins performants que les autres. Pour rappel, les élèves ayant redoublé la 6^e n'ont pas été suivis, ce qui explique cet écart de 0,51 points entre l'ensemble de départ et ceux observés aux trois sessions. L'intérêt du tableau est de montrer que cet écart est comparable parmi les élèves de l'échantillon témoin (0,52 = 24,79 - 24,27). Par conséquent, l'attrition a bien concerné des élèves plus faibles que la moyenne, mais ce phénomène est identique pour l'échantillon EIST et pour l'échantillon témoin.

Source : MENJVA-DEPP

des élèves de ce groupe contre 18,9 % des élèves du groupe 3.

En se restreignant maintenant aux 2 942 élèves ayant passé les trois sessions, les scores moyens de la session 1 sont légèrement supérieurs, ce qui s'explique notamment par le non-suivi des élèves redoublants. Cette augmentation s'observe de manière très comparable pour les deux groupes : EIST et témoin (de 27,75 à 25,26 pour les élèves du groupe EIST, et de 27,47 à 24,79 pour les élèves du groupe témoin). De la même façon, les différences de répartition entre les deux groupes, selon le retard scolaire ou selon le secteur de scolarisation, sont conservées lorsqu'on se restreint aux élèves ayant passé les trois sessions d'évaluation. Il n'y a donc pas lieu de conclure à une attrition différentielle selon les groupes, EIST ou témoin.

Résultats descriptifs

La figure 1 représente l'évolution des scores moyens obtenus par les deux groupes – EIST et témoin – en fonction des trois temps de mesure. Sept dimensions sont étudiées : le score cognitif en sciences et les

six facteurs conatifs. Les scores ont été standardisés de manière à ce que, pour les élèves du groupe témoin lors de la première session, la moyenne soit de 0 et l'écart-type de 1.

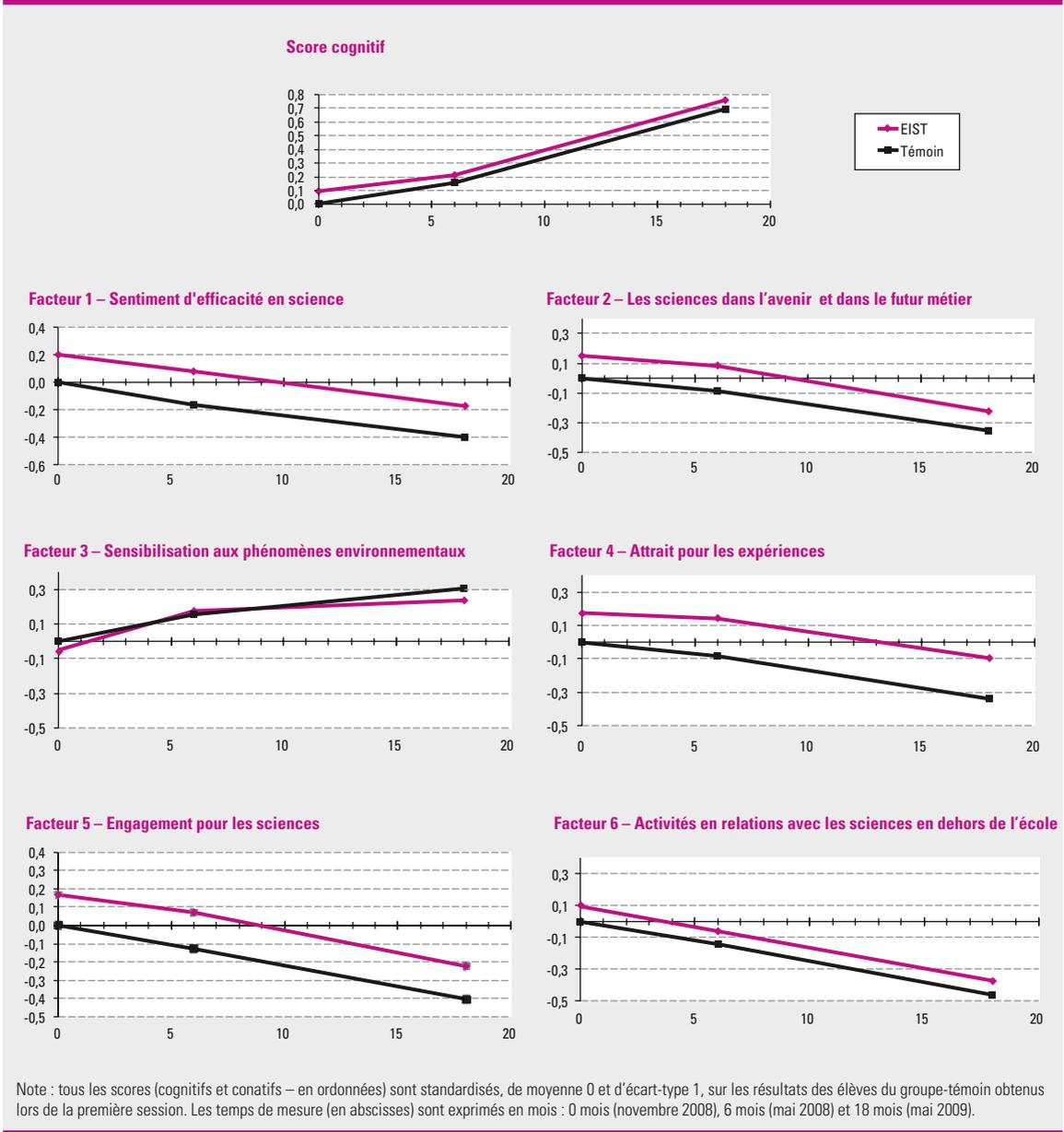
Il ressort que les scores moyens évoluent en fonction du temps de manière remarquablement proches pour les deux groupes (EIST et témoin). Les élèves ayant suivi le dispositif EIST affichent globalement une supériorité, à la fois sur le score cognitif et sur les facteurs conatifs. Mais l'écart qui sépare les deux groupes reste stable au cours du temps. Cette première analyse descriptive tend à montrer l'absence d'effet lié au dispositif EIST, que ce soit en termes de performances cognitives que d'attitudes à l'égard des sciences.

S'agissant du score cognitif, il apparaît une augmentation presque constante du niveau moyen au cours du temps sur l'échelle commune de sciences. L'année scolaire de 5^e correspond à une augmentation d'environ 0,55 écart-type de score, quel que soit le groupe considéré. Cette progression, associée à une année d'étude supplémentaire, est supérieure à ce

que l'on peut observer dans d'autres enquêtes : dans PISA par exemple, cette progression est estimée à environ 0,4 écart-type du score. Il faut rappeler que les écarts calculés ici ne portent pas sur les élèves ayant redoublé la 6^e, ce qui peut expliquer cette surestimation.

Le phénomène le plus frappant mis en exergue par ces premiers résultats est sans doute l'évolution des scores obtenus sur les variables de motivation et d'intérêt. À l'exception notable de la dimension « sensibilisation aux phénomènes environnementaux », toutes les autres affichent une tendance à la baisse du score moyen, et ce, quel que soit le groupe considéré (EIST et témoin). Une illustration de ce phénomène est fournie par le tableau 5 qui présente les réponses données à la question « Quand je serai adulte, j'aimerais continuer à faire de la science » du facteur « Les sciences dans l'avenir et dans le futur métier ». En début de 6^e, 77 % des élèves du dispositif EIST répondaient « d'accord » ou « tout à fait d'accord » à cette affirmation. En fin de 5^e, ces mêmes élèves n'étaient plus que 54 % à

Figure 1 – Évolution des résultats moyens selon le groupe (EIST versus témoin)



donner ces réponses positives. Parmi les élèves du groupe-témoin, ce taux passe de 61 % à 47 %.

Modélisations

Afin de préciser cette première analyse descriptive et de tester statistiquement l'absence ou la présence d'effet de l'EIST, des modélisations sont engagées. Notre objectif est d'étudier les progressions des élèves,

Tableau 5 : « Quand je serai adulte, j'aimerais continuer à faire de la science »

% de réponse	Non-réponse	Pas du tout d'accord	Pas d'accord	D'accord	Tout à fait d'accord
EIST					
Session 1 (début 6 ^e)	0,8	10,6	21,5	40,8	26,2
Session 2 (fin 6 ^e)	2,1	10,1	19,0	38,7	30,1
Session 3 (fin 5 ^e)	2,0	17,7	26,6	36,4	17,3
Témoin					
Session 1 (début 6 ^e)	1,1	13,0	24,9	42,3	18,7
Session 2 (fin 6 ^e)	2,9	12,8	23,0	41,5	19,8
Session 3 (fin 5 ^e)	2,1	19,3	31,1	34,4	13,1

Lecture : parmi les élèves du dispositif EIST et ayant participé aux trois sessions d'évaluation, 26,2 % déclaraient en début de 6^e être « tout à fait d'accord » avec l'affirmation « quand je serai adulte, j'aimerais continuer à faire de la science » ; ils ne sont plus que 17,3 % en fin de 5^e à être dans ce cas.

c'est-à-dire la relation entre le score – cognitif ou conatif – et le temps de mesure. Cette relation est variable d'un élève à l'autre, d'une classe à l'autre et éventuellement d'un groupe à un autre, EIST et témoin.

Nous avons eu recours à un modèle multiniveau de croissance (*multilevel growth curve model*). Ces modèles prennent en compte la structure hiérarchique des données, en l'occurrence organisée sur trois niveaux : temps de mesure/élève/collège. Ils sont très adaptés aux données recueillies ici. Pour une présentation en français de ces modèles et de leurs applications possibles dans le champ de l'éducation, le lecteur est invité à lire Bressoux (2010).

Le tableau 6 présente les résultats de la modélisation appliquée au score cognitif ainsi qu'aux six facteurs conatifs. Concernant le premier

modèle, qui porte sur le score cognitif, la partie « Effets fixes » indique une relation positive et significative entre le score et le temps de mesure, ce qui traduit le développement au cours du temps des performances des élèves sur l'échelle commune de sciences. La variable indicatrice EIST (1 pour EIST, 0 pour témoin) indique une absence de différence entre EIST et témoin, lors de la première mesure. Graphiquement, il apparaît que les élèves bénéficiant de l'EIST affichent des scores supérieurs, quel que soit le temps de mesure (*figure 1*). En fait, cette supériorité semble être due à un effet de structure, puisqu'elle devient non significative lorsqu'est introduite dans le modèle la répartition selon le type de scolarisation (privé, éducation prioritaire).

La question d'importance ici est celle de l'interaction entre le

fait d'avoir bénéficié de l'EIST et le temps. Il ressort que cette interaction n'est pas significative, indiquant que la relation entre le score et le temps n'est pas différente selon que les élèves ont ou non bénéficié de l'EIST. Plus généralement, cette interaction n'est pas significative, quelle que soit la dimension interrogée.

La partie « Effet aléatoire » rapporte quant à elle les variances des différents paramètres aléatoires du modèle. La variance du niveau 1 indique la variabilité intra-élèves, celle du niveau 2 la variabilité inter-élèves et celle du niveau 3, la variabilité inter-collèges. Pour le score cognitif par exemple, des variances significatives sont observées à ces trois niveaux, et concernent à la fois les écarts de performance entre élèves ou entre collèges (les constantes) et les écarts de progressions (les pentes).

Tableau 6 – Évaluation de l'effet (modèles de croissance)

	score cognitif	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5	facteur 6
Effets fixes							
Constante	0,002 (0,061)	0,035 (0,046)	0,098 (0,039)**	0,069 (0,042)	0,023 (0,036)	-0,058 (0,039)	0,079 (0,037)**
Temps	0,039 (0,003)***	-0,022 (0,002)***	-0,02 (0,002)***	0,016 (0,002)***	-0,019 (0,002)***	-0,022 (0,002)***	-0,026 (0,002)***
EIST	0,103 (0,083)	0,2 (0,059)***	0,155 (0,047)***	-0,012 (0,054)	0,195 (0,044)***	0,185 (0,048)***	0,082 (0,044)*
EIST*Temps	-0,002 (0,005)	-0,0003 (0,004)	-0,0015 (0,003)	-0,0014 (0,002)	0,0028 (0,003)	-0,0011 (0,004)	0,0001 (0,002)
EP	-0,58 (0,108)***	0,077 (0,077)	0,067 (0,062)	-0,197 (0,069)***	0,044 (0,057)	0,06 (0,063)	0,041 (0,057)
Privé	0,033 (0,110)	0,035 (0,079)	0,01 (0,064)	0,057 (0,071)	0,051 (0,059)	0,061 (0,064)	0,036 (0,059)
Fille	0,058 (0,033)*	-0,128 (0,032)***	-0,19 (0,031)***	-0,078 (0,029)***	-0,055 (0,029)*	0,084 (0,031)***	-0,163 (0,032)***
Effets aléatoires							
Niveau 3 constante	0,124 (0,022)***	0,0463 (0,011)***	0,0203 (0,007)***	0,0372 (0,009)***	0,016 (0,006)***	0,0201 (0,007)***	0,0118 (0,005)**
(interclasses) pente	0,00047 (0,00009)***	0,0001 (0,00004)***	0,0001 (0,00003)**	0,00002 (0,00002)	0,00006 (0,00002)***	0,00013 (0,00004)***	0,00003 (0,00002)*
Niveau 2 constante	0,582 (0,021)***	0,4744 (0,020)***	0,4889 (0,019)***	0,3957 (0,017)***	0,4052 (0,016)***	0,4362 (0,019)***	0,5341 (0,020)***
(interélèves) pente	0,00103 (0,00009)***	0,0011 (0,0001)***	0,0006 (0,0001)***	0,0002 (0,0001)**	0,0002 (0,0001)**	0,0009 (0,0001)***	0,0001 (0,0001)**
Niveau 1	0,433 (0,010)***	0,5475 (0,013)***	0,5202 (0,013)***	0,6067 (0,014)***	0,504 (0,012)***	0,6016 (0,014)***	0,5238 (0,013)***
-2logL	23920.1	23651.2	23213.4	23145.3	21919.8	23998.0	23102.9

*significatif au seuil de 10 %

**significatif au seuil de 5 %

***significatif au seuil de 1 %

Note : le tableau présente pour chacune des dimensions visées les principaux résultats des modèles multiniveaux de croissance, à savoir les valeurs des paramètres ainsi que leurs erreurs associées (entre parenthèses) et leur degré de significativité indiqué par des « * ». Dans la partie « Effets fixes », il ressort que le coefficient associé au temps de mesure est très significatif pour toutes les dimensions, ce qui traduit leur développement au cours du temps : les progressions sont positives pour les performances cognitives et le facteur 3 (sensibilisation aux phénomènes environnementaux) et négatives pour les autres dimensions (attitudes à l'égard des sciences). La variable indicatrice EIST (1 pour EIST, 0 pour témoin) indique la différence entre les élèves des deux groupes (EIST et témoin), lors de la première session. L'interaction entre la variable EIST et le temps n'est pas significative, quelle que soit la dimension : les progressions ne sont pas différentes selon que les élèves ont ou non bénéficié de l'EIST. Dans la partie « Effet aléatoire », la variance du niveau 1 indique la variabilité intra-élèves, celle du niveau 2 la variabilité interélèves et celle du niveau 3, la variabilité intercollèges.

CONCLUSIONS ET PERSPECTIVES

Les premières analyses de l'effet de l'expérimentation EIST sur les progressions des élèves, observées sur trois temps de mesure, du début de 6^e à la fin de 5^e, ne font pas ressortir d'effet significatif associé au fait d'avoir bénéficié du dispositif de l'EIST. Cette absence d'effet porte tant sur les progressions en termes d'acquis cognitifs dans le domaine des sciences qu'en termes d'attitudes à l'égard des sciences.

Plusieurs remarques importantes méritent d'être apportées pour éclairer ces premiers résultats. Tout d'abord, rappelons que les collègues engagés dans l'EIST sont des collègues volontaires, et qu'à ce titre, on pouvait craindre l'apparition d'un effet positif de l'expérimentation qui soit « artefactuel », c'est-à-dire ne traduisant pas les effets du dispositif lui-même mais celui de caractéristiques non observées des collègues, en lien avec leur adhésion à l'EIST. Cet effet n'apparaît pas, ce qui évacue mécaniquement la question de la surestimation de l'effet liée au volontariat.

Une deuxième série de commentaires concerne la temporalité de l'évaluation, qui concerne l'évolution des performances et des attitudes du début de 6^e à la fin de 5^e. Il est tout à fait possible que les effets ne soient pas immédiats et puissent s'observer avec un temps de latence. À cet égard, il conviendra de poursuivre les analyses de progressions, en fin de 4^e et en fin de 3^e. Cette question est d'autant plus importante que les instruments de mesure des performances en sciences intègrent à ces deux derniers niveaux, un volume plus important de questionnement ouvert et des éléments plus fournis sur la démarche d'investigation.

Troisièmement, nous avons traité ici sur le même plan tous les élèves ayant suivi le dispositif EIST. Mais le temps d'exposition à l'expérimentation semble variable : certains élèves auraient profité du dispositif EIST pendant toute l'année scolaire, tandis que d'autres en auraient bénéficié seulement pendant un ou deux trimestres. Cette variable sera précisée et introduite dans les futures analyses. Cependant, il est

peu probable que ce phénomène soit de nature à dissimuler des effets importants.

Parallèlement, ce travail a été l'occasion de présenter des approches liées à la mesure et au suivi des dimensions, en ayant recours à des modélisations psychométriques. Ce champ sera également développé en visant une modélisation intégrée comprenant à la fois le modèle de mesure et le modèle d'évaluation de l'effet. Ceci permettra de mieux tenir compte des erreurs de mesure.

Enfin, signalons que ce suivi longitudinal a permis de constituer une base de données très riche sur le développement des performances et des attitudes à l'égard des sciences au cours des années de collège. Un premier enseignement original concerne la désaffection pour les sciences qui apparaît dès le début du collège à travers la baisse observée sur les variables d'intérêt et de motivation à l'égard des sciences. Ainsi, ces données mériteront d'être analysées plus largement, au-delà de la question de l'évaluation des effets de l'EIST. ■

BIBLIOGRAPHIE

- Bressoux, P.** (2010), *Modélisation statistique appliquée aux sciences sociales*, De Boeck, 2^e édition, Bruxelles.
- Dickes P., Tournois J., Flieller A. et Kop J.-L.** (1994), *La psychométrie : théorie et pratique de la mesure en psychologie*, PUF, Paris.
- Stocking M.L. et Lord F.M.** (1983), Developing a Common Metric in Item Response Theory, *Applied Psychological Measurement*, n° 7, pp. 201-210.

Annexe 1 : détail des groupes suivis

Groupe 1

Tous les élèves de France métropolitaine recevant un enseignement EIST sont évalués de façon exhaustive. Les collèges participant à l'EIST se situent dans les 12 académies suivantes : Clermont-Ferrand, Nice, Toulouse, Montpellier, Poitiers, Bordeaux, Orléans-Tours, Nancy-Metz, Lille, Strasbourg, Versailles, Créteil. S'agissant des élèves de 6^e, cela concerne 2 035 élèves répartis dans 35 collèges (cf. tableau 3).

Groupe 2

Dans les établissements participant au programme EIST, une classe suivant un enseignement « traditionnel » est tirée au hasard, et tous les élèves de cette classe participent à l'évaluation.

Groupe 3

Ce groupe constitue l'échantillon-témoin. Il regroupe des élèves d'établissements ne participant pas à l'expérimentation de l'EIST. Dans la mesure où les collèges participant à l'EIST se situent dans 12 académies, et afin de simplifier la communication avec les différents interlocuteurs, l'échantillon-témoin a été tiré dans ces 12 académies uniquement. Une analyse a d'ailleurs montré que la population des 12 académies est représentative de la population totale (France métropolitaine). Un échantillon stratifié selon le secteur de l'établissement (public hors éducation prioritaire, éducation prioritaire, privé) est alors obtenu par un tirage à deux degrés. Le premier degré consiste à tirer des établissements proportionnellement à leur taille (en nombre de classes), et le deuxième consiste à tirer, de façon aléatoire, une classe dans chacun de ces établissements. L'échantillon-témoin est constitué de manière à être de même taille que celui des élèves bénéficiant du programme d'EIST (soit environ 2 000 élèves en 6^e et environ 500 élèves en 5^e), et que la répartition selon les trois strates soit représentative.

Annexe 2 : questionnaire conatif

Facteur 1 : « sentiment d'efficacité en science »

Ce que je fais en science est intéressant.
La science, c'est trop difficile pour moi.
Je pense que j'ai un bon niveau en science.
Je comprends bien ce que nous faisons en science.
Je pense que je peux réussir en science.

Facteur 2 : « Les sciences dans l'avenir et dans le futur métier »

Quand je serai adulte, j'aimerais continuer à faire de la science.
Quand je serai adulte, j'aimerais exercer une profession dans le domaine scientifique.
Quand je serai adulte, j'aimerais concevoir des objets techniques.
Quand je serai adulte, j'aimerais être chercheur.

Facteur 3 : « Sensibilisation aux phénomènes environnementaux »

*L'augmentation des gaz à effet de serre dans l'atmosphère**.*
*Les pénuries d'eau**.*
*Les déchets nucléaires**.*
*La destruction des forêts**.*
*La disparition de certaines espèces de plantes et d'animaux**.*

Facteur 4 : « attrait pour les expériences »

Je refais chez moi des expériences réalisées en classe.*
J'aime les jouets à caractère scientifique.*
Je suis capable d'imaginer une expérience.
Faire des sciences au collège, c'est faire des expériences par moi-même.

Facteur 5 : « engagement pour les sciences »

Je participe en science parce que j'aime bien chercher.
Je travaille en science parce que j'aime bien cette discipline.
Faire des sciences au collège, c'est se poser des questions sur le monde qui nous entoure.
Faire des sciences au collège, c'est chercher des informations dans les livres ou sur internet.
Faire des sciences au collège, c'est expliquer ce que j'ai compris.

Facteur 6 : « Activités en relations avec les sciences en dehors de l'école »

Je regarde des émissions scientifiques à la télévision.*
J'aime lire des livres ou des revues scientifiques.*
Je cherche des documents scientifiques sur internet.*

Note : questions sans astérisque : les élèves devaient répondre « tout à fait d'accord » « d'accord », « pas d'accord » ou « pas du tout d'accord » ; questions marquées d'un astérisque « * » : les élèves devaient répondre « très souvent », « souvent », « parfois » ou « jamais » ; questions marquées de deux astérisques « ** » : les élèves devaient répondre « je connais ce sujet et je pourrais assez bien l'expliquer », « j'en ai entendu parler mais je serais incapable d'expliquer ce que c'est exactement », « je n'en ai jamais entendu parler ».

PACEM : une expérimentation sur l'utilisation d'évaluations standardisées des acquis des élèves par les enseignants

Jean-François Chesné, Sandrine Prost

Bureau de l'évaluation des actions éducatives et des expérimentations
DEPP, MENJVA

Avec la collaboration d'**Arthur Souletie** (ENS Cachan)

et de **Christelle Raffaelli** (Bureau de l'évaluation des actions éducatives et des expérimentations, DEPP, MENJVA)

Le Projet pour l'acquisition de compétences par les élèves en mathématiques (PACEM) est un projet de la direction de l'évaluation, de la prospective et de la performance (DEPP) mené en CM1 et en sixième en mathématiques.

Il intègre quatre dimensions principales : l'évaluation des élèves, la formation des enseignants, l'observation de leurs pratiques individuelles dans les classes et l'étude des dynamiques entre les enseignants au sein d'un même établissement. Son objectif principal vise des changements de pratiques induits par une formation, elle-même ancrée sur les résultats des élèves à un test de début d'année scolaire.

À l'issue de la première année d'expérimentation, nous pouvons dire que, par comparaison à la progression d'élèves témoins qui n'ont pas suivi le protocole d'expérimentation, la performance des élèves expérimentateurs est significativement meilleure pour la sixième et très significativement meilleure pour le CM1. Cela tendrait à montrer qu'un investissement relativement faible mais spécifique de formation continue et d'accompagnement des enseignants conduit à une amélioration sensible des performances scolaires des élèves.

L'expérimentation PACEM, conçue et développée par Jean-François Chesné, avec l'appui statistique de Sandrine Prost, est un projet mené par le bureau de l'évaluation des actions éducatives et des expérimentations de la DEPP. Celui-ci a pris à sa charge l'intégralité de son développement : conception et mise en œuvre des évaluations des élèves et des formations à destination des enseignants. Si cette singularité caractérise l'étude, elle ne la rend pas moins représentative d'une partie de l'activité des différents bureaux de la DEPP. Innovante et experte, elle s'inscrit dans la dynamique d'autres études comme CEDRE¹ ou Victimation, conduites respectivement par le bureau de l'évaluation des élèves et le bureau des études sur les établissements et l'éducation prioritaire de la DEPP. Soutenue par les membres du groupe de mathématiques de l'inspection générale de l'Éducation nationale (IGEN), elle a bénéficié des appuis académiques et départementaux indispensables à son implantation et à son déploiement dans les établissements scolaires sélectionnés, ainsi que de la

collaboration de chercheurs des universités Denis Diderot et René Descartes de Paris.

Après avoir exposé les fondements et les enjeux de l'expérimentation PACEM, nous en décrivons la mise en œuvre et le développement au cours de l'année 2010-2011 au CM1, puis nous en présenterons les premiers résultats sur ce niveau. Nous énoncerons enfin nos premières interprétations, ainsi que les perspectives envisagées à court et à moyen termes et les nouvelles questions sur les conditions d'intégration des évaluations nationales dans les pratiques des enseignants.

FONDEMENTS ET ENJEUX

Un contexte d'évolution des pratiques

Cette expérimentation s'inscrit d'une part dans un contexte de questionnement sur la réussite des

NOTE

1. CEDRE : cycle d'évaluations disciplinaires réalisées sur échantillon

élèves et sur l'évolution des pratiques enseignantes, d'autre part dans un contexte d'accompagnement de la mise en place de nouveaux objectifs institutionnels.

En effet, un certain nombre d'évaluations nationales ou internationales antérieures conduites par la DEPP a montré des résultats convergents : parmi eux, nous citerons l'augmentation du nombre d'élèves de bas niveau et l'impact de leur environnement socioéconomique sur leur réussite scolaire [1]. Bien que diffusés par le biais de notes d'informations et de dossiers publiés par la DEPP, il est légitime de s'interroger sur leur prise en compte par les enseignants dans leurs pratiques quotidiennes. De façon générale, nous faisons l'hypothèse que, si les évaluations à large échelle (sur échantillons ou exhaustives) ont de plus en plus d'importance dans le pilotage du système éducatif français, la plupart des enseignants semblent considérer qu'elles constituent davantage un outil externe de mesure des acquis des élèves plutôt qu'un ensemble de ressources pour la classe [2].

Dans le même temps, les six dernières années ont vu, avec l'apparition du socle commun de connaissances et de compétences, et la réécriture des programmes de l'école et du collège, des évolutions sensibles dans l'organisation des enseignements et des pratiques pédagogiques souhaitées [3]. L'expérimentation PACEM permet, en elle-même, de porter un regard sur la question de la pénétration de ces nouveaux textes auprès du corps enseignant, de leur interprétation et de l'esprit dans lequel ils peuvent être mis en œuvre.

Enfin, cette expérimentation renvoie à la mise en œuvre des évaluations nationales de CE1 et de CM2 (et bientôt de 5^e) qui posent la ques-

tion, au-delà du retour chiffré de leurs résultats, de leur exploitation locale par les enseignants [4].

Parallèlement à ces questionnements, des dynamiques institutionnelles existent avec une volonté affirmée de faire progresser les acquis des élèves en mathématiques et d'accompagner les enseignants dans l'évolution de leurs pratiques : documents ressources pour les enseignants [5], action nationale en direction des inspecteurs de l'Éducation nationale (IEN) visant à piloter des missions mathématiques dans chaque département.

Un cadre d'expérimentation étayé

Si une expérience de la formation initiale continue et de la liaison école/collège, et une connaissance précise du monde enseignant, ont motivé la nature et les contenus de cette démarche d'expérimentation, il est nécessaire de préciser que celle-ci prend ancrage dans un cadre théorique qui est celui de la double approche, conçu et développé par Robert et Rogalski [6]. Celui-ci permet d'analyser les pratiques des enseignants, relativement aux résultats de leurs élèves en prenant en compte les tâches proposées et les contraintes liées à l'exercice du métier au quotidien, afin de déterminer ce qui peut varier dans ces pratiques [7].

Des hypothèses sur la formation des enseignants mettant en jeu l'efficacité du travail dans une « Zone proximale de développement des pratiques » nous ont conduits ensuite à élaborer des éléments de formation [8].

Des résultats institutionnels, ainsi qu'issus de la recherche, servent également de fondements à ce pro-

jet, notamment les rapports de l'IGEN sur l'école primaire [9], les travaux de Douady et de Perrin-Glorian sur les aires [10], ceux de Butlen sur le calcul mental [11] ou encore ceux de Peltier [12] et de Van Zanten sur l'éducation prioritaire [13].

Enfin, cette étude s'appuie sur la conviction forte que l'on peut « changer les choses » lorsque l'on est enseignant, ce qui n'est en fait qu'une façon de réaffirmer l'impact primordial et identifié par de nombreuses études de l'effet enseignant dans la réussite des élèves [14].

Dès lors, il apparaît très clairement que l'enjeu de cette étude, qui fait s'entrecroiser un travail sur la formation des enseignants et un travail sur les évaluations des élèves, sera d'apporter des réponses sur la question de l'appropriation des résultats d'évaluations standardisées par les enseignants et de s'interroger sur la nature des effets qu'elle peut produire sur leurs pratiques [15].

PRÉSENTATION DU DISPOSITIF

Lancée en septembre 2010, l'expérimentation se terminera en juin 2012. Sa mise en œuvre en 2010-2011 s'est faite dans deux académies différentes et a concerné deux niveaux de classe : le CM1 dans l'académie d'Aix-Marseille, dans des écoles relevant presque toutes de l'éducation prioritaire, et la sixième, dans l'académie de Créteil, sur un échantillon de collèges relevant pour moitié environ de l'éducation prioritaire. En 2011-2012, le dispositif est reconduit en CM1 et en sixième avec le même protocole et étendu aux niveaux CM2 et cinquième en suivant les élèves

testés en 2010-2011. Nous aurons donc ainsi à la fois la possibilité d'étudier son impact sur deux cohortes d'élèves suivies sur deux années et de comparer l'évolution des acquis des élèves. Complétant ces données sur les acquis des élèves, nous disposons également d'éléments d'information sur les pratiques des enseignants : observations et enregistrements audio ou vidéo de séances, réponses à un questionnaire en ligne.

Dans cet article, outre la présentation du dispositif, les résultats exposés ci-après concerneront le premier volet de l'expérimentation menée dans l'académie d'Aix-Marseille en 2010-2011. Après une brève description de la population, nous étudierons plus précisément l'évolution des taux de réussite des élèves aux items d'ancrage d'une part, l'évolution des scores de réussite entre les deux phases d'expérimentation d'autre part, dans le but de voir l'impact de l'expérimentation sur la réussite des élèves de CM1 en mathématiques.

La mise en œuvre de PACEM sur le niveau CM1 en 2010-2011

L'étude s'est concentrée sur un domaine du programme de mathématiques – Grandeurs et mesures – choisi par le groupe de pilotage des Bouches-du-Rhône en mathématiques pour le 1^{er} degré. On peut considérer qu'il s'agit d'un domaine nouveau puisqu'après plus de 30 ans d'absence, il fait sa réapparition en 2002 dans les programmes de mathématiques comme domaine d'étude. L'immense majorité des enseignants actuels n'en possède donc qu'une culture assez restreinte, tant du point de vue du savoir mathématique que des contenus à enseigner.

L'observation des résultats a porté sur trois circonscriptions : deux sont engagées dans l'expérimentation, Marseille 11 et Marseille 12, tandis que l'autre, celle d'Istres, constitue la circonscription témoin.

L'expérimentation PACEM repose sur un certain nombre de principes organisationnels :

- deux tests proposés aux élèves : l'un en début (test 1), l'autre en fin d'année scolaire (test 2) ;
- une formation proposée à certains enseignants ;
- l'accès à un site collaboratif pour les enseignants ;
- un travail de mutualisation des pratiques pour les enseignants des écoles engagées, dans le projet ;
- une collaboration de la DEPP avec les IEN, les conseillers pédagogiques et le groupe de pilotage départemental mathématiques.

La population concernée par le projet

Tous les enseignants ayant une classe de CM1, ou un double niveau en 2010-2011, dans une école des circonscriptions de Marseille 11, Marseille 12 et d'Istres, sont impli-

qués, soit 86 enseignants. Parmi ceux-ci, nous distinguerons trois catégories :

- 25 enseignants expérimentateurs qui ont bénéficié directement d'une formation : ce sont les « correspondants » du projet ;
- 27 autres enseignants expérimentateurs, dits « associés » au projet : ce sont les enseignants des écoles engagées dans le projet qui n'ont pas participé à une formation spécifique liée au projet ;
- 34 enseignants « témoins ».

Au total, l'expérimentation concerne 11 écoles et 24 classes pour Marseille 11, 14 écoles et 28 classes pour Marseille 12, et 15 écoles pour 34 classes dans la circonscription d'Istres.

Le nombre d'élèves ayant participé à l'expérimentation diffère entre les deux tests : 1 568 élèves ont participé à la première phase et 1 527 à la deuxième phase. Afin de comparer les résultats des élèves lors des deux phases d'expérimentation, nos analyses porteront uniquement sur les élèves qui ont été présents aux deux tests, soit exactement 1 400 élèves.

Graphique 1 – Caractéristiques démographiques des élèves de CM1 testés dans l'expérimentation PACEM en 2010-2011 (en %)

Marseille 12	Zone d'éducation	EP 70,6		Non EP 29,4
	Age de scolarité	2,2	À l'heure 80,1	
		Avance		Retard 17,7
	Sexe	Garçon 47,8	Fille 52,2	
Marseille 11	Zone d'éducation	EP 92,4		Non EP 7,6
	Age de scolarité	1,5	À l'heure 76,8	
		Avance		Retard 21,7
	Sexe	Garçon 49,5	Fille 50,5	
Istres	Zone d'éducation	10,8	Non EP 89,2	
	Age de scolarité	2,8	À l'heure 86,6	
		Avance		Retard 10,6
	Sexe	Garçon 53,8	Fille 46,2	

Source : MENJVA-DEPP.

Le groupe des élèves dits « expérimentateurs » représente 58,9 % de l'effectif total réparti sur deux circonscriptions Marseille 11 (N = 406) et Marseille 12 (N = 418). Quant au groupe des élèves dits « témoins », il représente 41,1 % des élèves (N = 576) participant à l'expérimentation.

Globalement, la part des garçons est équivalente à celle des filles (50,8 % vs 49,2 %)². La grande majorité des élèves scolarisée est « à l'heure » (81,9 %), 15,9 % sont en retard et 2,2 % en avance. C'est à Marseille 11 que la part des élèves scolarisés en retard est la plus importante (21,7 % vs 17,7 % à Marseille 12 et 10,6 % à Istres (*graphique 1*), alors qu'Istres enregistre une plus grande proportion d'élèves scolarisés en avance (2,8 % contre 2,2 % à Marseille 12 et 1,5 % à Marseille 11).

Enfin, les élèves scolarisés en éducation prioritaire représentent un peu plus de la moitié de notre population (52,3 %). Ils sont très largement surreprésentés à Marseille 11 et sous-représentés à Istres (92,4 % à Marseille 11 contre 10,8 % à Istres).

Les cahiers d'évaluation

Les cahiers des tests 1 et 2 ont une structure similaire, seul le nombre d'items change. Ils se décomposent en deux parties de 30 minutes de passation : la première, sur les unités de mesures et la seconde sur la résolution de problèmes. Le test 1 comporte 43 items (31 items dans la première partie et 12 dans la seconde) alors que le test de la seconde phase comprend 45 items (30 items dans la première partie et 15 dans la deuxième).

Dans la première partie, les élèves ont, soit à répondre à des questions à choix multiples (QCM), soit à produire des réponses courtes. La deuxième partie ne comporte que des QCM.

Les contenus des deux tests sont évidemment différents, à la fois pour ne pas favoriser une préparation des élèves au test 2 aux dépens de leur apprentissage, mais aussi pour tenir compte de l'année scolaire écoulée. Cependant, 7 items sont communs aux deux tests dans la perspective d'une comparaison temporelle (items d'ancrage).

Les contenus sont des tâches correspondant aux connaissances et aux capacités indiquées dans les programmes de CE2 et de CM1 : la connaissance des unités de mesure et les relations qui les lient, la lecture de l'heure sur une montre à aiguilles ou une horloge et l'utilisation de la règle graduée pour mesurer des longueurs (partie 1 des tests) ; le calcul du périmètre d'un polygone et la résolution de problèmes (parties 2a et 2b). Le rythme des exercices n'était pas imposé. Des indications régulières sur la durée restante étaient données aux élèves (toutes les 10 minutes, et 2 minutes avant la fin des tests).

Des indications pouvaient être fournies aux élèves seulement à partir du moment où elles favorisaient une activité mathématique sans la déformer. Par exemple, les enseignants pouvaient indiquer aux élèves ce qu'est « un écolier » dans l'exercice 1, ou « un alpiniste » ou « une paroi » dans l'exercice 13, mais ne devaient pas répondre aux questions du type « Qu'est-ce que c'est un centilitre ? », « Qu'est-ce que c'est les kilogrammes », « Il y a combien d'heures dans une journée ? », « Qu'est-ce que c'est un segment ? » ou « Qu'est-ce que c'est le périmètre ? ».

Dans l'exercice 10, les enseignants pouvaient, si les élèves le demandaient, dire que « polygone » signifie

« figure ». Chaque question posée par un élève devait être reprise pour toute la classe et les indications données à l'ensemble des élèves.

Les conditions habituelles de travail d'élèves à profil particulier (non lecteurs ou en situation de handicap) ont été respectées : selon les cas, les énoncés des exercices ont été lus individuellement, une durée supplémentaire de passation a été accordée ou l'élève a été aidé par un(e) assistant(e) de vie scolaire (AVS).

Le contenu et la passation des tests sont donc très similaires à d'autres tests nationaux passés par les élèves. Nous sommes toutefois conscients des limites de ce type d'évaluation, en particulier sur le thème des grandeurs et des mesures, puisque les modalités d'évaluation retenues ne renvoient sans doute que partiellement aux compétences potentiellement acquises par les élèves sur l'ensemble d'une année ou d'un cycle scolaire.

La formation

La caractéristique principale de cette expérimentation est de lier la formation des enseignants à l'analyse des items proposés dans les tests et de leurs résultats, et à l'intégration de leurs progressions annuelles dans leur classe. Cette formation est destinée à favoriser, chez les enseignants, l'appropriation d'évaluations standardisées et de les conduire progressivement à modifier leurs pratiques ordinaires individuelles. Elle a également vocation à accentuer la dynamique d'équipe en suggérant des échanges plus réguliers

NOTE

2. Les différences observées au sein de chaque circonscription ne sont pas significatives.

sur l'utilisation d'outils ou la mise en œuvre de pratiques innovantes.

Le projet, perçu dans sa globalité, comporte une dimension à la fois individuelle et collective. Au niveau de la classe, il vise la prise en compte, par chaque enseignant, des réussites des élèves, mais aussi la compréhension de leurs difficultés afin de mettre en œuvre des stratégies pour les traiter. Les apports reçus et les échanges entre collègues lors de la formation ont pour objectif de les amener à adapter leur enseignement quotidien au sein de la classe, à construire de nouvelles séquences pour leur classe, à les penser différemment dans la progression annuelle. Ce questionnement des pratiques personnelles devrait permettre d'enclencher une dynamique à une autre échelle, celle de l'école, puisque ces démarches demandent à être partagées, questionnées, transférées. Ce travail peut même aller jusqu'à la construction d'un projet commun d'enseignement sur le cycle 3.

L'accompagnement des enseignants pendant le projet se décline en trois composantes. La première est une formation minimale en présentiel : une demi-journée de lancement en septembre 2010 pour tous les enseignants expérimentateurs, suivie de deux jours début octobre 2010 réservés aux correspondants. Au cours de l'année, les enseignants expérimentateurs ont accès à une plate-forme collaborative et un suivi local est effectué dans les écoles par les IEN et les conseillers pédagogiques. Le cœur de ce dispositif, le moment de formation en présentiel, s'articule autour de quatre temps :

- une phase de déconstruction : il s'agit d'amener les enseignants à accepter de lâcher prise sur cer-

taines de leurs pratiques et de leurs représentations, de « faire de la place pour du nouveau ». L'utilisation d'éléments statistiques lors de ce premier temps de formation s'avère très utile, permettant d'interroger et donc d'objectiver la difficulté d'une tâche ;

- une phase d'apport de contenus de la part du formateur : contenus mathématiques et éléments de didactique « simples », liens avec le programme, explicitation des capacités attendues ;

- une phase d'appropriation : des tâches alternatives à celles des fichiers et manuels utilisés sont proposées aux enseignants, qu'ils ont à mettre en situation, à travers différents scénarios, potentiellement adaptables à la classe ;

- une phase d'organisation, de structuration et de contextualisation : ce dernier temps de la formation fait écho au premier puisqu'il s'agit de proposer aux enseignants des éléments leur permettant d'une part, de reconstruire des séances et, à plus longue échéance, des séquences, et d'autre part, de réviser la façon dont ils concevaient jusqu'alors leurs progressions.

PREMIERS RÉSULTATS AU CM1 POUR 2010-2011

Résultats par item

Les scores de réussite obtenus au test 1 ont fait l'objet d'une attention particulière puisque les items des cahiers d'évaluation des élèves ont été choisis à la fois comme caractéristiques des attendus (partiels) au regard du programme de CE2 et dans la perspective d'une utilisation

en formation. Une analyse item par item a donc été réalisée, et communiquée aux enseignants expérimentateurs dans un document transmis aux IEN et déposé sur le site collaboratif.

Analyse des items d'ancrage

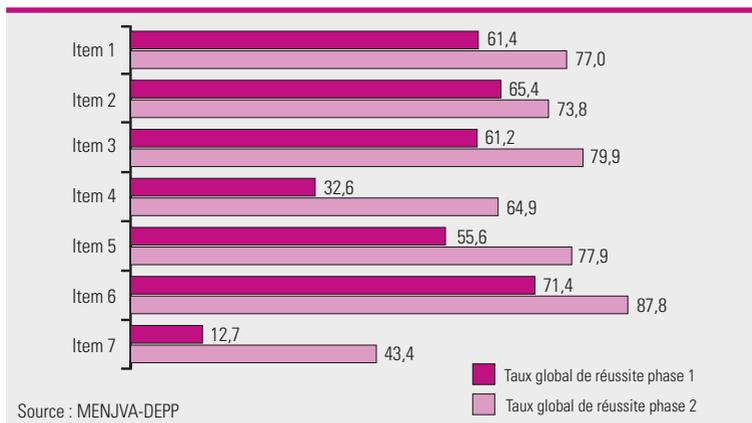
Les scores obtenus aux sept « items d'ancrage » – dont les énoncés sont présentés dans l'encadré page suivante –, permettent de comparer les résultats obtenus à ces deux tests, et donc de déterminer l'impact de l'expérimentation sur l'évolution des acquis des élèves [16].

Pour chaque item d'ancrage, on observe globalement une très nette amélioration des résultats des élèves entre les phases 1 et 2, allant de + 8,4 points pour l'item 2 à + 32,4 points pour l'item 4 (*graphique 2*). Ce sont les items 4, 5 et 7 qui enregistrent la plus forte progression.

En effet, l'item 4, qui avait été réussi par 32,6 % des élèves lors de la phase 1, l'a été par 64,9 % lors de la phase 2. De même, les items 5 et 7, réussis respectivement par 55,6 % et 12,7 % d'élèves lors de la phase 1, l'ont été par respectivement 77,9 % et 43,4 % des élèves à la phase 2. Notons que l'item 7, seul item d'ancrage portant sur la résolution de problèmes, est nettement moins bien réussi aux deux phases que les autres items d'ancrage qui portaient sur des unités de mesure. Il enregistre néanmoins une progression importante entre les deux phases de 30,7 points.

Cette progression n'est cependant pas homogène entre les trois circonscriptions (*graphique 3*). En effet, c'est à Marseille 11 que les résultats des élèves progressent le plus, et ceci sur chaque item d'an-

Graphique 2 – Taux de réussite aux items d’ancrage de l’ensemble des élèves.

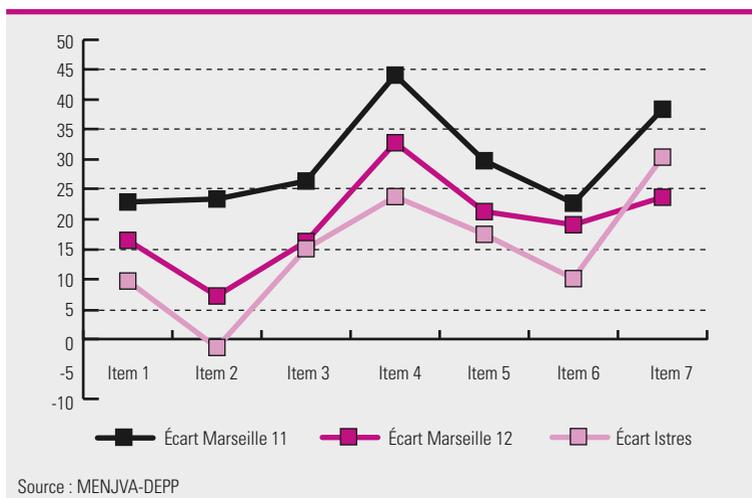


crage, alors qu’au contraire ceux des élèves d’Istres (échantillon témoin) présentent la plus faible progression (mais à partir de scores initiaux plus élevés). Par exemple, sur l’item 4, alors qu’on observe une progression de + 44,1 points pour Marseille 11 ou de + 32,8 points pour Marseille 12, la progression n’est que de + 23,8 points pour Istres. Quant à l’item 2, alors que la progression est égale à + 23,4 points pour Marseille 11, on constate pour Istres un écart de - 1,4 point entre le taux de réussite de la phase 1 et celui de la phase 2.

Analyse des scores de réussite

Lors de la première phase de l’expérimentation, les élèves totalisaient un score moyen de réussite de 27,4 à Istres, 21,4 à Marseille 11 et 23,1 à Marseille 12 sur 43 items. Lors de la deuxième phase, ils avaient en moyenne réussi 30,4 items à Istres, 28,8 items à Marseille 11 et 27,1 items à Marseille 12 sur 45. Les tests étant différents et ne comportant pas le même nombre d’items, nous ne sommes pas en mesure, dans le cadre d’une restitution de premiers résultats, de produire une

Graphique 3 – Évolution des taux de réussite aux items d’ancrage entre la phase 1 et la phase 2 par circonscription.



analyse approfondie de l’évolution des scores entre les deux phases de l’expérimentation. Toutefois, la standardisation des scores de réussite des élèves expérimentateurs sur ceux des témoins, les élèves d’Istres, nous permet de confirmer nos observations sur les taux de réussite aux items d’ancrage entre les trois circonscriptions. Le *graphique 4* montre que les mesures de tendance centrale¹ du score de réussite des élèves de Marseille 11 ont très nettement progressé par rapport à celles d’Istres : lors de la première phase, elles étaient inférieures à celles d’Istres, alors qu’à la deuxième phase, elles sont quasiment au même niveau. Quant à celles de Marseille 12, leur évolution est légèrement supérieure à celle d’Istres.

Pour Marseille 11, l’intervalle interquartile est plus étendu lors de la phase 2 : il augmente de 11 points. De plus, la médiane qui était inférieure à la moyenne lors de la phase 1, lui devient supérieure lors de la phase 2. Ce renversement de tendance indique que la distribution des scores qui était plus allongée vers les grandes valeurs devient plus allongée vers les petites valeurs. Enfin, la distribution des scores est plus étendue lors de la phase 2.

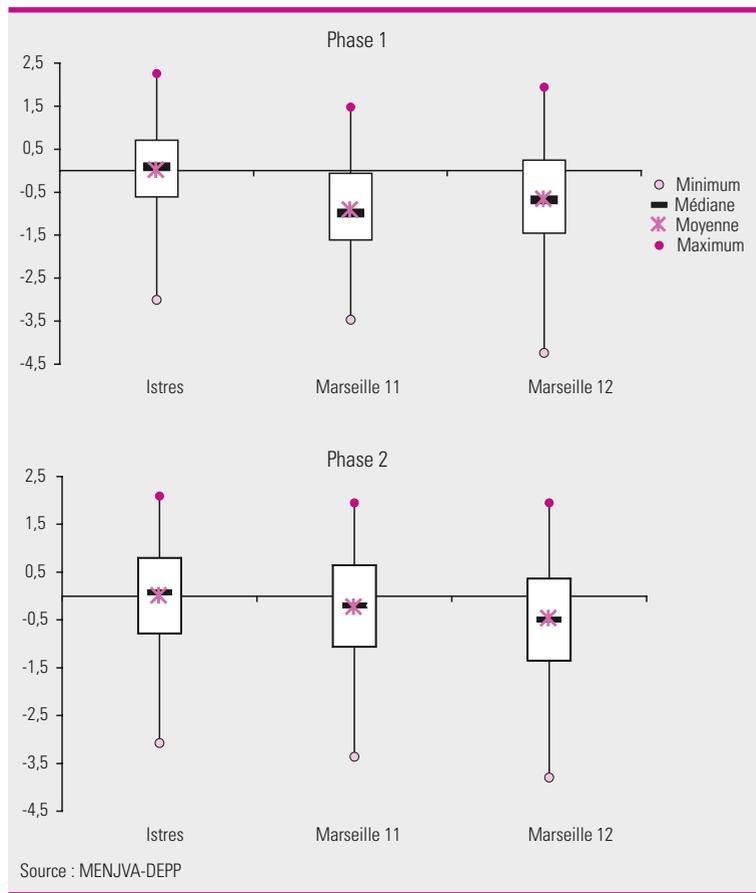
À Marseille 12 en revanche, l’intervalle interquartile reste quasiment le même. La distribution des scores y est symétrique lors des deux phases : la moyenne et la médiane se confondent et l’étendue est stable.

Lors de la première phase, l’écart observé entre le score de réussite de Marseille 11 et celui d’Istres

NOTE

3. Il s’agit ici de la moyenne et de la médiane.

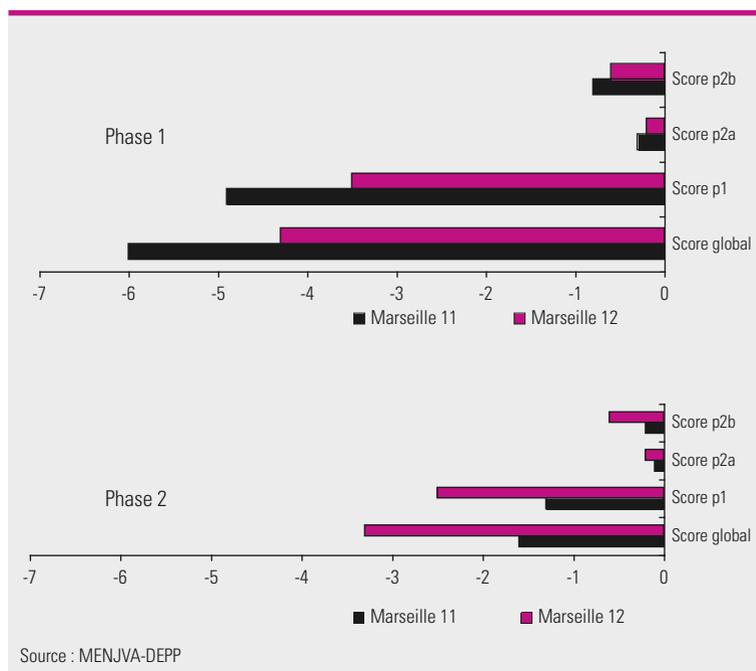
Graphique 4 – Scores de réussite globaux par circonscription



est beaucoup plus grand que celui observé entre Marseille 12 et Istres. Cette tendance s'inverse lors de la deuxième phase de l'expérimentation alors même que l'écart Marseille 12 – Istres s'est lui-même réduit (*graphique 5*). Lors de la première phase de l'expérimentation, Marseille 11 enregistre un score inférieur de - 6 points par rapport à Istres, alors que la différence de Marseille 12 avec Istres est égale à - 4,3 points. Lors de la deuxième phase, Marseille 11 n'a plus qu'un écart de - 1,6 point avec Istres, alors que Marseille 12 en a un de - 3,3 points. C'est principalement sur la première partie du test que les écarts se creusent. Sur la deuxième partie du test, les écarts existent, mais sont beaucoup moins marqués.

Si l'on s'intéresse aux groupes d'élèves définis par les quartiles des scores de réussite standardisés (*graphique 6*), on remarque que les scores des élèves de Marseille 11 augmentent plus que ceux des élèves des deux autres circonscriptions pour tous les groupes, du groupe 1, constitué par les élèves qui ont les acquis les plus fragiles, au groupe 4 des élèves qui obtiennent les scores les plus élevés. De plus, alors que les scores des élèves de Marseille 12 se rapprochent eux aussi de ceux d'Istres pour tous les groupes sauf le groupe 3, on peut observer que, si les scores les plus bas étaient ceux de Marseille 11 à la phase 1, ce sont ceux de Marseille 12 qui le sont à la phase 2. On remarque enfin que le bénéfique s'effectue surtout en faveur des élèves dont les résultats sont bas, en particulier à Marseille 11. En revanche, pour les élèves ayant les meilleurs scores à la phase 1 de l'expérimentation, on relève certes une progression, mais moins importante.

Graphique 5 – Écarts entre les scores des élèves expérimentateurs (Marseille 11 et 12) et ceux des élèves témoins (Istres)



Le graphique 7 confirme ces analyses : les élèves dont les résultats étaient les plus bas (groupe 1) améliorent très nettement leur score à la phase 2 contrairement à ceux dont les résultats étaient les plus élevés (groupe 4). Les élèves ayant un score moyen (groupe 2 et 3) suivent la même tendance que les élèves ayant un score élevé, mais de façon moins marquée, excepté à Marseille 11 où ceux du groupe 2 progressent quasiment autant que ceux du groupe 1.

CONCLUSION

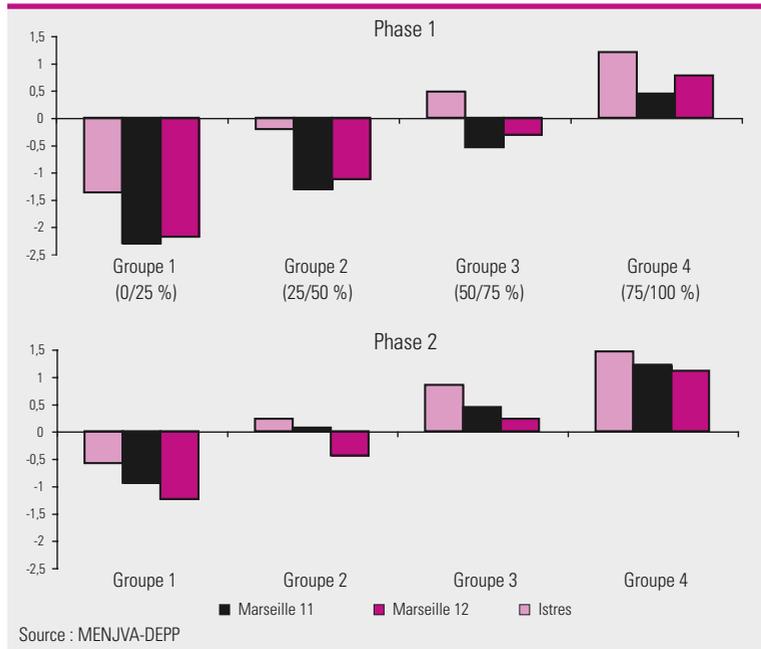
Nous sommes conscients qu'un tel projet, par son approche pluridimensionnelle et les nombreuses variables auxquelles nous avons été confrontés, nécessite une certaine prudence dans la méthodologie à adopter et la présentation de ses résultats.

En effet, la démarche adoptée qui consiste à lier l'évaluation des acquis des élèves aux pratiques des enseignants et à leur formation, tend à en complexifier l'analyse et à en inscrire le traitement dans une temporalité longue. Il est également nécessaire d'avoir à l'esprit que, dans toute étude de ce type, l'effet établissement, classe ou enseignant, sur les résultats des élèves – tout en restant difficile à mesurer – est non négligeable.

Par ailleurs, l'influence d'un tel dispositif dans le temps, sur les apprentissages et sur les pratiques n'est pas sans poser question [17]. Faut-il, du point de vue enseignant, se placer du côté des pratiques efficaces ou des « bonnes pratiques » ? Un tel dispositif est-il efficient à moyen terme ?

L'analyse des taux de réponse aux items d'ancrage et celle des scores

Graphique 6 – Scores moyens de réussite en fonction des quartiles de score de la phase 1, du groupe 1 (élèves aux scores les plus bas) au groupe 4 (élèves aux scores les plus élevés)

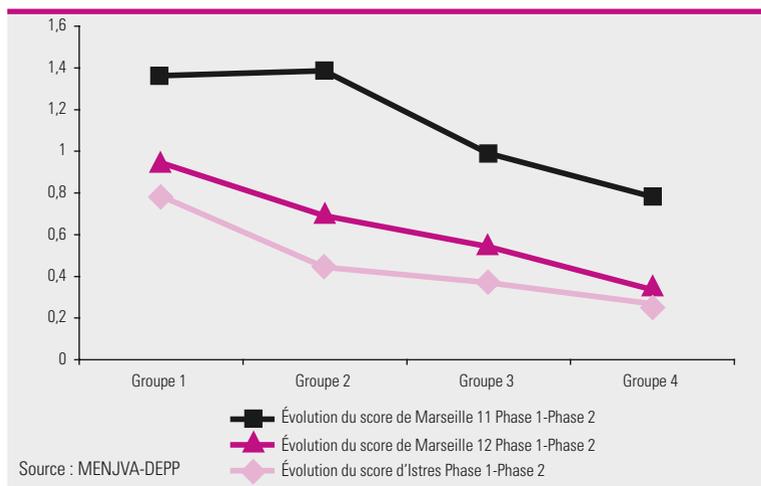


de réussite convergent toutes deux vers les mêmes conclusions. Les résultats des élèves de Marseille 11 et de Marseille 12 ont davantage progressé entre les deux tests que ceux d'Istres, Marseille 11 étant la circonscription où les scores ont le plus augmenté. Notons qu'il s'agit de la circonscription où la part d'élèves relevant de l'éducation prioritaire est la plus importante, avec des élèves

scolarisés en retard plus nombreux et partant dans l'ensemble d'un niveau plus bas.

Sur la base de ces premiers résultats, il semble donc que le projet PACEM ait produit un impact positif sur les acquis des élèves de CM1 ayant bénéficié du dispositif. La progression enregistrée concerne surtout les connaissances sur les unités de mesure des élèves les plus

Graphique 7 – Évolution des scores moyens de réussite entre les deux phases en fonction des quartiles de score de la phase 1



en difficulté, alors que la réussite des élèves dans la résolution de problèmes, peu élevée au départ, a moins progressé. Même si cette différenciation des progrès consta-

tée chez les élèves ne reflète que partiellement des changements de pratiques des enseignants, cela tendrait à montrer qu'un investissement relativement faible mais spécifique

de formation continue et d'accompagnement des enseignants pourrait conduire à une amélioration sensible des performances scolaires. ■

BIBLIOGRAPHIE

- [1] **A. Brun et J.-M. Pastor** (2010), Les compétences en mathématiques des élèves en fin d'école primaire. *Note d'information* n° 10.17, http://media.education.gouv.fr/file/2010/23/5/NIMEN1017_158235.pdf
- [2] **IGEN** (2010), *La mise en œuvre des programmes dans les classes et les évaluations nationales des élèves comme outil de pilotage du premier degré*, Note n° 2010-092, http://media.education.gouv.fr/file/2010/93/9/notereformeprimaire_154939.pdf
- [3] **Programmes de l'école et du collège** (2008) <http://eduscol.education.fr/pid23391/programmes-de-l-ecole-et-du-college.html>
- [4] **Hyerl J.-H.**, (2011) *If you build it will they come? Teacher use of student performance data on a web-based tool*. Working Paper 17486, <http://www.nber.org/papers/w17486>
- [5] **Le nombre au cycle 2** (2011), Ressources pour faire la classe, SCEREN http://media.eduscol.education.fr/file/ecole/00/3/Le_nombre_au_cycle_2_153003.pdf
- [6] **Robert A. et Rogalski J.** (2005) *A cross-analysis of the mathematics teacher's activity. An example in a French 10th-grade class*. *Educational Studies in Mathematics* volume 59, 1 pp 269-298.
- [7] **Vandebrouck F.**, (2007) *La classe de mathématiques : activités des élèves et pratiques des enseignants*. Octares éditions, Toulouse
- [8] **Chesnè J.-F., Pariès M., Robert A.** (2009) *Partir des pratiques en formation professionnelle des enseignants de mathématiques des lycées et collèges*. *Petit x* n° 80, pp 25-46, Grenoble
- [9] **IGEN** (2006) *L'enseignement des mathématiques au cycle 3 de l'école primaire* ; Rapport n° 2006-034 ; <http://media.education.gouv.fr/file/46/0/3460.pdf>
- [10] **Douady R. et Perrin-Glorian M.-J.** (1983), *Mesure des longueurs et des aires*, brochure n° 48, I.R.E.M. université Paris VII, Paris,
- [11] **Butlen D.** (2007) *Le calcul mental entre sens et technique*. Presses universitaires de Franche-Comté, Besançon
- [12] **Peltier M.-L. et al**, (2004) *Dur, dur d'enseigner en ZEP – Analyse de pratiques des professeurs des écoles*. La Pensée Sauvage, Grenoble
- [13] **Van Zanten A.** (2001) *L'école de la périphérie – Scolarité et ségrégation en banlieue*. PUF, Paris
- [14] **Gauthier C. et Dembélé M.** (2005) *Qualité de l'enseignement et qualité de l'éducation : revue des résultats de recherche*. UNESCO Report, 2005/ED/EFA/MRT/PI/18
- [15] **Nye B., Konstantopoulos S., Hedges L.**, (2004) *How Large Are Teacher Effects?* *Educational Evaluation and Policy Analysis*, Vol. 26, n° 3, pp. 237-257
- [16] **Laveault D. et Grégoire J.** (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. De Boeck université, Paris.
- [17] **Kostantopoulos S.** (2007) *How long do teacher effects persist?* Institute for the study of labor (IZA), n° 2 893, Mahwah, Erlbaum, New Jersey

L'accompagnement des collégiens dans leur recherche de stage : une évaluation expérimentale¹

Yannick L'Horty

Emmanuel Duguet

université Paris-Est, ERUDITE et TEPP-CNRS

Pascale Petit

université d'Evry-Val d'Essonne, EPEE, et TEPP-CNRS

Les jeunes en classe de troisième semblent éprouver davantage de difficulté pour effectuer leur stage obligatoire d'une semaine dans les collèges de l'Éducation nationale qui relèvent de l'éducation prioritaire. Alors que la « séquence d'observation en entreprise » est un élément central du dispositif d'orientation et de découverte du monde professionnel, cette première rencontre des jeunes avec l'entreprise peut aussi nourrir un sentiment d'échec et d'exclusion. Nous évaluons, par une méthode aléatoire, un dispositif d'accompagnement à la recherche de stage pour des collégiens qui résident dans des quartiers socialement défavorisés. Le dispositif d'accompagnement n'a d'effet ni sur l'effectivité du stage, ni sur la qualité du stage, évaluée au travers de la satisfaction du jeune, alors qu'il s'agit des objectifs visés par l'accompagnement. En revanche, il influence les choix d'orientation des élèves, ce qui correspond à l'objectif du stage lui-même. Les élèves accompagnés refusent moins fréquemment l'orientation vers des filières courtes et professionnalisées.

Depuis la rentrée 2005, l'ensemble des élèves de troisième suivent un stage d'une semaine en entreprise appelé « séquence d'observation en milieu professionnel ». Pour le ministère en charge de l'Éducation nationale qui l'organise, l'objectif est de sensibiliser les élèves « à l'environnement technologique, économique et professionnel en liaison avec les programmes d'enseignement, notamment dans le cadre de l'éducation à l'orientation » (circulaire n° 2008-092 du 11 juillet 2008). Ce stage constitue la première rencontre entre les jeunes et l'entreprise, qui coïncide avec la première formulation d'un choix d'orientation par l'élève en fin de troisième. Si les modalités peuvent différer selon les établissements, le stage s'étend généralement sur 5 jours consécutifs, en dehors des vacances scolaires, et a lieu dans une entreprise privée ou un établissement public (les élèves âgés de moins de 14 ans ne peuvent effectuer leur séquence d'observation qu'au sein d'un établissement public²).

Ce stage obligatoire pose un problème particulier dans les collèges qui relèvent de l'éducation prioritaire.

Après avoir été créées en 1981, les zones d'éducation prioritaire (ZEP) ont été structurées en réseaux d'éducation prioritaire (REP) en 1999, et ces derniers ont, eux-mêmes, été restructurés en réseau ambition réussite (RAR) de 2006 jusqu'au moment de cette expérimentation. En 2010, on dénombrait 254 collèges classés réseau ambition réussite parmi l'ensemble des 7 000 collèges français (soit 3,6 % des établissements). Pour les collégiens en RAR, le stage obligatoire peut produire des effets qui vont à l'encontre de ceux poursuivis par le ministère en charge de l'Éducation nationale. Ces jeunes

NOTES

1. Ce travail a bénéficié du soutien du Fonds d'expérimentation pour la jeunesse, dans le cadre de son appel à projet n° 1 (Axe 1, prog 6 : améliorer les dispositifs d'orientation et lutter contre l'échec scolaire). Il a également bénéficié de remarques des participants à l'école thématique du CNRS « Évaluation des politiques publiques » (Aussois, mars 2011).

2. Les dispositions de l'article L.4153-1 du code du travail ne permettent pas à un élève de moins de 14 ans d'effectuer la séquence d'observation dans les établissements de droit privé. Seuls les employeurs tels que les administrations, les établissements publics administratifs et les collectivités territoriales peuvent accueillir les élèves sans restriction d'âge.

peuvent éprouver plus de difficultés à trouver un stage que les autres jeunes, avantagés par leur réseau social. Certains peuvent échouer dans leur recherche de stage, beaucoup peuvent trouver des stages de mauvaise qualité qui ne vont pas leur ouvrir de nouvelles perspectives. Certains établissements peuvent même renoncer à mettre en place la séquence d'observation. Leur sentiment d'échec et d'exclusion peut s'en trouver renforcé, ce qui va à l'encontre de l'objectif poursuivi par la politique publique.

Ces constats rejoignent les travaux de la sociologie économique qui concernent l'influence des réseaux sociaux sur la recherche d'emploi [4, 5]. Pour les jeunes qui résident dans des quartiers prioritaires de la politique de la ville, les réseaux sociaux exerceraient donc très tôt un effet défavorable dans la trajectoire professionnelle. Dès la classe de troisième, la première rencontre des jeunes avec l'entreprise, organisée dans tous les collèges par l'Éducation nationale, produirait des effets contre-productifs pour les jeunes situés dans les quartiers les plus défavorisés. Une politique nationale conçue pour faciliter l'orientation des jeunes et leur insertion professionnelle contribuerait aussi à renforcer la ségrégation sociale et spatiale.

C'est la raison pour laquelle des rectorats d'académie ont décidé d'expérimenter un dispositif d'accompagnement spécifique pour la recherche de stage des jeunes collégiens qui résident dans les quartiers socialement défavorisés. L'accompagnement prend la forme de séances collectives avec les collégiens, animées par des professionnels venus du monde de l'entreprise

dans le but d'aider à la recherche de stage. L'expérimentateur est une association qui intervient pour le compte du conseil général de l'Essonne, de la préfecture des Yvelines, de l'inspection d'académie de l'Essonne et de celle des Yvelines.

Dans cette étude, nous proposons d'évaluer les effets de ce dispositif d'accompagnement sur trois variables d'intérêt : le taux d'accès, la qualité du stage et les choix d'orientation. Nous procédons à une évaluation aléatoire avec un protocole qui nous permet de neutraliser les effets du lieu de résidence, des caractéristiques des établissements scolaires, et des caractéristiques des élèves et de leurs familles, pour mettre en évidence l'effet spécifique au dispositif d'accompagnement.

LE CONTENU DE L'EXPÉRIMENTATION ET DE L'ÉVALUATION

L'accompagnement dont bénéficient les élèves de troisième des établissements classés en éducation prioritaire consiste en quatre demi-journées passées avec un *coach* spécialisé, issu du monde de l'entreprise. Les trois premières séances se déroulent avant le stage et visent :

- à faire émerger la représentation d'un stage idéal, et à créer l'envie de réaliser ce stage ;
- à définir les moyens de trouver ce stage idéal ;
- à valider que chaque jeune a obtenu son stage idéal ou le meilleur stage possible.

Pendant toutes ces séances, le collégien tient à jour un dossier de recherche de stage, et est immergé dans un groupe de recherche (avec deux groupes par classe). Il s'agit à la

fois de soutenir la motivation du jeune à mener une recherche active, et de lui apporter une aide effective qui prend la forme de mise à disposition d'adresses d'entreprises dans des secteurs d'activité ciblés par le jeune. La quatrième et dernière séance a lieu à l'issue du stage. Elle a pour thème l'élaboration du rapport de stage et fournit l'occasion d'un nouvel échange où les jeunes se projettent dans l'avenir.

L'objectif de l'évaluation est de mesurer l'effet causal de ce dispositif d'accompagnement sur trois ensembles de variables d'intérêt : le taux d'accès au stage, la qualité du stage et l'évolution des choix d'orientation. Les deux premières variables correspondent à l'objectif assigné au dispositif d'accompagnement. La troisième correspond aux objectifs assignés au stage lui-même.

La difficulté de l'évaluation est de pouvoir mesurer l'effet spécifique de l'accompagnement, indépendamment de celui du stage, alors qu'il a lieu dans des collèges particuliers et pour des élèves particuliers, dont les caractéristiques peuvent influencer à la fois l'effet du stage et celui de l'accompagnement. De plus, on ne peut s'appuyer sur aucune base de données dans la mesure où aucune source statistique de données ne couvre le suivi des stages des jeunes de troisième scolarisés dans des quartiers prioritaires. Comment, sans données, mesurer les effets de l'accompagnement en contrôlant à la fois l'effet établissement et celui des caractéristiques des élèves (observables et non observables) ?

Nous avons retenu une méthodologie d'évaluation aléatoire parce qu'elle peut s'appliquer à de petits échantillons et parce qu'elle permet effectivement de mesurer l'effet causal d'un programme social, toutes

choses égales par ailleurs [6, 3, 1 et 7]. Cette méthodologie est encore très peu utilisée en France, alors qu'elle l'est dans de nombreux pays anglo-saxons et dans de nombreux pays émergents ou en développement. Notons qu'elle a été mise en œuvre avec succès sur des dispositifs d'accompagnement de demandeurs d'emploi dans une étude de Behaghel, Crépon et Gurgand [2].

Afin de neutraliser les effets d'établissement, nous tirons au sort des classes dans des collèges. Par exemple, si un collège comprend 6 classes de troisième, on procède à

un tirage aléatoire simple avec une probabilité de $\frac{1}{2}$, pour décider des classes qui seront effectivement accompagnées et de celles qui ne le seront pas (3 seront accompagnées, 3 ne le seront pas et constitueront nos classes témoins). Pour que le tirage au sort des classes soit accepté par les enseignants et les principaux des collèges, toutes les classes sont effectivement accompagnées, mais sur deux années, la moitié des classes la première année, l'autre moitié la deuxième année. La collecte des données s'est donc étendue sur deux années scolaires : 2009-2010 et 2010-2011.

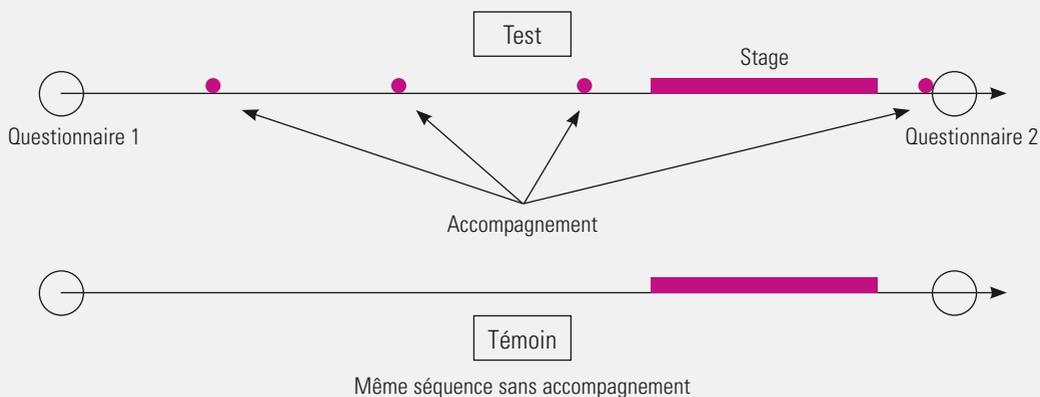
QUALITÉ DES TÉMOINS

La première étape du travail statistique consiste à vérifier si le tirage au sort des classes a permis de produire un échantillon de bonne qualité. On souhaite surtout vérifier que les élèves des classes-témoins ont bien les mêmes caractéristiques moyennes que ceux des classes-tests. Sur les petits échantillons sur lesquels cette étude a été réalisée, il y a un risque non négligeable d'avoir tiré un « échantillon aberrant ». Pour le vérifier, nous comparons les réponses des deux groupes

Méthodologie

Le schéma ci-dessous présente le protocole retenu pour l'évaluation. Nous avons réalisé deux questionnaires qui ont été administrés à l'ensemble des collégiens des classes tests et témoins avant et après l'expérimentation. Le premier questionnaire permet de collecter des informations sociodémographiques sur les élèves et sur leur situation familiale, sur leurs résultats scolaires, sur leurs choix d'orientation, sur leurs attentes relativement au stage et sur leurs représentations du monde de l'entreprise. Le deuxième questionnaire complète ces informations par des variables sur le déroulement du stage, sur le ressenti de l'élève relativement à son stage et, à nouveau, sur ses choix d'orientation.

Le protocole de l'évaluation



L'expérimentation a eu lieu sur deux années consécutives, de la rentrée 2009 à la fin de l'année scolaire 2010-2011. Elle a porté sur 10 collèges, 44 classes et 970 élèves. Néanmoins, les deux questionnaires n'ont été remplis de façon satisfaisante que dans 6 collèges. Dans certains établissements, les questionnaires n'ont pas été distribués dans les classes-témoins. Dans d'autres, un questionnaire sur les deux était manquant ou les questionnaires n'ont pas été renvoyés à l'équipe d'évaluation. Les questionnaires ont été donnés aux principaux des collèges qui les ont transmis aux enseignants puis les ont collectés avant de les retransmettre, une fois remplis, à l'équipe d'évaluation. Le dépôt et la collecte des questionnaires ont été effectués avec plusieurs mois d'intervalles et leur passation dans les classes ont nécessité plusieurs intermédiaires. Cette complexité a été à l'origine de dysfonctionnements dans la chaîne de collecte qui s'est traduite par une diminution sensible du nombre d'observations. Au final, l'échantillon des données effectivement exploitables porte sur 6 collèges classés RAR, soit 28 classes et 550 élèves. Parmi ces élèves, 271 sont dans le groupe-test, 279 sont dans le groupe-témoin. Les collèges de l'échantillon final sont situés à Carrières-sous-Poissy (78), aux Mureaux (78), à Évry (91) et à Épinay-sous-Sénart (91).

au premier questionnaire³ (Q1 : avant Expérimentation). Les tests de significativité ont été réalisés en recalculant les écart-types de chaque variable par la méthode du *bootstrap* avec 100 000 tirages.

Les collégiens de notre échantillon ont des caractéristiques très spécifiques. Leurs parents sont fréquemment bi-actifs et ils sont issus de fratries nombreuses. En moyenne, chaque élève a trois frères et/ou sœurs. On dénombre 15 % d'enfants immigrés et les deux tiers ont des parents qui ne sont pas nés en France. Mais sur tous ces aspects, il n'y a pas de différence significative entre le groupe-test et le groupe-témoin.

Il y a cependant plusieurs variables pour lesquelles nous trouvons des différences significatives entre les deux groupes. Les collégiens du groupe-test déclarent moins fréquemment que leur père travaille et, lorsqu'il travaille, ils déclarent moins fréquemment sa profession. En outre, leur moyenne générale est un peu plus faible (de 0,65 point sur 20) et ils considèrent moins souvent que la meilleure filière pour réussir est la filière générale. Par ailleurs, ils portent un regard moins positif sur leur collègue : moins d'élèves du groupe-test considèrent que la discipline est respectée dans leur établissement et qu'il y a des bonnes relations entre élèves.

Les attentes vis-à-vis du stage et les attentes en matière d'orientation font l'objet d'un grand nombre d'items en fin de questionnaire. Près de 75 %

des collégiens déclarent qu'ils ont besoin de renseignements sur leurs prochaines années de scolarité. La proportion de ceux qui déclarent avoir une idée du métier qu'ils veulent faire plus tard est identique. Notons également que plus de la moitié des collégiens ont un regard positif sur l'entreprise, et plus de 40 % ont un regard indifférent. Il y a également sur ce point une différence significative entre tests et témoins. L'opinion des élèves du groupe-test sur l'entreprise est moins fréquemment positive. Parmi près de 20 items décrivant ce qu'ils attendent du stage, le plus espéré est « une bonne ambiance de travail », suivi « d'un travail intéressant », puis « le respect des autres membres de l'entreprise » et « l'opportunité d'apprendre de nouvelles choses », « être traité de manière juste ». Le moins souhaité est « d'être dans un milieu d'adultes », « que les salariés me consacrent du temps », et « découvrir un autre univers ». Sur tous ces items, il n'y a pas de différences significatives (au seuil de 5 %) entre les élèves du groupe-test et ceux du groupe-témoin.

Parmi une autre vingtaine d'items indiquant ce que les jeunes peuvent apporter à leur employeur, le plus cité est « respecter mes supérieurs », puis « respecter mes collègues ». Le moins cité est « aller chercher un café à mon tuteur s'il me le demande », ou « prendre des notes sur ce que je vois ». Sur quelques-uns de ces items, il y a une différence significative entre le groupe-test et le groupe-témoin. Les élèves du groupe-test déclarent plus fréquemment qu'ils sont prêts à « adapter leur langage » pour se conformer aux attentes de leur futur employeur. Ils indiquent aussi plus

fréquemment qu'ils sont prêts à aller chercher un café à leur tuteur s'il leur demande.

Au total, les différences entre le groupe-témoin et le groupe-test paraissent non négligeables. Elles sont suffisamment importantes pour suggérer qu'une simple comparaison de moyenne entre les deux groupes ne serait pas pleinement convaincante. Le risque est de voir les éventuelles différences, qui seront relevées dans l'exploitation du deuxième questionnaire, après l'expérimentation, plutôt provoquées par des biais de composition que liées à l'effet causal de l'accompagnement.

Même si les deux groupes ont été constitués sur la base d'un tirage au sort des classes, il n'y a pas suffisamment d'observations pour que les deux groupes aient statistiquement la même composition. Pour remédier à ce risque de biais de sélection, nous avons reconstruit un groupe témoin avec la méthode du score canonique (ou « *propensity score* », en anglais) proposée par Rosenbaum et Rubin [8, 9 et 10].

LES EFFETS DE L'ACCOMPAGNEMENT SUR L'ACCÈS ET LA QUALITÉ DU STAGE

Nous exploitons maintenant le deuxième questionnaire qui est administré à l'issue de l'accompagnement. Pour chacune des questions, les différences significatives entre le groupe-test et le groupe-témoin sont recherchées. Mais, compte tenu des différences entre les deux groupes avant l'expérimentation, les résultats de l'exploitation du deuxième questionnaire ne

NOTE

3. Les résultats détaillés figurent dans le tableau 1 de notre rapport de recherche, disponible sur le site www.tepp.eu/.

peuvent être attribués *ex abrupto* à l'effet causal de l'accompagnement. Des traitements complémentaires ont donc été réalisés pour estimer l'effet causal de l'accompagnement à l'aide de la méthode de Rubin. Pour présenter nos résultats, nous explorons successivement les trois aspects qui nous intéressent : l'effet de l'accompagnement sur le taux d'accès en stage, l'effet sur la qualité du stage, et l'effet sur les choix d'orientation.

Concernant le taux d'accès au stage, on trouve une différence dans les statistiques brutes entre le groupe-test et le groupe-témoin. On dénombre 92,5 % de stagiaires au sein du groupe-test contre 97,8 % au sein du groupe-témoin. Si l'on ne tenait pas compte du biais de sélection, on en déduirait que l'accompagnement exerce un effet négatif sur les chances de trouver un stage. En réalité, il n'en est rien. On ne trouve plus de différence dans l'accès au stage une fois que sont considérées les différences de composition entre les deux groupes avec la méthode du score de propension. On ne trouve pas non plus de différence dans les canaux grâce auxquels le stage a été trouvé, même si les statistiques brutes faisaient apparaître que les accompagnés déclaraient moins fréquemment avoir trouvé leur stage grâce à leur famille que les collégiens non accompagnés. L'accompagnement n'aurait donc eu aucun effet significatif sur la capacité des jeunes à trouver un stage et les modalités par lesquelles ce stage a été trouvé.

Pour la suite des traitements, on ne s'intéresse qu'au sous-échantillon des collégiens qui ont effectué le stage. Concernant la qualité

de ce stage, on est confronté à un problème de définition et de mesure. Qu'est-ce qu'un stage de bonne qualité ? Pour répondre à cette question, nous nous sommes placés du point de vue du ressenti du collégien. Nous avons donc choisi d'interroger les élèves sur différentes dimensions de la qualité du stage, qui vont de la satisfaction globale à la découverte d'un métier, en passant par l'ambiance de travail. Il était demandé aux collégiens d'indiquer dans un tableau la réponse qui correspondait le mieux à ce qu'ils pensaient, en se positionnant sur une échelle de Likert à cinq choix (de « pas du tout d'accord » à « tout à fait d'accord »). Les modalités retenues pour apprécier la qualité du stage sont présentées dans le *tableau 1*.

Sur la plupart de ces items, nous n'avons pas relevé de différences significatives entre les collégiens des classes accompagnées et ceux des classes qui ne l'ont pas été. Sur certains items, nous avons relevé des différences significatives dans les moyennes entre les deux groupes (les collégiens accompagnés déclarent

plus fréquemment avoir une vision plus claire de ce qu'ils veulent faire, avoir découvert le métier qu'ils veulent faire plus tard, et avoir découvert un autre univers). Mais ces différences ne sont plus significatives une fois pris en compte les biais de composition avec la méthode de Rubin. Les seules modalités pour lesquelles des différences existent (au seuil de 10 %) sont celles qui ont trait à la qualité de la relation établie avec le tuteur (« J'ai été soutenu par mon tuteur », « J'ai pu communiquer de façon ouverte avec mon tuteur ») qui est meilleure pour les élèves accompagnés, et celles qui ont trait au choix d'orientation : les élèves accompagnés déclarent plus fréquemment avoir conforté, grâce au stage, leur choix d'orientation. La différence est de plus de dix points de pourcentage dans les statistiques brutes (48,8 % contre 38,3 % pour le groupe-témoin) et reste significative après contrôle. Nous retenons que l'accompagnement ne produit globalement pas d'effet sur la qualité du stage telle qu'elle est appréciée par le collégien, mais qu'il améliore la relation avec

Tableau 1 – Les modalités retenues pour apprécier la qualité du stage

Globalement, je suis satisfait par ce stage
Le stage m'a apporté des informations sur un métier
Le travail était intéressant
J'ai été soutenu par mon tuteur
J'ai pu communiquer de façon ouverte avec mon tuteur
Il y avait une bonne ambiance de travail
Les autres membres de l'entreprise ont fait preuve de respect vis-à-vis de moi
J'ai eu l'opportunité d'apprendre de nouvelles choses
J'ai maintenant une vision plus claire de ce que je veux faire
J'ai conforté mon choix d'orientation après la troisième
J'ai participé aux activités de l'entreprise
J'ai découvert le métier que je veux faire plus tard
Je me sens plus adulte
Je sais ce que c'est que d'aller au travail
J'ai fait des choses plus concrètes qu'à l'école
J'ai eu une expérience du monde du travail
J'ai découvert un autre univers
J'ai vécu dans un milieu d'adultes

le tuteur et qu'il conduit à ce que le stage conforte plus fréquemment les choix d'orientation.

LES EFFETS SUR LES CHOIX D'ORIENTATION

L'accompagnement semble effectivement influencer assez nettement les choix d'orientation des collégiens. Tout d'abord, il modifie l'attitude des jeunes relativement aux filières professionnelles. À la question, « quelle est la meilleure filière pour réussir ? », une majorité des collégiens accompagnés indiquent une filière professionnelle (40,9 % vs 38,5 %), alors qu'une majorité des collégiens qui n'ont pas bénéficié de l'accompagnement indiquent une filière générale (46,5 % vs 35,7 %). Les écarts-types ont été estimés par *bootstrap* sur 100 000 tirages, et la préférence des collégiens non accompagnés pour la filière générale comme étant meilleure filière pour réussir, relativement aux collégiens accompagnés, est effectivement significative⁴. Toutefois, ces différences brutes ne sont plus significatives une fois pris en compte les biais de sélection avec la méthode de Rubin.

Leur attitude vis-à-vis du CAP a, elle aussi, été modifiée par l'accompagnement. Ils refusent aussi fréquemment la perspective d'effectuer un CAP. À la question, « envisages-tu de faire un CAP ? », 75,7 % des accompagnés comme des non accompagnés répondent non. Mais lorsqu'on

leur demande pour quelles raisons ils envisagent un CAP, les jeunes accompagnés répondent majoritairement « Pour gagner de l'argent plus vite et être autonome », alors que ce n'est le cas que d'une minorité des non accompagnés qui répondent majoritairement « Pour faire le métier dont j'ai envie ». La différence est significative entre les deux groupes et elle se maintient avec la méthode du score de propension.

CONCLUSION

Dans cette étude, nous avons présenté les résultats d'évaluation d'une expérimentation sociale consistant à mettre en œuvre un dispositif d'accompagnement pour aider les collégiens à la fois scolarisés en éducation prioritaire et résidant dans des zones prioritaires de la politique de la ville, dans la recherche de leur stage obligatoire de troisième. L'expérimentation a porté sur 10 collèges et 44 classes. Des dysfonctionnements dans la chaîne de collecte ont conduit à une diminution sensible du nombre d'observations exploitables qui porte finalement sur 6 collèges classés RAR, soit 28 classes et 550 élèves. Parmi ces élèves, 271 sont dans le groupe-test, 279 sont dans le groupe-témoin. Les collèges sont situés à Carrières-sous-Poissy (78), aux Mureaux (78), à Évry (91) et à Épinay-sous-Sénart (91). Même si les deux groupes ont été constitués sur la base d'un tirage au sort des classes, il n'y a pas suffisam-

ment d'observations pour qu'ils aient statistiquement la même composition. Pour remédier au risque de biais de sélection, nous avons reconstruit un groupe témoin avec la méthode du score de propension proposée par Rosenbaum et Rubin [10].

Nous montrons que ce dispositif n'a d'effet ni sur l'accès au stage, ni sur la qualité du stage, évaluée au travers de la satisfaction du stagiaire (qui sont les objectifs visés par le dispositif complémentaire). En revanche, il joue un rôle sur les choix d'orientation des élèves (qui est l'objectif du dispositif principal, le stage obligatoire). Les élèves accompagnés, issus de quartiers ségrégués, refusent moins fréquemment l'orientation vers des filières courtes et professionnalisées et sont plus fréquemment confortés dans leur choix d'orientation à l'issue du stage.

Cette étude illustre l'intérêt d'une évaluation d'impact appliquée à des microprogrammes sociaux locaux. Alors que les méthodes d'évaluation non expérimentales ne pourraient pas mesurer les effets *ceteris paribus* de ce type de programme, une évaluation expérimentale permet d'identifier cet impact y compris sur la base d'un échantillon restreint. ■

NOTE

4. Les résultats détaillés figurent dans le tableau 2 de notre rapport de recherche, disponible sur le site www.tepp.eu/.

BIBLIOGRAPHIE

- [1] **Banerjee A. et E. Duflo** (2008). *The Experimental Approach to Development Economics*. Mimeo MIT J PAL.
- [2] **Behaghel L., Crépon B. et Gurgand M.** (2009). *Évaluation d'impact de l'accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise*, rapport final. Miméo, septembre.
- [3] **Burtless G.**, (1995). The Case for Randomized Field Trials in Economic and Policy Research. *Journal of Economic Perspectives*, Vol 9, n°2, pp 63-84.
- [4] **Fontaine F.** (2006). Les réseaux de relations : quelles perspectives pour l'économie du marché du travail ?. *Revue française d'économie*, volume 21, n° 21-1, pp. 127-172.
- [5] **Granovetter M.** (1995). *Getting a job: A study of contacts and careers*, The University of Chicago Press,
- [6] **Heckman J.** (1992). Randomization and social policy evaluation, in *Evaluating Welfare and Training Programs*. Editors Charles Manski and I. Garfinkel. Cambridge, MA: Harvard University Press. (also available as NBER Technical Working Paper No.107, 1991).
- [7] **L'Horty Y., et Petit P.** (2011). Évaluation aléatoire et expérimentations sociales. *Revue française d'économie*, 26 (1), p. 13-48.
- [8] **Rosenbaum P. R. and Rubin D. B.** (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.
- [9] **Rosenbaum P. R. and Rubin D. B.**, (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70, 41-55.
- [10] **Rubin D. B.** (2006). *Matched Sampling for Causal Effects*. New York: Cambridge University Press.

Améliorer l'accès des jeunes ayant moins d'opportunités à la mobilité internationale

Le cas appliqué d'une stratégie territoriale

Sophie Carel

directrice du centre régional associé (CAR)
Céreq¹ de Besançon, ThéMA-UMR 6049

Frédéric Déloye

Céreq de Besançon, ThéMA-UMR 6049,
université de Franche-Comté

L'une des priorités de la Commission européenne consiste à assurer à des jeunes, nommés « jeunes ayant moins d'opportunités » (JAMO), un accès au dispositif existant d'aide à la mobilité internationale. Cette terminologie, créée par la Commission européenne² pour désigner les jeunes handicapés ou issus d'un milieu culturel, géographique ou socioéconomique défavorisé, permet de cibler au mieux les jeunes concernés, tout en développant et favorisant leur accès aux dispositifs d'aide à la mobilité et d'éducation non formelle développés dans le cadre des programmes européens dédiés à la jeunesse dont le Programme jeunesse couvrant la période 2000-2006 et son successeur, le Programme européen jeunesse en action (PEJA) établi pour 7 ans de 2007 à 2013.

La mission locale du bassin d'emploi du Grand Besançon a souhaité travailler en respectant la philosophie de la Commission européenne pour permettre à des jeunes, peu voire non mobiles, de pouvoir bénéficier d'expériences de séjours à l'étranger.

Partant du constat que la mobilité internationale des jeunes reste encore aujourd'hui limitée et qu'elle profite essentiellement aux jeunes majoritairement diplômés, sans difficultés supposées de future insertion, notamment professionnelle, la mission locale du bassin d'emploi du Grand Besançon a investi dans une expérimentation sociale en réponse à l'appel à projets lancé par Martin Hirsch en avril 2009. Elle a souhaité se concentrer sur la mobilité des jeunes âgés de 16 à 25 ans peu ou pas qualifiés (niveau IV et *infra*), apprentis ou jeunes accompagnés par la mission locale et habitant la Communauté d'agglomération du Grand Besançon (CAGB), territoire à la fois urbain et rural.

Seront pointés, ci-après, les éléments facilitateurs et les logiques de blocage induits par les expérimentations sociales à destination de la jeunesse,

lancées par le gouvernement en prenant à témoin cette initiative de la mission locale du Grand Besançon, porteur d'un projet d'« aide à la mobilité internationale » (dénommée porteur ou expérimentateur dans la suite du propos). L'évaluation a été menée par le Céreq (dénommé évaluateur), par le biais de son centre associé de Besançon, localisé au sein de l'université de Franche-Comté et intégré au laboratoire ThéMA-UMR 6049 du CNRS.

LES JAMO, CIBLE PRIORITAIRE DES PROGRAMMES EUROPÉENS

Le Programme européen jeunesse en action (PEJA) s'adresse à tous les jeunes âgés de 15 à 28 ans, éventuellement de 13 à 30 ans dans certains cas. Il vise à développer le sens de la citoyenneté active, de la solidarité et de la tolérance des

NOTES

1. Centre d'études et de recherches sur la qualification.

2. Le 15 novembre 2006, le Parlement européen et le Conseil ont adopté la décision n° 1719/2006/CE

jeunes et à leur faire jouer un rôle actif dans la création de l'avenir de l'Union européenne. Il favorise leur mobilité au sein de celle-ci et au-delà de ses frontières, promeut l'apprentissage non formel³ et le dialogue interculturel, et encourage l'inclusion de tous les jeunes, sans tenir compte de leur milieu éducatif, social et culturel. Pour le programme en cours « Jeunesse en action », il est particulièrement important que tous les jeunes, et surtout ceux qualifiés de plus défavorisés, aient accès à ces activités et soient intégrés dans la société en général. En cela, le programme est un outil au service de leur intégration. Pour désigner cet effort d'intégration, on utilise le terme « *inclusion* ». Une stratégie d'inclusion spécifique a donc été développée à cet effet. La stratégie d'inclusion du PEJA poursuit un certain nombre d'objectifs : veiller à ce que les jeunes ayant moins d'opportunités bénéficient d'un accès équitable au programme ; améliorer l'inclusion sociale, la citoyenneté active et l'employabilité des jeunes ayant moins d'opportunités ; et, enfin, contribuer à la cohésion sociale.

Avant d'aborder les bénéfices pour les JAMO d'un accès facilité à la mobilité, il est important de revenir sur la (les) définition(s) donnée(s) à ce type de public.

Qui sont les jeunes avec moins d'opportunités ?

Définir cette catégorie de personnes n'est pas aisée. Premièrement, la définition des JAMO varie en fonction de la profession exercée, des situations auxquelles on est confronté, de la zone géographique où l'on exerce ou encore de son resenti personnel. En effet, chacun voit

les choses différemment selon qu'il soit animateur jeunesse, professionnel de l'insertion ou fonctionnaire de l'État. Deuxièmement, il est encore plus ardu d'élaborer une définition commune à tous les pays concernés par le programme : en Europe, on dénombre 35 agences nationales du PEJA. En France, depuis les incidents de novembre 2005, la tendance est d'assimiler les jeunes ayant moins d'opportunités aux jeunes des « quartiers sensibles ». En Allemagne, on retient pour cette définition les jeunes qui n'ont pas eu accès à la mobilité (études courtes, emplois peu qualifiés, etc.). Le terme de JAMO est donc protéiforme voire polysémique.

La décision du Parlement européen et du Conseil européen du 13 avril 2000 établissant le Programme jeunesse (pour rappel, le prédécesseur du PEJA) définissait les jeunes ayant moins d'opportunités comme les « *jeunes qui, pour des raisons d'ordre culturel, social, physique, mental, économique ou géographique, ont le plus de difficultés à participer aux programmes d'action qui les concernent tant au niveau communautaire qu'aux niveaux national, régional et local* ». Dans cette définition, les jeunes sont caractérisés selon les différents obstacles auxquels ils sont confrontés, au regard des difficultés voire des handicaps identifiés.

La définition la plus utilisée est celle listant les différents obstacles auxquels peuvent être confrontés les jeunes que l'on peut trouver dans le mémo-guide sur le Service volontaire européen (SVE) réalisé par l'Agence française pour le programme européen jeunesse (AFPEJ) : « *La terminologie européenne utilise l'expression "jeunes avec moins*

d'opportunités", c'est-à-dire des jeunes qui, pour une ou plusieurs raisons liées à la précarité de leur situation personnelle, rencontrent des obstacles dans l'accès à la mobilité et donc au programme. Ces obstacles peuvent être :

– *des obstacles socio-économiques liés à la situation sociale et/ou économique rencontrée par le jeune et/ou sa famille ;*

– *des obstacles familiaux passagers ou durables ;*

– *des obstacles géographiques liés au milieu de vie : isolement en milieu rural ou urbain, quartier dit « sensible », éloignement géographique avec la métropole ;*

– *des obstacles dans la scolarité et/ou la formation : situation d'échec, de rupture, de faible niveau ;*

– *des obstacles liés à la culture, à l'origine, ou à la discrimination sous toutes ses formes ;*

– *des obstacles liés à des problèmes de santé et/ou de handicap. »*

La définition des JAMO sur le programme « Jeunesse en action » est sensiblement la même : elle définit les jeunes défavorisés comme des

NOTE

3. Définition et objectifs de l'apprentissage non formel : en raison d'inégalités rencontrées au cours des premiers cycles d'enseignement, les élèves issus de milieux défavorisés n'obtiennent bien souvent pas le niveau de qualification requis pour s'insérer sur le marché du travail ou accéder à l'enseignement supérieur. L'éducation non formelle permet de remédier en partie à cette situation. Destinée aux jeunes ayant moins d'opportunités, elle contribue à rendre le processus éducatif agréable et attrayant. En proposant une source d'apprentissage supplémentaire, elle permet d'offrir une nouvelle voie vers l'éducation formelle ou la formation à des jeunes qui n'avaient pas eu l'occasion d'en bénéficier auparavant.

jeunes « ayant moins d'opportunités » que leurs pairs en raison de facteurs socioéconomiques, de handicaps physiques ou mentaux, de problèmes de santé, de différences culturelles ou encore de situations et d'obstacles éducatifs ou géographiques.

La Commission européenne a adopté en 2007 la « stratégie d'inclusion des jeunes avec moins d'opportunités » retenue par l'agence française en ajoutant des spécificités nationales selon trois axes d'intervention :

- les « groupes cibles » prioritaires au niveau national sont les jeunes des quartiers urbains sensibles, et de certaines zones rurales, et les jeunes handicapés, physiques ou mentaux ;

- la démarche retenue est de faire travailler en réseau au niveau local, régional et national, les différents acteurs de terrain susceptibles d'accompagner ces jeunes et de monter des projets avec des partenaires européens. Le travail de proximité, animé par le correspondant régional PEJA et son réseau, est très important ;

- la participation des JAMO aux activités de mobilité éducative en Europe doit être réellement une étape importante de leur parcours d'insertion. C'est la raison pour laquelle l'agence considère comme stratégique la formation et la mise en réseau des acteurs du programme susceptibles d'intervenir auprès de ces jeunes.

En somme, un jeune ayant moins d'opportunités est un jeune qui – en raison de divers obstacles ou handicaps – n'a pas naturellement accès à la mobilité transnationale ou internationale, qui doit fournir plus d'efforts pour s'informer puis participer à un des dispositifs du programme qui lui serait pourtant particulièrement bénéfique.

Le pari de la mobilité spatiale pour encourager une mobilité sociale

La mobilité européenne revêt de nombreux avantages : une rupture avec le quotidien pouvant agir comme un outil de redynamisation ; une mixité sociale, culturelle et linguistique modifiant profondément la rencontre avec l'autre ; une ouverture sur l'Europe ; l'acquisition de compétences non formelles (indépendance, prise de responsabilité, etc.). Dépassant leurs limites géographiques et culturelles, une (re)mobilisation de ces jeunes est attendue. En changeant de regard sur eux-mêmes, ils acquièrent généralement une meilleure confiance en eux. Enfin, au retour, la mobilité européenne laisse supposer une redynamisation des « jeunes ayant moins d'opportunités » et les incite, au moins pendant un temps, à se projeter un peu plus dans l'avenir.

Au-delà de la définition des « jeunes ayant moins d'opportunités » proposée précédemment, et sans chercher spécifiquement à caractériser les obstacles auxquels ils sont confrontés, nous pouvons nous intéresser aux avantages présentés de la mobilité européenne, aux effets attendus dans un projet construit, dans une démarche volontaire et espérée positive, et aux changements qu'elle pourrait induire chez un jeune mobilisé. Dans cette perspective, la mobilité spatiale est supposée entraîner une mobilité sociale.

Une expérience de mobilité européenne est avant tout, pour un « jeune avec moins d'opportunités », un séjour de rupture. Un service volontaire européen⁴ de plusieurs mois, tout comme un échange de jeunes de courte durée,

permet de rompre avec le quotidien, de quitter sa famille, ses amis, son environnement. Il peut ainsi prendre de la distance par rapport à sa situation personnelle et professionnelle. L'espoir d'un déclic pour le jeune et de la constitution d'un « puissant outil de redynamisation » sont permis. Avec les dispositifs habituels d'insertion, les jeunes en difficulté rencontrent rarement d'autres jeunes, ils restent entre eux, partagent souvent les mêmes problèmes, rencontrent les mêmes obstacles et les mêmes freins. Partir à l'étranger pour quelques jours ou quelques mois peut représenter une importante ouverture, une réelle opportunité, en même temps qu'une expérience enrichissante.

La mobilité volontaire encadrée peut, tout à la fois, être événement, expérience, interaction et projection. Elle est « socialement construite, individuellement vécue, intersubjectivement partagée » [M. Lussault, 2004]. La mobilité a donc, en cela, des effets individuels et sociaux. En outre, la mixité sociale, culturelle et linguistique, modifie totalement l'appréhension même de la question de la rencontre, et donc de l'ouverture. En ce sens, l'immersion dans un pays de langue étrangère, premier obstacle au dialogue, force l'adaptation pour permettre échanges et

NOTE

4. La mobilité européenne dans le cadre du SVE est une forme d'apprentissage non formelle. Elle est propice à l'acquisition de nouveaux savoir-faire : des capacités et des compétences apprises hors du cadre scolaire, et réutilisables dans une démarche d'insertion professionnelle. Se développent aussi de nouvelles aptitudes sociales : devenir plus indépendant, prendre des responsabilités, aider les autres, contribuer à la vie de la communauté locale, etc. On construit ainsi « le socle d'un processus d'accès à l'autonomie ».

compréhension et « impose » le recours à de nouvelles formes d'expression et de communication. Un certain phénomène de « décentration » peut ainsi se produire.

Arrêter, un temps, de voir le reste du monde sous le prisme de son quotidien et des limites qu'il impose parfois, peut ainsi être rendu possible. La rupture que constitue l'expérience même de mobilité laisse supposer une certaine remobilisation. Les jeunes reviennent d'abord avec une meilleure idée de ce qu'ils souhaitent faire, ou tout au moins de ce qu'ils ne veulent pas faire. Leur projet personnel et professionnel étant mieux défini, ils peuvent recalibrer les outils et méthodes à déployer dans le cadre de leur recherche active d'emploi ou de formation. Ce dynamisme retrouvé ne dure parfois peut-être pas, mais « il est indubitablement moteur, au retour ». Enfin, les anciens volontaires reviennent changés : ils ne portent plus le même regard sur leur situation. Ils ont généralement pris confiance en eux et pourront prendre des initiatives, seront prêts pour de nouvelles expériences. À la confiance s'ajoute une meilleure estime de soi, doublée de la fierté d'avoir achevé un projet personnel construit en ayant su gérer craintes, appréhensions, dépaysement et possibles difficultés.

Confrontés à une autre culture, à des pratiques différentes, ils ont dû s'adapter à ce nouvel environnement. Voyager signifie dès lors pour eux, connaître, apprendre à se connaître, comprendre ses capacités et ses limites. Le Service volontaire européen (SVE) permet ainsi de « potentialiser tout le travail effectué par les structures habituelles dans le suivi et l'accompagnement des jeunes et de gagner en rapidité et en efficacité ».

En cela, la mobilité est « un outil pédagogique à part entière et à forte valeur ajoutée ».

INNOVATIONS SOCIALES : UN SYSTEME COMPLEXE

En substance, ce projet d'aide à la mobilité internationale porté par la mission locale du bassin d'emploi du Grand Besançon est novateur. En effet, il souhaite faciliter le départ à l'étranger de jeunes âgés de 16 à 25 ans, apprentis ou suivis par la mission locale, population dont on sait qu'elle a peu de propension à la mobilité. Plusieurs dispositifs coexistent : chantier international, bourses individuelles, échanges culturels et création d'un portail collaboratif. Ce champ de l'existant au niveau des instruments de l'action publique sera, dans cette expérimentation, combiné. Cette combinatoire constitue l'aspect innovant de la démarche mis en œuvre. Par ailleurs, expérimentateur et évaluateur se sont accordés sur une démarche d'ensemble qui vise à réaliser un compromis entre, d'une part, les directives nationales données pour la déclinaison sur le territoire français du programme communautaire, et d'autre part, les réalités du terrain spécifiques au territoire du grand Besançon. Cette recherche de compromis est à l'origine d'un système complexe d'actions.

Dynamique partenariale et logique de co-construction

Dans la candidature à l'appel à projets d'expérimentations sociales portées par le Fonds d'expérimentation pour la jeunesse (FEJ), soulignons la présence du registre collaboratif, souvent présent dans les discours liés

à l'intervention publique volontariste. Dès la conception, la dynamique partenariale est souhaitée, la logique de co-construction affichée. Ainsi, nous remarquons que le porteur envisage que « *la synergie des acteurs et la conjugaison des savoir-faire permettra dans une logique de guichet unique de co-construire et d'organiser avec le jeune son projet de mobilité internationale* » (In Dossier de candidature HCJ⁵ Aide à la mobilité, portée par la mission locale de Besançon et évaluée par le centre associé Céreq de Besançon).

La dynamique partenariale doit se traduire par une combinaison des compétences des différents services, ce qui devrait permettre d'encadrer au mieux les demandes et projets des jeunes grâce à « *la création d'une plate-forme collaborative dédiée à la mobilité des jeunes regroupant les acteurs et les compétences nécessaires à la réussite des projets individuels ou collectifs* ». En concordance avec les orientations du Haut commissaire à la jeunesse, le porteur a insisté sur la mise en réseau nécessaire et la complémentarité entre acteurs.

Le projet a démarré grâce à l'initiative propre du directeur de la structure ; sa personnalité dynamique a énormément joué dans le portage de la réponse à l'appel à projets. Le porteur, de par son expérience d'administrateur territorial, a activé des partenariats qui préexistaient, des réseaux qui lui étaient familiers et opérationnels. À ce titre, il a sollicité des structures et

NOTE

5. À partir de maintenant et jusqu'à la prochaine note, toutes les citations en italique proviendront de cette même source.

des collectivités pour les impliquer dans le projet (centre régional d'information jeunesse (CRIJ), conseil régional, conseil général, ville de Besançon, centre communal d'action social) ; la candidature s'est enrichie sur le mode itératif, ce qui a permis aux partenaires de se saisir du projet en amont ; mais pas forcément de se l'approprier et d'en devenir vraiment acteurs. Dans le cadre de la démarche participative, les différents responsables de service se sont retrouvés pour les réunions (mensuelles) du comité de pilotage.

Cette image reste bien entendu à nuancer, dans la mesure où la traduction du registre collaboratif se heurte à des difficultés multiples. Bien souvent, dans la conception d'un projet, il n'y a pas d'effort de conceptualisation des difficultés, des dérives et écueils potentiels ni de réflexion sur les mesures palliatives à mettre en œuvre le cas échéant.

On retrouve ainsi, à différentes étapes, des difficultés caractéristiques de la gestion de projet, que ce soit des divergences stratégiques entre directeurs de structure, des oppositions techniques sur l'utilisation des ressources technologiques, des revendications quant au *leadership* ou l'animation du réseau, du travail collaboratif, etc.

Force est de constater que les intentions volontaristes, si honorables soient-elles, se heurtent à des difficultés empiriques multiples. Par ailleurs, plusieurs acteurs associatifs ont déploré, malgré leur expérience de terrain dans l'accompagnement de projets des jeunes, de ne pas avoir été sollicités en amont, à l'instar d'autres partenaires institutionnels. Le porteur, dès qu'il a eu connaissance de cette « frustration » *via* l'évaluateur, a invité

ces acteurs « de terrain » à intervenir dans la préparation au départ, et à réaliser des volets de formation concernant, notamment, le choc culturel.

Les intentions collaboratives fonctionnent souvent de façon stratifiée, entre directeurs, entre responsables informatiques, entre services qui se connaissent bien, qui se pratiquent. Les collaborations avec des structures hors premiers cercles partenariaux, plus éloignées de leurs prérogatives et missions, sont parfois beaucoup plus difficiles à instaurer. Seuls le temps, l'information et la communication feront leur œuvre. Les possibles lenteurs et réticences dans la mise en place de collaborations opérationnelles non sectorielles témoignent de la difficulté de traduire un discours volontariste en pratique, en vue de réunir des compétences plurielles et s'assurer ainsi de ne pas réinventer, seul, l'existant.

Des bénéficiaires éloignés des offres de mobilité

Malgré les politiques incitatives de l'Union européenne, la popularité des programmes de mobilité⁶ et l'augmentation des bénéficiaires, les objectifs initiaux sont loin d'être atteints. La mobilité reste encore marginale. La mobilité géographique reste toujours un signal fort de réussite et de prestige.

De plus en plus de jeunes, ces dernières années, se sont rendus à l'étranger pour réaliser un séjour, que ce soit principalement dans les différents pays de l'Union européenne, au Québec, dans le reste du Canada ou encore en Australie. Néanmoins, si de plus en plus de jeunes se déplacent, le chiffre relatif reste faible. Une fois de plus, la mobilité concerne encore les

jeunes les mieux informés, les mieux qualifiés, les plus diplômés.

Il est vrai que, malgré l'existence de dispositifs d'aide à la mobilité et les efforts de la Commission européenne, force est de constater que ceux-ci profitent la plupart du temps aux étudiants diplômés ou aux jeunes en recherche de stage. Ceux qui sont très faiblement diplômés, en décrochage ou exclus du système scolaire, avec des situations professionnelles précaires, sont très peu, voire pas du tout, la cible des programmes d'aide à la mobilité. L'enjeu pour l'expérimentation était donc de réussir à capter un public n'ayant jamais eu d'expérience de mobilité internationale *via* des dispositifs existants, mais aussi de développer des mesures d'accompagnement adaptées pour baliser les différentes étapes et sécuriser les jeunes dans leurs démarches. Le montage et la réalisation d'un tel projet ne pouvait se faire sans un partenariat fort, l'information étant aussi importante que l'accompagnement. Le pari pour les porteurs du projet était de déclencher, par le voyage, la remobilisation du jeune, son implication dans la construction de son insertion professionnelle, l'apprentissage de l'estime et de la confiance en soi, et du respect des autres, et une ouverture socioculturelle. L'idée de la mission locale du Grand Besançon était d'ouvrir la mobilité internationale, les séjours à l'étranger, à un public qui en

NOTE

6. La mobilité internationale est surtout connue et appréciée au travers des programmes communautaires (Erasmus, Socrates, Leonardo Da Vinci, Erasmus Mundus, Tempus, Promoteurs de Bologne, Label européen des langues, Europass, etc.).

est éloigné, les apprentis et jeunes en insertion professionnelle, typiquement la cible des JAMO que nous avons évoquée dans nos propos précédents. Le voyage devait déclencher la remobilisation des jeunes, permettant leur implication dans la construction de leur insertion professionnelle. Les qualités et compétences sollicitées dans le cadre d'une mobilité sont un outil de formation dans la mesure où, si la démarche est balisée, elle implique la compréhension et la maîtrise de notions et de valeurs. Si bien qu'au retour, il est espéré que la plus-value liée à la mobilité serve directement les jeunes dans leur démarche et se caractérise, à terme, par des retombées sur leur insertion professionnelle.

Empiriquement, l'inscription dans une démarche de mobilité nécessite une capacité de projection, une lecture et compréhension des dispositifs et financements disponibles, des ressources financières et un certain appui de l'entourage. Ainsi, l'investissement personnel intervient dès la phase de compréhension des conditions et critères d'éligibilité des dispositifs. Il faut pouvoir s'inscrire dans un processus souvent long et complexe, composé de plusieurs étapes : contact d'un « conseiller mobilité », présentation d'un projet cohérent, constitution du dossier avec regroupement des pièces et documents administratifs, préparation au départ, remise à niveau dans la langue du pays, recherche d'information sur la destination, préparation de l'arrivée sur place (recherche de logement, information sur la ville, les transports, etc.). De ce fait, un jeune, qui est incapable de se projeter dans un futur proche, ne se préoccupera pas de mobilité, de séjour à l'étranger. Tout comme les intentions volontaristes concernant une dynamique partenariale, la volonté

affichée d'encourager des jeunes faiblement diplômés ou en transition professionnelle doit se traduire par un effort conséquent non seulement en matière d'information mais également en termes d'accompagnement des projets. En effet, les acteurs de la mobilité s'impliquent pour améliorer l'accès à l'information mais ont du mal à cerner le besoin d'accompagnement pourtant exprimé par les jeunes.

L'enjeu pour l'expérimentation, réussir l'élargissement de la mobilité internationale à un public qui en est éloigné, consisterait, au-delà de la simple mise à disposition de programmes sur mesure, à développer des mesures d'accompagnement adaptées pour baliser les différentes étapes et sécuriser les jeunes dans leurs démarches. Sans quoi, cet enjeu restera de l'ordre des incantations volontaristes.

LE PARI GAGNÉ DE LA MISSION LOCALE DU GRAND BESANÇON

L'expérimentation sociale portée par la mission locale du bassin d'emploi du Grand Besançon est réellement innovante dans le sens où le projet d'action ne préexistait pas, ni dans la conduite, ni dans le financement. Il a été créé de toute pièce par le directeur de la mission locale, qui s'est immédiatement saisi de la thématique de la mobilité internationale comme outil d'insertion des jeunes.

Une expérimentation réussie grâce à l'implication des acteurs

Pour répondre aux difficultés et promouvoir la mobilité internationale, la mission locale du bassin d'emploi du Grand Besançon et ses partenaires⁷

ont proposé trois formes de séjours à l'international, comportant chacun un accompagnement pédagogique (évaluation linguistique, acculturation au pays de destination, sensibilisation aux valeurs civiques, etc.) :

- un séjour international (en forêt amazonienne) à vocation pédagogique et citoyenne pour un groupe de 21 jeunes de niveau IV et *infra* IV, majeurs, dont 8 inscrits en mission locale et 13 apprentis en formation « travaux forestiers » ;

- un stage socioprofessionnel à Berlin autour des métiers de la restauration pour un groupe de 9 jeunes inscrits en mission locale ;

- 16 projets individuels permettant à des jeunes, dont les dispositifs classiques d'aide à la mobilité internationale ne suffisaient pas, de construire leur propre projet.

À ces trois actions s'est ajoutée une campagne de sensibilisation et de promotion de la mobilité internationale, organisée conjointement avec les journées de l'apprentissage, en valorisant notamment des expériences positives de jeunes dans ce domaine (participation à des manifestations, vidéo) et en utilisant le support du portail Jeunesse du CRIJ.

Enfin, une plate-forme collaborative dédiée à la mobilité des jeunes⁸ regroupe les acteurs nécessaires à la réussite des projets individuels ou

NOTES

⁷. Le conseil régional de Franche-Comté, l'association régionale des missions locales de Franche-Comté, le centre régional d'information jeunesse, la ville de Besançon et son CCAS, le groupe partenarial de promotion de l'apprentissage et le centre de formation des apprentis agricole du Doubs.

⁸. PHILEAS : <http://www.jeunes-fc.com/phileas/>

collectifs. Au-delà des outils (portail informatique, guide, documents, etc.), la synergie des acteurs et la conjugaison des savoir-faire ont permis, dans une logique de guichet unique, de co-construire et d'organiser avec le jeune son projet de mobilité internationale en lui donnant toutes les chances de réussite.

Cette expérimentation a été testée sur un territoire restreint (communauté d'agglomération du Grand Besançon) mais souhaite s'étendre au territoire régional. L'outil créé est également transposable à d'autres thématiques, telles que le logement, et transférable à d'autres partenaires de l'expérimentation, comme le CRIJ.

Une posture d'évaluation : l'évaluation embarquée⁹

Tout au long du processus, l'évaluateur s'est positionné d'emblée comme une force de proposition vis-à-vis de l'expérimentateur, afin de s'adapter aux besoins de l'expérimentation elle-même. De son côté, l'expérimentateur était en amont conscient des bénéfices que pouvait procurer un processus évaluatif embarqué. Cette congruence entre les deux parties impliquées a notamment permis le développement d'un langage partagé co-construit à chaque étape du protocole. Cette co-construction est innovante pour l'espace territorial dans lequel elle se déroule.

Les bénéfices de cette démarche sont grands : le rôle d'intermédiaire entre expérimentateur et évaluateur est primordial ; il facilite les échanges d'informations, la connaissance partagée et les adaptations nécessaires aux impératifs de chacun. « L'enjeu est [ainsi respecté de manière optimale] de concilier des programmes sociaux

innovants avec des méthodes d'évaluation rigoureuses » (id.). La connaissance que les porteurs de projet ont de l'évaluation est primordiale pour engager une relation de confiance entre les deux parties car « toute évaluation prend place dans une situation qu'elle perturbe. » [Potier, Bencivenga, 2005 p. 113].

Dans la conception des outils d'évaluation, la proximité géographique et la relation de confiance tissée entre expérimentateur et évaluateur permettent de discuter des outils pour mieux les ajuster à la spécificité du public jeune accueilli dans les missions locales. L'obligation d'évaluer les expérimentations n'a pas posé de problème particulier à l'expérimentateur, ce dernier ayant la culture de l'évaluation.

Le projet d'expérimentation mettait en place conjointement plusieurs actions distinctes, pas toutes destinées au même public, ce qui rendait l'exercice de l'évaluation difficile. Aucune évaluation quantitative n'a pu se révéler véritablement pertinente. Les raisons sont aisément compréhensibles. Premièrement, la faiblesse des effectifs (environ 50 jeunes répartis dans trois dispositifs) ne nous autorisait pas à mener des analyses statistiques robustes. Deuxièmement, le temps court de l'expérimentation – un an –¹⁰ ne nous permettait pas de mesurer la portée d'une telle action et l'impact sur les individus en termes d'insertion professionnelle, ces éléments ne pouvant être saisis qu'à plus long terme. L'évaluation s'est donc basée sur une méthode qualitative ; le but de l'évaluation proposée a été de repérer les obstacles au bon fonctionnement du programme et les leviers qui permettraient d'en améliorer l'efficacité dans l'optique d'une généralisation. L'évaluation qualita-

tive a été menée « au fil de l'eau », de manière parallèle à l'expérimentation sous trois formes :

- l'interrogation des acteurs partenaires de l'expérimentation : deux entretiens semi-directifs, au cours de l'année, avant et à l'issue de l'expérimentation ;

- l'interrogation des bénéficiaires de l'expérimentation : entretiens semi-directifs ou enquête papier avant et au retour du séjour. Cette interrogation n'a parfois pas été réalisée lorsque les bénéficiaires étaient indisponibles ;

- l'observation directe : immersion dans le projet, présence lors des réunions de travail, des manifestations, des comités de pilotage, etc.

Dans la conception des outils d'évaluation, la proximité géographique et la relation de confiance tissée entre expérimentateur et évaluateur ont permis de discuter des outils pour mieux les ajuster à la spécificité du public jeune accueilli dans les missions locales. Un langage et des informations partagés ont véritablement servi le projet par leur capacité à approfondir des questions sensibles, mettre en lumière des contradictions ou blocages, et à interpeller les acteurs sur des lacunes ou limites. L'évaluateur a pu naturellement prendre sa place au sein du comité de pilotage et se voir attribuer

NOTES

9. Évaluation embarquée (ou « *on-going* ») : elle se déroule pendant toute la période de mise en œuvre de l'intervention. Elle permet à l'évaluateur de comprendre le contexte et les conditions de mise en œuvre de l'intervention. Elle implique l'évaluateur et l'expérimentateur dans une démarche d'échange et favorise ainsi leur collaboration. Cf. « Les démarches d'évaluation en question », Rouaud et Sulzer, 2011.

10. L'expérimentation est désormais terminée.

un temps de parole spécifique afin de faire un point sur l'évaluation. Ce temps de parole n'excluait pas une intervention au cours des discussions afin d'échanger autour des outils d'évaluation dans le déroulement des actions.

Dans le respect du cahier des charges, en amont du démarrage du projet, une enquête en ligne a été construite à destination des jeunes de centres de formation d'apprentis (CFA), du centre régional information jeunesse (CRIJ) et de la mission locale afin de mieux connaître le public cible, d'appréhender sa connaissance des structures d'aide à la mobilité internationale. Cette enquête a notamment révélé que 18 % des 392 jeunes ayant répondu, déclaraient ne pas être intéressés par un départ à l'étranger. Ils évoquaient principalement l'absence de désir et déclaraient se trouver bien en France. Une enquête auprès du même public devait avoir lieu à la fin de l'expérimentation. Cependant les difficultés de mise en œuvre de l'enquête en ligne (mobilisation des équipes mission locale/CRIJ et CFA) et la quasi inexistence de résultats mesurables sur l'impact de la mobilisation des acteurs autour du jeune (portail collaboratif, journée de l'apprentissage et de la mobilité internationale) nous ont conduit à abandonner cette action.

Les effets probants de l'expérimentation sur le développement des partenariats

En termes de mesure d'impacts sur les bénéficiaires, les enquêtes qualitatives ont montré une certaine difficulté à obtenir des informations sur leur perception du séjour à l'étranger et l'impact sur leurs comporte-

ments. Néanmoins, il est important de souligner deux effets positifs. Tout d'abord, les jeunes partis à l'étranger se sont réellement impliqués dans la construction d'un projet (les attentes formulées ont d'ailleurs été satisfaites dans la majorité des cas). Par ailleurs, la préparation avant départ, pour ce type de public, s'est révélée essentielle pour le bon déroulement du séjour (le but étant d'éviter au maximum un double échec pour les jeunes en mission locale ayant déjà connu auparavant un échec scolaire). Le problème lié à l'analyse de l'impact de la mobilité sur le jeune réside dans la formulation par le jeune de l'évolution de son comportement (dans la construction d'un projet professionnel, dans la recherche d'emploi, dans la confiance en soi, etc.). Les conclusions qu'ils tirent de leur séjour sont souvent formulées de manière peu étoffées, mais il reste dans leur propos les traces d'un impact de ce séjour sur la maturation de leur projet professionnel, voire leur projet de réorientation. Ces tendances se confirment également de part le contenu des entretiens avec les conseillers des missions locales qui soulignent la difficulté d'obtenir, parfois, de ce public, des informations véritablement traitables. En outre, l'expérience d'un débriefing en groupe, d'un séjour collectif, ne s'est pas révélée plus constructive. L'évaluation a reposé également, pour pallier les difficultés précitées à obtenir des informations satisfaisantes, sur un entretien avec les référents de la mission locale et les encadrants pendant le séjour.

C'est sur la coordination des acteurs régionaux de la mobilité que les effets les plus importants de l'expérimentation se manifestent.

On peut donc relever des effets plus que positifs sur la coordination des acteurs régionaux de la mobilité. Au-delà de l'organisation des séjours à l'étranger, l'objet principal de l'expérimentation était de co-construire un outil permettant aux structures de travailler ensemble autour d'une des problématiques rencontrées par le public des missions locales.

L'expérimentation a permis de nombreux rapprochements, une meilleure connaissance des différents acteurs et services locaux (que ce soit autour de la mobilité mais aussi d'autres thématiques liées). Les Journées de l'apprentissage et de la mobilité internationale avec la présence des partenaires agissant dans le champ de la mobilité ont été une première, une réussite en matière de travail en réseau. Ainsi, autour du portail collaboratif, une vingtaine de partenaires se sont associés afin de rassembler tous les acteurs de la mobilité internationale (allant des dispositifs classiques à des chantiers de jeunes, des projets de solidarité internationale, etc.). Ce rapprochement de structures qui se méconnaissaient auparavant a permis de nouer de nouveaux partenariats entre structures. À ce titre, l'expérimentation a su insuffler une dynamique endogène qui reste à animer et à pérenniser. L'outil qu'est le portail interactif et collaboratif doit permettre de réunir les acteurs de façon durable autour d'un accompagnement concerté des projets des jeunes. Les acteurs ont donc su recentrer la problématique autour du jeune afin de promouvoir la mobilité de tous les jeunes.

L'évaluation de cette expérimentation a permis de mettre en lumière les aspects très positifs attachés à la

démarche d'une part, les effets bénéfiques qui mériteraient d'être consolidés d'autre part. La mise en réseau des acteurs, l'organisation collective autour d'un projet commun clairement défini sont à l'évidence une réussite – sentiment partagé par les acteurs. Par ailleurs, les résultats obtenus peuvent apparaître modestes dans une certaine mesure eu égard aux effectifs et à la difficile mesure des effets de l'expérimentation. Cependant, parce qu'ils ont permis notamment des réorientations possibles, ils laissent augurer l'identification et la multiplication de résultats probants dans le cadre de la mise en place d'un dispositif renforcé, soutenu et régulièrement reconduit.

CONCLUSION

La volonté d'impulser une dynamique de réseau a porté ses fruits dans la mesure où plusieurs acteurs de la mobilité se connaissent et ont désormais une expérience de travail en commun. Il reste des perspectives et des chantiers pour entretenir cette dynamique : accompagnement personnalisé, valorisation de la mobilité, géolocalisation des boursiers, etc. Les acteurs sont volontaires et mobilisés pour étendre l'expérience à toute la région et à d'autres thématiques, le logement par exemple.

Cette expérience a finalement profité au territoire en entraînant une meilleure coordination et coopération entre acteurs autour de la mobilité des jeunes. Les bénéficiaires pour le public cible ne peuvent pas, en revanche, être perceptibles à court terme. Cependant, le concept d'intelligence territoriale prend ici tout son sens et il est certain qu'à long terme, les dynamiques de col-

laboration et de coordination initiées avec ce projet au niveau local auront des répercussions positives à terme sur les JAMO.

Les programmes d'expérimentations envers la jeunesse, initiés par le gouvernement, sont sensés créer des dynamiques vertueuses et innovantes dans les territoires expérimentaux, en particulier à travers des dynamiques collaboratives, des rapprochements d'acteurs et une évaluation au fil de l'eau. Le résultat doit être une innovation sociale, généralisable.

La première vague des appels à projets HCJ a retenu 165 projets, disséminés sur tout le territoire national, mais de nature diversifiée, tant sur les publics, les thématiques, que sur les territoires eux-mêmes.

On le voit, il ne peut y avoir d'évaluation « unique », d'évaluation-type. Il s'agit d'une opération sur mesure, propre à chaque expérience.

Le HCJ, dans son « guide d'évaluation des expérimentations sociales » évoque pourtant le recours aux méthodes classiques qualitatives et/ou quantitatives, avec une préférence affichée pour les méthodes quantitatives avec assignation aléatoire et constitution de deux groupes, cible (bénéficiant du dispositif) et témoin (évoluant sans intervention et non bénéficiaires). D'un point de vue empirique, la démarche quantitative est nécessaire à des fins de généralisation mais elle est difficile à mettre en œuvre dans le cadre strict des projets HCJ. Les raisons sont diverses : les effectifs ne sont souvent pas suffisants, la constitution d'un groupe témoin pose des problèmes éthiques ou tout au moins des réticences, etc.

Les méthodes qualitatives sont en conséquence une bonne alternative et sont dans l'esprit de la commande du

HCJ. Elles permettent une cartographie des réseaux d'acteurs, évaluant au fil du temps en révélant et signalant les relations réelles entre acteurs désignés, les oppositions possibles rencontrées, les associations induites ou envisageables, les éléments bloquants par principe ou par nécessité stratégique ou structurelle, les facilitateurs, « défricheurs », et volontaires mobilisables, les évolutions dans les jeux d'acteurs, etc.

Il n'existe pas, en tout cas à ce jour, de recherches sur des méthodes « transposables » d'évaluation, pas de cadre commun, consensuel et adaptable à toute expérience sociale. Seul un travail *a posteriori* à partir des résultats des expériences permettra peut-être de réfléchir, d'ouvrir un champ de recherche sur l'évaluation d'expérimentations sociales, de créer des typologies d'innovations sociales, d'aller vers des méthodes communes et originales, applicables à différents contextes.

Des recherches plus approfondies sur ces nouvelles lois expérimentales et leur généralisation seraient bienvenues, mais il est encore prématuré de porter des jugements.

En l'état actuel de nos recherches (nous évaluons plusieurs projets d'expérimentations sociales en Franche-Comté), nous voyons se dégager quelques facteurs-clés pour la réussite des projets et quelques obstacles réhibitoires. La réussite semble résider dans la diffusion des connaissances (tacites surtout, et codifiées), dans le transfert, les réseaux, les partenariats, la gestion participative, et dépend également du tempérament des porteurs de projets en termes de créativité, *leadership*, charisme et prise de risque. Les freins que nous avons pour le moment rencontrés

parfois sur d'autres projets, reposent principalement sur la perception des principes même de l'expérimentation et sur l'appréhension des impé-

ratifs attachés à l'évaluation. Ces obstacles ne sont à aucun moment apparus à l'occasion de notre évaluation, illustrant ainsi l'adhésion initiale des

acteurs et la cohésion effective autour d'un projet concerté, construit, et de par sa nature et ses objectifs, fédérateur. ■

BIBLIOGRAPHIE

Baslé M. (2008) Suivi et évaluation des politiques publiques et des programmes, *Économica*, Paris.

Baslé M. (2010) Connaissance et action publique, *Économica*, Paris.

Carel S., Déloye F., Mazouin A., (2011) La mobilité internationale des « jeunes avec moins d'opportunités » : retour d'expérience. *Bref du Céreq* n° 293.

Carel S., Dubos-Paillard E., Giannotti M., Mazouin A. (2010), Générer de l'innovation sociale : le pari des expérimentations du Haut Commissaire à la Jeunesse. Un exemple : dynamique endogène et synergie des acteurs autour de la mobilité des jeunes à Besançon, *Évaluation et données longitudinales : quelles relations ?*, coord. T. Couppié, D. Epiphane *et al*, Relief n° 30/Céreq.

Divay S., (2011) L'évaluation qualitative : objectiver sans compter, *Bref du Céreq* n° 286.

Haut Commissariat à la Jeunesse, (2009). *Guide méthodologique pour l'évaluation des expérimentations sociales*, Paris.

Lussault M. (2004), la mobilité comme évènement in *Le sens du mouvement, modernité et mobilités des sociétés urbaines contemporaines*, p. 116, coord. S. Allemand, F. Hascher, J. Lévy. Éditions Belin, Paris.

Rouaud P., Sulzer E., (2011) Les démarches d'évaluation en question. *Bref du Céreq* n° 281.

Potier V. et Bencivenga M. (2005), *Évaluation des politiques locales. Évaluation interne et externe, principes et dispositifs, méthodes*. Éditions Le Moniteur, Paris.

Autres publications de la DEPP

n° 78 – novembre 2008

Thème : Comparaisons internationales

Le niveau d'huile, le moteur et la voiture : les enjeux d'une évaluation de la qualité de l'enseignement par les indicateurs

Norberto Bottani

La situation éducative de la France comparée à celle d'autres pays de l'Union européenne et de l'OCDE

Nadine Dalsheimer, Claude Sauvageot

Les indicateurs de l'OCDE sur les dépenses d'éducation en 2005 : quelques tendances sur la situation comparée de la France

Christine Ragoucy

Que nous apprennent les évaluations internationales sur le fonctionnement des systèmes éducatifs ? Une illustration avec la question du redoublement

Thierry Rocher

La mesure de la littératie dans PISA : la méthodologie est la réponse, mais quelle était la question ?

Pierre Vrignaud

L'évaluation des compétences des adultes : des méthodes en plein développement

Fabrice Murat

Évaluation des langues étrangères : présentation du projet coopératif d'évaluation EFLUSL

Paul Caffrey

L'importance des enseignants, leur situation et leurs conditions de travail en Suède

Sten Söderberg, G. Iselau, et al.

Les enseignants en France : une spécificité en Europe ?

Nadine Esquieu

La taille des structures pédagogiques dans les établissements français du second degré en France en 2007

Paola Serries

Mesurer et comparer l'équité des systèmes éducatifs en Europe

Marc Demeuse, Ariane Baye

Analyse des classements internationaux des établissements d'enseignement supérieur

Nadine Dalsheimer, Denis Desprésaux

Les sorties précoces en Europe

Pascale Poulet-Coulibando

Genre et valorisation des compétences sur les marchés du travail en Europe

Jean-François Giret, et al.

Les études combinées avec les emplois par les jeunes : comparaison entre pays européens

Pascale Poulet-Coulibando

Un outil pour les comparaisons internationales : la CITE

Claude Sauvageot

n° 79 – décembre 2010

Thème : Les élèves : connaissances, compétences et parcours

La maîtrise de la langue française à l'épreuve du socle commun de connaissances et de compétences

Marlène Guillou

Les acquis des élèves en calcul à l'issue de l'école primaire

Jean-François Chesné

La place et le rôle de l'écrit et de l'oral au collège

Catherine Régnier

Le travail en dehors de la classe donné par les professeurs de collège

Catherine Régnier

L'attestation de la maîtrise du socle commun est-elle soluble dans le jugement des enseignants ?

Jeanne-Marie Daussin, et al.

Les cours particuliers en première année de collège : un entrant en sixième sur dix bénéficie de soutien scolaire payant

Jean-Paul Caille

Sortir sans diplôme du système éducatif : une nouvelle approche des déterminants socio-économiques

Rachid Bouhia, Thibaut de Saint-Pol

Impact de la catégorie sociale sur des performances moindres dans les petits collèges

Pascale Guillois, Jean-Pierre Astoul

Une meilleure mesure du contexte socio-éducatif des élèves et des écoles – Construction d'un indice de position sociale à partir des professions des parents

Noémie Le Donné, Thierry Rocher

La carte scolaire, un éclairage international

Nadine Dalsheimer-Van Der Tol

n° 80 – décembre 2011

Thème : Méthodes internationales pour comparer l'éducation et l'équité

Mesurer l'éducation à l'ère de la « société de la connaissance » : les usages de la Classification Internationale Type de l'Éducation

Rémi Tréhin-Lalanne

Les notions statistiques d'éducation, de « l'enseignement habituel » à « l'éducation formelle »

Pascale Poulet-Coulibando

Comparaison entre pays des coûts de l'éducation : des sources de financement aux dépenses

Luc Brière, Marguerite Rudolf

La CITE 2011 : quelles perspectives ? Quels défis ?

Claude Sauvageot

L'évaluation de l'équité scolaire : perspectives nationales et internationales

Estelle Herbaut

Goûts, habitudes et performances en lecture des élèves de 15 ans d'après PISA

Sylvie Fumel, Bruno Trosseille

La mesure de l'équité dans PISA : pour une décomposition des indices statistiques

Saskia Kespaik, Thierry Rocher

Les parents et l'école en France et en Europe

Nadine Dalsheimer-Van Der Tol, Fabrice Murat

De la lutte contre les inégalités à l'adaptation aux « besoins spécifiques »

Daniel Frandji, Jean-Yves Rochex

Décrire l'éducation et la formation pour comprendre les caractéristiques de la main-d'œuvre : les questions des « enquêtes sur les forces de travail » en Allemagne, en France et au Royaume-Uni

Annick Kieffer, Rémi Tréhin-Lalanne

Niveau d'éducation : que mesurent les statistiques internationales ?

Pascale Poulet-Coulibando

Niveau d'éducation : expérience d'une collaboration entre chercheurs et statisticiens

Annick Kieffer

Enseignement scolaire

180 (mars 2007) – L'évaluation internationale PISA 2003 : compétences des élèves français en mathématiques, compréhension de l'écrit et sciences, 245 p.
Anne-Laure Monnier

181 (mars 2007) – L'image des sciences physiques et chimiques au lycée (LEGT et LP), 90 p.

François Alluin

182 (mars 2007) – Les représentations de la grande difficulté scolaire par les enseignants, 116 p.

Chi-Lan Do

183 (mars 2007) – Image de la discipline et pratiques d'enseignement en histoire-géographie et éducation civique au collège, 168 p.

Jean-Claude Guillaume

185 (juillet 2007) – Compréhension à l'écrit et à l'oral des élèves en fin d'école primaire – Évaluation 2003, 224 p.

Jean-Marc Pastor

186 (juillet 2007) – Les attitudes à l'égard de la vie en société des élèves en fin d'école primaire et en fin de collège, 302 p.

Saïd Yacoub

187 (septembre 2007) – Les acquis des élèves en anglais et en allemand en fin d'école en 2004 et les contextes favorables à ces apprentissages, 288 p.

Annick Marquis

188 (septembre 2007) – Image de la discipline et pratiques d'enseignement en histoire-géographie et éducation civique, juridique et sociale au lycée professionnel, 204 p.

Marion Billet

190 (octobre 2007) – Image du sport scolaire et pratiques d'enseignement au collège et au lycée – 2005-2006, 136 p.

Jeanne Benhaïm-Grosse

196 (août 2010) – Les compétences des élèves en histoire, géographie et éducation civique en fin de collège, 120 p.

Michel Braxmeyer, Nicole Braxmeyer, Séverine Dos Santos

197 (octobre 2010) – Les technologies de l'information et de la communication (TIC) en classe au collège et au lycée : éléments d'usage et enjeux, 84 p.

François Alluin

Enseignants et personnels de l'éducation

189 (septembre 2007) – La formation des nouveaux personnels de direction – Enquête auprès de la promotion des lauréats 2002, 160 p.

Michèle Thauvel-Richard

192 (juin 2008) – Portrait des nouveaux personnels de direction du second degré. Interrogation de la promotion 2004, six mois après sa nomination en février-mars 2005, 253 p.

Michèle Thauvel-Richard

193 (septembre 2008) – Les inspecteurs territoriaux et leur formation à l'école supérieure de l'éducation nationale (ESEN), 196 p.

Nadine Esquieu et Alain Lopes

194 (octobre 2009) – Enseigner en collège et lycée en 2008 – Interrogation réalisée en septembre-octobre 2008 auprès de 1 200 enseignants du second degré dans les collèges et lycées publics, TMO-Régions, 222 p.

Patricia Gambert, Jacques Bonneau (TMO-Régions)

195 (février 2010) – Être professeur d'éducation physique et sportive en 2009, TMO-Régions, 146 p.

Patricia Gambert, Jacques Bonneau (TMO-Régions)

Insertion, éducation et société

184 (mai 2007) – Pratiques citoyennes des lycéens dans et hors établissement, 136 p.
Roseline Verdon

191 (novembre 2007) – Points de vue sur l'apprentissage – Actes du colloque « Points de vue sur l'apprentissage » du 28 novembre 2006, 110 p.

Pascale Pollet

198 (novembre 2010) – Éclairages internationaux sur quelques questions d'actualité, 114 p.

Nadine Dalsheimer-Van Der Tol

Moyens et coûts

199 (mai 2011) – Le compte de l'éducation – Principes, méthodes et résultats pour les années 1999 à 2009, 212 p.

Martine Jeljoul, Jean-Pierre Dalous, Luc Brière

D'autres publications de la DEPP

Repères et références statistiques sur les enseignements, la formation et la recherche

Des repères concernant tous les aspects de l'appareil de formation, des références indiscutables. Un vaste ensemble d'informations statistiques selon une présentation unique : des textes clairs et synthétiques commentant les tableaux statistiques, cartes et graphiques. Les définitions, les sources, les références documentaires en font un outil indispensable. Tous publics : enseignants, chercheurs, journalistes, étudiants, etc. L'édition 2011 présente les données statistiques de l'année 2010-2011.

Parution annuelle. Septembre 2011.

L'état de l'École n° 21

Une analyse synthétique des principales caractéristiques de notre système éducatif. 30 indicateurs renseignent sur les coûts, les activités et les résultats de l'école.

Des comparaisons internationales permettent de mieux situer la France par rapport aux États-Unis et aux principaux pays européens.

Parution annuelle. Novembre 2011.

... et toujours

la Note d'information

Parution hebdomadaire

Consulter les publications de la DEPP sur Internet

www.education.gouv.fr/statistiques

Consignes aux auteurs

La saisie du texte

Les projets d'articles doivent comporter 12 pages, maximum souhaité, tableaux et graphiques compris.

Les textes doivent être saisis en Arial 11 sous Word 97, 2000 ou 2003 (compatible PC), marges droite et gauche de 2,5 cm.

Les titres et intertitres doivent être numérotés clairement en respectant leur hiérarchie.

L'auteur indiquera, sous son nom, ses fonctions et l'établissement dans lequel il les exerce.

Les tableaux et graphiques doivent être réalisés sous Excel (compatible PC), numérotés (tableau 1, 2, 3, ...) et comporter un titre.

Les graphiques doivent être liés aux tableaux de données et accompagnés du fichier Excel à partir duquel ils ont été réalisés.

Les schémas et cartes doivent être transmis au format EPS ou WMF.

Nous ne pouvons traiter **aucun tableau ni graphique composé dans Word.**

Les notes et références bibliographiques

Les « notes de bas de page » sont saisies normalement par l'auteur **en fin de document sans utiliser la fonction « renvois » ou « notes »**. Elles sont appelées dans le texte par un numéro entre parenthèses : (1), (2), (3) ..., la rédaction de la revue les remplacera sur la bonne page lors de l'édition.

Les références bibliographiques sont appelées dans le texte par leur numérotation entre crochets : **[1], [2], [3]**, ..., et listées en fin de document de la même manière en veillant à ce qu'elles soient complètes :

- **pour les ouvrages** : nom de l'auteur, titre de l'ouvrage, ville, éditeur, date d'éditions et si possible nombre de pages de l'ouvrage ;
- **pour les revues** : nom de l'auteur, titre de l'article, nom de la revue, numéro, mois, année.

Le chapeau

Chaque article doit obligatoirement être accompagné d'un chapeau. Le chapeau est un résumé de l'article dégageant sa problématique et ses principaux résultats. Il ne doit pas dépasser 12 lignes (environ 165 mots), marges 2,5 cm comme le texte. Il est placé après le titre et le nom de l'auteur en tout début d'article.

Remise des textes, graphiques, cartes, etc.

Les projets d'articles sont à transmettre à la rédaction d'*Éducation & formations* par mail, adressés à :

Alain Lopes, chef du département de la valorisation et de l'édition, revue *Éducation & formations*,

DEPP-DVE, 61-65, rue Dutot – 75732 Paris Cedex 15.

☎ 01 55 55 76 31.

Par mail : alain.lopes@education.gouv.fr

Pour plus d'informations, vous pouvez contacter :

Marc Saillard

☎ 01 55 55 72 63.

marc.saillard@education.gouv.fr

*Achévé d'imprimer
en mars 2012*

Commander



À retourner à :

Direction de l'évaluation, de la prospective et de la performance

Département de la valorisation et de l'édition

61-65 rue Dutot – 75015 Paris

☎ 01 55 55 72 04 Fax 01 55 55 72 29

*Toute commande
d'ouvrage ou
souscription
d'abonnement doit
être accompagnée
d'un titre de
paiement libellé
à l'ordre de*
**Régie de recettes
MEN-DEPP**

Mme, Mlle, M. Établissement (s'il y a lieu)	
Adresse :	
Boîte postale :	
Code postal : <input type="text"/>	Ville :

La facture devra être envoyée à l'adresse ci-dessous
(si elle est différente de la précédente)

*Pour les administrations
uniquement, règlement
possible par mandat
administratif, avec
rappel impératif du
numéro de facture
que nous vous
ferons parvenir*

Nom ou dénomination :	
Adresse :	
Boîte postale :	
Code postal : <input type="text"/>	Ville :

Revue **Éducation & formations**

Prix de vente au numéro :

du n° 45 au n° 48 : **10,67 euros**

du n° 49 au n° 52 : **11,43 euros**

du n° 53 au n° 64 : **12,20 euros**

À partir du n° 65 : **13 euros**

Commandes au numéro

Référence	Quantité	Prix unitaire	Prix total

collection

Éducation & formations

éditeur

Direction de l'évaluation, de la prospective
et de la performance

date de parution

Mars 2012

conception, réalisation

DVE / Département de la valorisation et de l'édition

impression

Ovation

accès internet

www.education.gouv.fr/statistiques

ISBN 978-2-11-099366-3



9 782110 993663

ISSN 0294-0868

ISBN 978-2-11-099366-3

DEPP 00512181

Prix 13 euros

