



ÉVALUATION DES EFFETS DU DISPOSITIF EXPÉRIMENTAL D'ENSEIGNEMENT INTÉGRÉ DE SCIENCE ET TECHNOLOGIE (EIST)

Marion Le Cam

MENESR-DEPP, bureau de l'évaluation des élèves

Olivier Cosnefroy

Université Grenoble Alpes, laboratoire des sciences de l'éducation

Remerciements à **Pascal Bessonneau** (MENESR-DEPP) pour la mise en œuvre d'une partie des analyses psychométriques.

Cet article présente les résultats de l'évaluation des effets du dispositif expérimental d'enseignement intégré de science et technologie (EIST). Ce dispositif a été évalué à partir de la rentrée 2008. Pendant quatre ans, une cohorte composée d'élèves ayant bénéficié de l'EIST en classe de sixième et d'élèves n'en ayant pas bénéficié, a été suivie jusqu'en fin de troisième et évaluée à cinq reprises. Ces évaluations permettent de mesurer l'évolution d'un score cognitif de performance en sciences des élèves, ainsi que de deux scores conatifs mesurant la motivation intrinsèque pour les sciences et l'intérêt pour les sciences en dehors de l'école. Des modélisations ont été engagées pour identifier l'existence d'un effet de l'EIST sur la progression des élèves en sciences, et sur l'évolution de leurs attitudes envers les sciences au cours du temps. Nous n'observons pas d'effet de l'EIST sur la progression des performances en sciences des élèves tout au long de leur scolarité au collège. En moyenne, l'intérêt et la motivation des élèves pour les sciences tendent à décroître au cours des années de collège. Si les élèves ayant bénéficié de l'EIST présentent un niveau de motivation légèrement plus élevé en début de sixième, cet écart moyen reste cependant stable jusqu'en fin de collège.

De nombreux rapports d'institutions internationales, parus au début des années 2000, font état d'une désaffection des étudiants pour les filières scientifiques. En 2000, le Conseil européen de Lisbonne avait émis le souhait que l'Union européenne devienne l'économie basée sur la connaissance

la plus compétitive et la plus dynamique d'ici 2010, ce qui se traduisait par la volonté d'attirer et de retenir des chercheurs de haut niveau. Or, un rapport de la Commission européenne de 2004, intitulé « L'Europe a besoin de scientifiques » [GAGO, CARO *et alii*, 2004], s'alarme du manque d'engouement pour les carrières scientifiques de la part des jeunes et émet des recommandations afin que l'Union européenne se donne les moyens d'atteindre les objectifs de recrutements, dans le secteur de la recherche, nécessaires pour mettre l'Europe en tête en matière d'excellence scientifique et technologique. En 2006, un rapport d'orientation de l'OCDE fait le même constat, cherche à en étudier les causes et émet également un certain nombre de recommandations pour modifier cette tendance [OCDE, 2006]. Selon ce rapport, il convient, notamment, de réformer l'enseignement des sciences et technologies, et les programmes, qui ne favoriseraient pas l'intérêt des jeunes pour les sciences. En particulier, la pédagogie devrait « être concentrée plutôt sur les concepts et les méthodes scientifiques que sur la seule mémorisation de l'information », ces objectifs étant « particulièrement importants dans l'enseignement secondaire ». D'autres rapports [OSBORNE et DILLON, 2008 ; ROCARD, CSERLELY *et alii*, 2007] relient le déclin de l'intérêt des jeunes pour les études scientifiques à la façon dont les sciences sont enseignées, et préconisent de développer les pratiques pédagogiques basées sur des méthodes d'investigation, qui seraient plus efficaces pour accroître l'intérêt des élèves pour les sciences, ainsi que leur niveau de réussite dans ces disciplines.

Toutefois, si les méthodes d'enseignement privilégiant les activités expérimentales et un rôle actif des élèves permettent effectivement de susciter un engagement de court terme de la part des élèves, elles ne suffisent pas forcément pour les motiver à poursuivre des études scientifiques [ABRAHAMS et REISS, 2012 ; ABRAHAMS et MILLAR, 2008], et leur impact sur les performances des élèves en sciences reste largement discuté [ALFIERI, BROOK *et alii*, 2011 ; KIRCHNER, SWELLER, CLARK, 2006]. Différentes études montrent que la mise en œuvre de la démarche d'investigation en cours de sciences ne va pas de soi, ni pour les enseignants, ni pour les élèves qui peuvent rencontrer des difficultés à faire le lien entre les expériences et les théories. Il est donc nécessaire d'apporter à ces derniers une aide appropriée pour engendrer des effets positifs [EURYDICE, 2006].

En France, la définition du socle commun de connaissances et de compétences arrêtée en 2006 vise une plus grande cohérence des enseignements de sciences et de technologie. Il regroupe les principaux éléments de mathématiques et la culture scientifique et technologique dans l'un de ses sept piliers. En 2008, les programmes sont adaptés discipline par discipline, mais sont alors, dans ce cadre, soumis à des objectifs communs plus généraux, notamment de formation de l'esprit et de la personne en tant que citoyen. En ce qui concerne les sciences, une approche transdisciplinaire est clairement explicitée dans ces nouveaux programmes, et l'utilisation de la démarche d'investigation est préconisée, non seulement comme modalité pédagogique, mais également comme un objectif de formation en soi. Cependant, si le socle cherche à mettre en évidence la complémentarité des disciplines, la nature des interactions entre disciplines reste implicite, et la structure des programmes disciplinaires, en chaînes notionnelles, peut constituer un frein à la mise en place d'une approche interdisciplinaire [DELSERIEYS-PEDREGOSA, BOILEVIN *et alii*, 2010].

Dans ce contexte international, de revalorisation de l'intérêt pour les sciences, et national, de rénovation de l'enseignement, notamment des sciences et technologies,

l'Académie des sciences et l'Académie des technologies, en collaboration avec la direction générale de l'enseignement scolaire (DGESCO) du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, portent depuis 2005 le projet d'une expérimentation d'un enseignement intégré de science et technologie (EIST) lors des premières années de collège, en classe de sixième et de cinquième. Dans la lignée du projet « La main à la pâte » lancé en 1996 à l'école primaire, l'EIST consiste en un enseignement intégré des matières scientifiques et technologiques, par opposition au découpage disciplinaire traditionnel de l'enseignement dans le secondaire dans lequel différents enseignants interviennent en physique-chimie, en sciences de la vie et de la Terre (SVT) et en technologie. L'EIST se veut donc particulièrement adapté à la mise en place de la démarche d'investigation, et à un rôle actif des élèves dans le cadre des cours de science, tout en s'inscrivant dans le respect des programmes.

Présenté à travers dix grands principes, l'EIST affiche comme objectif « *l'appropriation progressive ou la consolidation par les élèves de concepts scientifiques et de techniques opératoires en même temps que l'amélioration de la maîtrise du langage et des qualités d'expression écrite et orale* » (voir rubrique EIST au collège, sur le site Internet de la fondation La main à la pâte). L'EIST est supposé « *accroître l'intérêt et l'autonomie* » des élèves, ainsi que favoriser « *une bonne acquisition des connaissances* », et « *une transition plus heureuse entre école et collège* ». Il suppose également « *un intense travail collectif et interdisciplinaire des professeurs* ».

L'EIST est expérimenté depuis la rentrée 2006, et a fait l'objet d'une évaluation, réalisée de 2008 à 2012 par la direction de l'évaluation de la prospective et de la performance (DEPP) du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. À l'aide d'évaluations répétées d'une cohorte d'élèves entrés en sixième en septembre 2008, et ce tout au long de leur scolarité au collège, nous avons cherché à mesurer l'impact de l'EIST sur les compétences en sciences des élèves, ainsi que sur leurs attitudes envers les sciences, par rapport à un enseignement disciplinaire traditionnel. Les premiers résultats, obtenus à partir des trois premiers temps de mesure réalisés entre 2008 et 2010, avaient fait l'objet d'un premier article [LE CAM et ROCHER, 2012]. À ce stade, l'analyse des progressions des élèves ne laissait apparaître aucun effet significatif du dispositif sur les performances des élèves.

Après un rappel sur les conditions de mise en œuvre de l'EIST et sur le dispositif d'évaluation, nous présenterons les résultats obtenus à partir des cinq temps de mesure réalisés entre le début de sixième à l'automne 2008 et la fin de troisième en juin 2012. Le caractère longitudinal des données permet d'étudier conjointement les progressions en sciences des élèves et l'évolution de leurs attitudes envers les sciences, tout au long de leur scolarité au collège. Nous nous sommes attachés à étudier l'existence d'un effet de l'EIST à la fois sur les domaines cognitifs et motivationnels.

LA MISE EN ŒUVRE DE L'EIST

Les conditions de mise en œuvre de l'EIST sont largement détaillées dans le rapport réalisé sur le sujet par l'inspection générale de l'Éducation nationale (IGEN) [PERROT, PIETRYK, ROJAT, 2009], ainsi que dans le guide de découverte réalisé par la fondation La main à la pâte [2011]. L'expérimentation de l'EIST a commencé au niveau sixième à partir de la rentrée 2006, et au niveau cinquième à partir de la rentrée 2007. À la rentrée 2008, 38 établissements volontaires expérimentaient le dispositif, au niveau sixième pour 35 d'entre eux et/ou cinquième pour 10 d'entre eux (7 établissements expérimentaient aux deux niveaux, 28 établissements uniquement au niveau sixième et 3 établissements uniquement au niveau cinquième). L'information étant passée par les recteurs, qui se sont appuyés sur les IA-IPR (inspecteurs d'académie – inspecteurs pédagogiques régionaux) des disciplines concernées pour solliciter les établissements, la communication auprès des établissements sur la mise en place du dispositif et la possibilité de se porter volontaire a été plus ou moins importante dans les différentes académies, selon l'intérêt porté par les recteurs à un tel dispositif. Les 38 établissements volontaires se situent dans 12 académies¹. Ces établissements s'engageaient alors à expérimenter l'EIST pendant quatre ans.

Le dispositif prévoit la constitution de trois groupes d'élèves à partir de deux classes, chaque groupe étant pris en charge pour l'enseignement de sciences et technologie par un unique professeur, de l'une des trois disciplines (physique-chimie, SVT, ou technologie). En sixième, chaque groupe d'élèves reçoit 3 h 30 d'enseignement des sciences et technologie par semaine, soit une demi-heure de plus que dans l'enseignement traditionnel dans lequel il n'y a pas de physique-chimie à ce niveau. Les enseignants concernés bénéficient d'une heure supplémentaire par semaine pour assurer le travail de concertation nécessaire à la préparation des cours, à partir de thèmes transversaux qui permettent d'intégrer les différents programmes disciplinaires. Pour deux classes de sixième, l'EIST représente un surcoût de 5 h 30 hebdomadaires par rapport à un enseignement traditionnel. En cinquième, ce surcoût est de 7 h 30 hebdomadaires. D'après le rapport de l'inspection générale, pour l'année 2006-2007, l'EIST n'était mis en place en sixième que pour un trimestre, mais au cours des années suivantes, il devait être mis en place sur l'année entière. D'après le cahier des charges que l'on peut trouver dans le guide de découverte de la fondation La main à la pâte, les établissements s'engageaient à assurer la mise en place de l'EIST « pendant au moins 30 % de l'année scolaire ». Cette souplesse accordée aux établissements rend délicate l'appréciation du temps effectif consacré à l'EIST au cours de l'année par chacun d'entre eux. De plus, l'organisation et la répartition des heures d'enseignement des sciences et technologie sont laissées à l'initiative des établissements, et sont donc aussi très variables d'un établissement à l'autre. À titre d'illustration, des éléments sur la mise en œuvre du dispositif dans les établissements ont été rendus disponibles grâce à un questionnaire envoyé en janvier 2009 aux chefs d'établissements et aux enseignants engagés dans l'EIST [BENHAIM-GROSSE, 2012]. Pour l'année 2008-2009, première année de l'évaluation du dispositif, l'expérimentation était prévue en classe de sixième pour toute l'année scolaire dans 32

1. Clermont-Ferrand, Nice, Toulouse, Montpellier, Poitiers, Bordeaux, Orléans-Tours, Nancy-Metz, Lille, Strasbourg, Versailles et Créteil.

établissements, et sur un semestre dans 3 établissements. Pour 80 % des groupes d'élèves de sixième participant, les 3 h 30 hebdomadaires d'EIST étaient réparties en deux séances par semaine. Ces séances, généralement de durée plus longue que dans l'organisation traditionnelle, permettent notamment de mettre en œuvre plus facilement la démarche d'investigation et des séquences de manipulations où les élèves sont actifs.

Les réponses au questionnaire des enseignants engagés dans l'EIST montrent que ces derniers se sont impliqués dans le dispositif principalement sur proposition des chefs d'établissement et/ou des IA-IPR, et assez peu, finalement, de leur propre souhait. Ils se répartissent selon le sexe de façon comparable au national. Ils sont un peu plus jeunes en physique-chimie et en SVT (avec une ancienneté moyenne de 12 et 13 ans respectivement), et sont comparables au niveau national en technologie (avec une ancienneté moyenne de 19 ans). Enfin, toujours suivant les réponses des chefs d'établissement au questionnaire, la sélection des élèves participant au dispositif en classe de sixième semble s'être effectuée de manière strictement aléatoire dans la moitié des établissements. Par ailleurs pour 7 collèges, essentiellement de petite taille, tous les élèves de sixième ont bénéficié de l'EIST. Enfin, pour 8 collèges la sélection s'est faite sur la base d'un ou plusieurs critères, et notamment le niveau des élèves pour 5 d'entre eux. Dans ce cas, les élèves en difficultés sont sous-représentés dans les classes d'EIST.

LE DISPOSITIF D'ÉVALUATION

L'évaluation d'expérimentations sociales n'est pas nouvelle, mais a connu un développement particulier en France suite à la diffusion des travaux d'Esther DUFLO en économie du développement [BANERJEE et DUFLO, 2009], et à la création du Fonds d'expérimentation pour la jeunesse (FEJ) en 2009. Celui-ci a largement contribué à diffuser et à promouvoir les méthodes d'évaluation d'expérimentation, et en particulier les méthodes d'évaluation randomisées (avec assignation aléatoire) qui sont particulièrement recommandées lorsqu'on souhaite mesurer, par une approche statistique, l'impact d'un dispositif expérimental [GURGAND et VALDENNAIRE, 2012 ; L'HORTY et PETIT, 2011] : si l'EIST apparaît comme un dispositif susceptible de favoriser l'acquisition des connaissances des élèves en sciences et leur intérêt pour ces disciplines, il est nécessaire de mettre à l'épreuve l'existence d'un réel effet du dispositif, et de s'assurer qu'il est bien la cause des différences observées chez les élèves. En effet, le seul jugement des acteurs impliqués (enseignants, IA-IPR, etc.) ne peut suffire à conclure, l'enjeu étant d'identifier si un tel dispositif peut et doit être plus largement diffusé, ou encouragé.

Le principe de l'expérimentation avec assignation aléatoire est de comparer la situation des individus ayant bénéficié du dispositif expérimental avec la situation qu'ils auraient connue s'ils n'en avaient pas bénéficié. L'écart entre les deux situations correspond alors à l'effet du dispositif expérimental. Comme on ne peut observer directement ce que serait la situation des bénéficiaires en l'absence d'intervention, on observe la situation d'un groupe d'individus non bénéficiaires, semblables aux

bénéficiaires, le groupe témoin. La validité de la comparaison repose notamment sur l'assignation aléatoire des individus dans chacun des deux groupes : les caractéristiques initiales des individus dans les deux groupes sont ainsi équivalentes en moyenne et toutes explications alternatives d'un potentiel effet sont ainsi écartées. La mise en œuvre de l'EIST s'est faite sur la base du volontariat des établissements, il ne s'agit donc pas d'une assignation aléatoire. Dans ce cas, la différence observée entre les élèves du groupe expérimental et ceux du groupe témoin ne correspond plus exactement à l'effet de l'expérimentation, mais peut être également composée d'un biais de sélection [GIVORD, 2010].

L'échantillon

Afin d'identifier l'existence d'un effet de l'EIST sur les connaissances et compétences en sciences des élèves, ainsi que sur leurs attitudes envers les sciences, la DEPP a mis en place un dispositif d'évaluation des élèves, à partir de l'année scolaire 2008-2009 et sur quatre années consécutives. L'objectif est de suivre une cohorte d'élèves, composée d'une part des élèves bénéficiant de l'EIST dans les 35 établissements expérimentateurs au niveau sixième en France métropolitaine en 2008-2009, et d'autre part d'un groupe témoin d'élèves de sixième suivant un enseignement de sciences et de technologie traditionnel dans des établissements non expérimentateurs. L'échantillon témoin a été sélectionné dans les 12 académies concernées par l'expérimentation de l'EIST (hors établissements expérimentateurs). Il est stratifié en respectant les proportions d'élèves issus d'établissements publics en éducation prioritaire, d'établissements publics hors éducation prioritaire, et d'établissements privés, observées au niveau national. Les différences de structure entre les deux groupes seront contrôlées lors des analyses. Afin qu'il soit de même taille, en nombre d'élèves, que le groupe expérimentateur, 81 établissements ont été sélectionnés, dans chacun desquels une classe de sixième a ensuite été sélectionnée par tirage aléatoire simple pour participer au dispositif d'évaluation.

Dans l'optique de pouvoir contrôler un éventuel effet de sélection, lié au volontariat des établissements engagés dans le dispositif expérimental, un échantillon d'élèves recevant un enseignement traditionnel mais scolarisés dans les établissements engagés dans l'EIST a également été sélectionné pour participer au dispositif d'évaluation. En pratique, dans chaque établissement expérimentateur, une classe entière de sixième ne pratiquant pas l'EIST en 2008-2009 a été échantillonnée par tirage aléatoire simple. La cohorte d'élèves participant au dispositif d'évaluation comportait donc trois groupes d'élèves à l'origine : celui des élèves qui ont reçu un enseignement EIST en sixième, le groupe témoin et enfin un groupe d'élèves scolarisés dans les établissements expérimentateurs, mais recevant un enseignement traditionnel. Ce dernier groupe étant fortement affecté dès la première session par une non-réponse de niveau établissement, a de ce fait un effectif particulièrement faible et ne pourra plus remplir sa fonction de groupe contrôle. Il n'est donc pas utilisé dans les analyses présentées ici.

Tous les élèves de la cohorte ainsi constituée ont été évalués à l'automne 2008, en début de sixième, avant leur participation à l'expérimentation. Ils ont ensuite été suivis tout au long de leur scolarité au collège et ont été à nouveau évalués à chaque fin d'année scolaire, et ce jusqu'en troisième, ce qui représente cinq moments d'évaluation au total. Les élèves redoublants n'ont pas été suivis, de même que les

élèves ayant fait l'objet d'un passage anticipé, et ceux ayant changé d'établissement en cours de scolarité.

Les instruments d'évaluation

Chaque moment de mesure consiste en une évaluation des élèves de l'échantillon, sous la forme d'un test papier-crayon standardisé. Chaque test, d'une durée totale de deux fois une heure, est composé de trois séquences. Les séquences une et trois sont des séquences évaluant les compétences et connaissances des élèves en sciences (dimensions cognitives), tandis que la séquence deux évalue les attitudes des élèves envers les sciences (dimensions conatives). Afin de respecter le cadre des programmes officiels en vigueur pour le niveau concerné et celui du socle commun, à chaque temps de mesure, les épreuves cognitives sont différentes. Ces dernières ont été conçues par des enseignants des trois disciplines (physique-chimie, SVT et technologie), dont certains venaient d'établissements engagés dans l'expérimentation de l'EIST, et d'autres pas.

La DEPP a été contactée pendant l'été 2008 pour débiter l'évaluation dès la rentrée de septembre 2008, ce qui imposait des délais très courts pour mettre en place les deux premiers temps de mesure, en début et en fin de sixième : les deux premières évaluations sont donc composées uniquement de questions à choix multiples (QCM) [LE CAM et ROCHER, 2012]. Les trois suivantes intègrent à la fois des QCM et des questions à réponses construites, qui supposent une réponse rédigée de la part des élèves.

La séquence deux est identique à chaque temps de mesure, et comporte 28 questions (aussi appelées items) réparties en six grands thèmes : « 1- motivation pour les sciences », « 2- activités en relation avec les sciences », « 3- les sciences dans l'avenir et le futur métier », « 4- sensibilisation aux phénomènes environnementaux », « 5- sentiment d'efficacité en sciences », et « 6- qu'est-ce que faire des sciences ? ».

ANALYSE DES DONNÉES

Participation et comparaison des groupes

Le **tableau 1 p. 290** présente la participation des établissements et des élèves aux cinq moments d'évaluation. Sur les 116 établissements visés au départ (35 expérimentateurs et 81 témoins), 102 ont participé à tous les moments d'évaluation (34 expérimentateurs et 68 témoins). Sur les 4 012 élèves visés au départ, 2 004 ont finalement participé aux cinq temps de mesure, soit 50 %. Ce taux de participation est un peu plus élevé dans le groupe EIST (53 %) que dans le groupe témoin (47 %).

Le groupe témoin présente un taux de non-réponse au niveau établissement plus important que le groupe EIST : à chaque session, entre 2 et 5 établissements n'ont pas participé, amenant à 68 le nombre d'établissements ayant effectivement participé aux 5 temps de mesure. Au-delà de cette non-réponse au niveau établissement, l'attrition

► **Tableau 1** Participation des établissements et des élèves aux cinq moments d'évaluation

		EIST			Témoins	
		Établissements EIST	Élèves EIST	Élèves (en %)	Établissements témoins	Élèves témoins
2008-2009	Nombre visé	35	2 035	100,0 %	81	1 977
Session 1 : début de 6 ^e	Nombre testé	35	1 961		79	1 845
Session 2 : fin de 6 ^e	Nombre testé	35	1 903		81	1 870
Participation sessions 1 et 2		35	1 853	91,1 %	79	1 755
2009-2010	Nombre visé	35	1 798		81	1 750
Session 3 : fin de 5 ^e	Nombre testé	34	1 647		76	1 547
Participation sessions 1, 2 et 3		34	1 542	75,8 %	74	1 402
2010-2011	Nombre visé	35	1 629		81	1 584
Session 4 : fin de 4 ^e	Nombre testé	35	1 435		78	1 361
Participation sessions 1, 2, 3 et 4		34	1 280	62,9 %	70	1 132
2011-2012	Nombre visé	35	1 460		81	1 401
Session 5 : fin de 3 ^e	Nombre testé	35	1 289		78	1 239
Participation sessions 1, 2, 3, 4 et 5		34	1 068	52,5 %	68	936

observée à partir du troisième temps de mesure est due notamment aux élèves ayant quitté leur établissement et aux élèves redoublants ou ayant fait l'objet d'un passage anticipé, qui ne sont plus suivis et sortent de la cohorte. À cette perte s'ajoute de la non-réponse au niveau élève, la part de chaque cause dans la perte totale est difficile à évaluer, car ces informations ne sont transmises par les établissements que de façon très partielle.

La principale différence initiale entre les deux groupes concerne la proportion d'élèves issus d'établissements privés qui s'élève à 4 % dans le groupe EIST contre 23 % dans le groupe témoin ► **Tableau 2**. La comparaison des caractéristiques des élèves ayant participé à la première session, et des élèves ayant participé aux trois

► **Tableau 2** Caractéristiques des groupes de répondants

	Répondants à la première session (N=3 840)		Répondants aux cinq sessions (N=1 894)	
	EIST	Témoins	EIST	Témoins
N	1 961	1 845	1 068	936
% d'élèves du secteur privé	4,04	22,89	3,89	23,21
% d'élèves en éducation prioritaire	14,00	14,29	12,06	9,47
% de filles	48,03	48,96	50,10	53,12
% d'élèves en retard	18,96	20,55	11,19	12,93
Indice de position sociale ¹ moyen du père	- 0,43	- 0,52	- 0,35	- 0,46
Indice de position sociale ¹ moyen de la mère	- 0,28	- 0,39	- 0,18	- 0,31

1. L'indice de position sociale est calculé à partir des professions et catégories socio-professionnelles [LE DONNÉ et ROCHER, 2010].

Lecture : les filles représentent 48 % des 1 961 élèves du groupe EIST ayant participé à la première session, et 50 % des 1 068 élèves du même groupe ayant participé aux cinq sessions.

Source : MENESR-DEPP.

Élèves (en %)	Ensemble		
	Établissements	Élèves	Élèves (en %)
100,0 %	116	4 012	100,0 %
	114	3 806	
	116	3 773	
88,8 %	114	3 608	89,9 %
	116	3 548	
	110	3 194	
70,9 %	108	2 944	73,4 %
	116	3 213	
	113	2 796	
57,3 %	104	2 412	60,1 %
	116	2 861	
	113	2 528	
47,3 %	102	2 004	50,0 %

Lecture : le suivi de cohorte visait 2 035 élèves de sixième participant au dispositif expérimental EIST à la rentrée 2008, dans 35 établissements. Parmi eux, 1 961 ont participé à l'épreuve du début de sixième, 1 903 ont participé à celle de fin de sixième, et 1 853 ont participé aux deux, ce qui représente 91 % des élèves visés au départ. Le nombre d'élèves visés dès la seconde année n'est pas celui de 2008-2009, car il ne prend pas en compte les élèves sortis de la cohorte (passage anticipé, redoublement, etc.).

Source : MENESR-DEPP.

premières sessions, avait montré que l'attrition ne présentait pas jusque-là de caractère différentiel entre le groupe EIST et le groupe témoin [LE CAM et ROCHER, 2012]. Ces analyses ont été complétées en ajoutant les deux derniers moments de mesure. L'attrition concerne un peu plus les garçons que les filles, en particulier dans le groupe témoin, et surtout les élèves qui présentaient déjà un retard scolaire à l'entrée en sixième. Ces caractéristiques seront cependant contrôlées dans l'analyse de l'impact de l'EIST sur les performances et les attitudes des élèves.

Analyse des items et construction des scores

Les évaluations passées par les élèves doivent permettre de mesurer leur performance en sciences, ainsi que leurs attitudes envers les sciences, à chaque moment de mesure. Le caractère longitudinal des données permet alors d'étudier la progression des acquisitions des élèves en sciences tout au long de leur scolarité au collège, de même que l'évolution de leurs attitudes envers les sciences.

Concernant la partie cognitive des évaluations, les épreuves étaient différentes à chaque session, puisqu'elles ont été conçues dans le respect des programmes du niveau considéré chaque année. Le nombre d'items cognitifs passés par les élèves varie de 44 items lors de la session 1, à 72 items lors des sessions 4 et 5. À chaque temps de mesure, un certain nombre d'items de la session précédente sont repris (de 8 à 18 items selon les sessions). Ils sont utilisés comme ancrage afin de construire des scores comparables d'un temps de mesure à l'autre. Une analyse des caractéristiques psychométriques des items a été réalisée dans un premier temps, afin d'apprécier la qualité des items. Une première sélection des items a été réalisée à l'aide du Rbis, indice mesurant la discrimination de l'item. Les items faiblement discriminants à une ou plusieurs évaluations ont été écartés des analyses ultérieures. La fidélité de chaque épreuve a été mesurée par l'alpha de Cronbach. Ce dernier varie de 0,85 à 0,89 selon les sessions d'évaluation, indiquant

une bonne cohérence interne de l'épreuve à chaque session. La dimensionnalité des épreuves a été étudiée grâce à des analyses factorielles. Ces analyses ont laissé apparaître une large dimension dominante, correspondant à la compétence en sciences testée par chaque épreuve. Enfin, une analyse a été réalisée concernant d'éventuels fonctionnements différentiels d'items (FDI). Un FDI peut être défini comme une probabilité de réussite différente, pour deux élèves, liée à une variable autre que la compétence de l'élève. À niveau de compétence égal, la réussite différente à certains items entre groupe témoin et groupe expérimental, ou entre deux années consécutives pour les items d'ancrage, ne doit pas perturber la comparabilité des épreuves. Les items présentant un FDI ont donc été écartés des analyses. Le nombre d'items sélectionnés pour le calcul des scores est présenté dans le **tableau 3**.

► **Tableau 3** Nombre d'items sélectionnés pour le calcul des scores cognitifs

Session	Nombre total	Communs avec la session précédente	Mauvaise discrimination	FDI	Nombre final	Communs avec la session précédente
1	44		6	3	35	
2	62	10	15	9	38	7
3	54	8	1	12	41	5
4	72	18	5	13	54	8
5	72	17	2	4	66	13

Lecture : parmi les 72 items de l'épreuve de sciences du cinquième moment de mesure, 17 étaient repris de la session précédente. Sur ces 72 items, 2 ont été écartés en raison d'un indice de discrimination (ou corrélation item-test) trop faible, 4 ont été écartés car ils présentaient un fonctionnement différentiel (FDI). Au final, le score de la session 5 est calculé sur 66 items, dont 13 communs avec la session précédente.

Source : MENESR-DEPP.

Une fois les items sélectionnés, un score, correspondant au niveau de compétence de l'élève, a été calculé à chaque moment de mesure en utilisant des modèles de réponse à l'item à deux paramètres. Cette technique de modélisation permet, grâce aux items d'ancrage, de mettre sur une même échelle et rendre comparable les scores d'une année sur l'autre, ce qui est indispensable pour étudier les progressions des élèves en sciences tout au long de leur scolarité au collège². Pour mettre sur une même échelle les scores des cinq moments de mesure, les paramètres sont estimés ensemble sur les trois premières sessions d'une part et sur les deux dernières d'autre part. Les items communs aux sessions 3 et 4, permettent de relier ces deux échelles par la méthode de STOCKING et LORD [1983]. Les scores sont standardisés, de moyenne 0 et d'écart-type 1, sur les résultats des élèves du groupe témoin à la première session.

Concernant la partie conative des évaluations, le questionnaire est identique à chaque session. Les thèmes 4, sur les questions environnementales, et 6, qui concerne les représentations des élèves sur ce qu'est « faire des sciences », ont été écartés, car ces items se prêtent mal aux analyses engagées ici ► **Encadré**. Sur les 18 items restants, sur la base d'analyses factorielles exploratoires et confirmatoires,

2. Voir Le CAM et ROCHER [2012], ainsi que l'article de ROCHER, dans ce numéro, p. 37, pour plus de détails sur les modèles de réponse à l'item, et les techniques d'*equating*.

13 sont retenus, organisés en quatre facteurs. L'invariance du construit au cours du temps n'étant satisfaisante que pour les facteurs « 1- motivation intrinsèque pour les sciences », et « 2- intérêt pour les sciences en dehors de l'école », seules ces deux dimensions ont été considérées dans les analyses.

CONSTRUCTION DES SCORES CONATIFS

Le questionnaire intitulé « attitudes envers les sciences » compte 6 sections comprenant chacune entre 4 et 5 items, soit au départ 28 items. Chaque section se rapporte à un grand thème :

- 1 – Motivation pour les sciences
- 2 – Activités en relation avec les sciences
- 3 – Les sciences dans l'avenir et le futur métier
- 4 – Sensibilisation aux phénomènes environnementaux
- 5 – Sentiment d'efficacité en sciences
- 6 – Qu'est-ce que faire des sciences ?

La section 4 qui comprend 5 items portant sur les questions environnementales a été écartée. En effet, les items s'intéressent davantage aux connaissances des élèves relatives aux thématiques environnementales qu'aux aspects affectivo-motivationnels. De la même manière, la section 6, qui concerne les représentations des élèves sur ce qu'est « faire des sciences » permet difficilement de faire une analyse globale, et se prête davantage à l'analyse individuelle et descriptive de l'évolution de chaque item. Les analyses portent donc sur 4 sections, soit 18 items.

Les élèves devaient répondre selon une échelle de Likert en quatre points, allant de « tout à fait d'accord » à « pas du tout d'accord », ou de « très souvent » à « jamais » pour les items de la section 2 sur les activités en relation avec les sciences.

L'échantillon des répondants au premier temps de mesure a été subdivisé aléatoirement en deux sous-échantillons de même taille. Le premier sous-échantillon est considéré comme l'échantillon de *calibration* et le second comme l'échantillon de *réplication*.

L'échantillon de *calibration* est utilisé dans un premier temps pour réaliser une analyse factorielle confirmatoire, afin de tester la dimensionnalité

des 18 items des quatre sections sélectionnées en 4 facteurs distincts. Les indicateurs d'ajustement ne sont pas complètement satisfaisants, et quatre items présentent une variance résiduelle élevée, ce qui indique qu'ils représentent mal la dimension à laquelle ils étaient associés dans le questionnaire.

Toujours en s'appuyant sur l'échantillon de *calibration*, une analyse factorielle exploratoire est alors mise en œuvre, sur les 18 items des quatre sections sélectionnées. Chaque nouvelle dimension ainsi identifiée est par ailleurs testée, en utilisant un modèle distinct (analyse confirmatoire congénérique) faisant l'hypothèse d'un seul facteur expliquant la variance commune à chacun des groupes d'items correspondants. Finalement, une structure factorielle en 4 facteurs est identifiée à partir de 13 items.

La structure factorielle trouvée est ensuite testée sur l'échantillon de *réplication*, à l'aide d'une nouvelle analyse factorielle confirmatoire. Les indicateurs d'ajustement étant satisfaisants, cette structure factorielle est celle qui est retenue.

Pour finir, nous avons interrogé la stabilité de ces construits au cours du temps, c'est-à-dire s'ils étaient bien mesurés de manière identique sur les cinq moments de mesure. Nous avons mis en œuvre une série d'analyses factorielles confirmatoires avec autocorrélation des erreurs (la réponse à une variable au temps T est corrélée avec la réponse à la même variable au temps T+1). Dans un premier temps, nous avons testé une invariance configurale, qui permet de s'assurer que le modèle de base pour chacun des construits est potentiellement spécifiable aux cinq moments de mesure (les mêmes items saturent bien sur un seul facteur). À ces cinq moments, on a autorisé l'estimation libre des moyennes et des variances d'items résiduelles. Les indices d'ajustement sont satisfaisants pour les quatre facteurs.

Le second temps a consisté à mettre en œuvre et tester un modèle d'invariance métrique (invariance faible) qui ajoute au modèle précédent la contrainte que les saturations soient égales à chaque moment de mesure. L'ajustement du modèle reste satisfaisant pour les facteurs 1 (motivation intrinsèque pour les sciences) et 2 (intérêt pour les sciences en dehors de l'école). Autrement dit, on peut avancer que pour ces deux facteurs, au cours du temps, la structure unidimensionnelle et les saturations des items correspondant à chaque construit restent identiques. Nous ne prendrons pas le risque d'avancer ce constat pour les deux autres facteurs.

Le score de motivation intrinsèque pour les sciences est calculé pour chaque élève, à chaque temps de mesure, en faisant la somme des trois items retenus pour ce facteur (« Je participe en science parce que j'aime bien chercher » ; « Je travaille en science parce que j'aime bien cette discipline » ; « Ce que je fais en science est intéressant »).

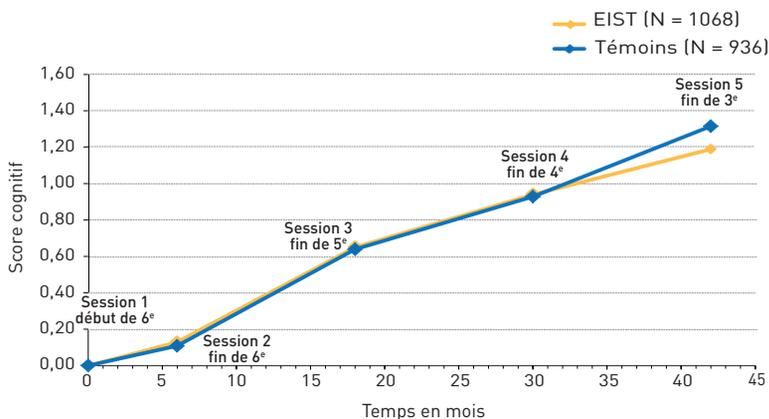
De même, le score d'intérêt pour les sciences en dehors de l'école est calculé à partir des trois items retenus pour ce facteur (« Je regarde des émissions scientifiques à la télévision » ; « J'aime lire des livres ou des revues de science » ; « Je cherche des documents scientifiques sur Internet »).

RÉSULTATS

Progression des performances en sciences

Le niveau moyen de performance en sciences des élèves connaît une augmentation régulière au cours du temps (mesuré en mois) depuis le début de sixième (novembre 2008 : temps = 0) jusqu'en fin de troisième (mai 2012 : temps = 42). En moyenne, le niveau initial et la progression des élèves du groupe EIST sont très proches de ceux des élèves du groupe témoin, sauf en fin de troisième où le score moyen des élèves du groupe EIST est inférieur à celui des élèves du groupe témoin ▶ **Figure 1**. Mais ces moyennes brutes ne tiennent pas compte des différences de structures entre les deux échantillons, qui peuvent être corrélées à la performance des élèves.

▶ **Figure 1** Évolution du score cognitif moyen selon le groupe (EIST vs témoin)



Note de lecture : le score cognitif (en ordonnées) est standardisé, de moyenne 0 et d'écart-type 1 pour le groupe témoin à la première session. Le temps (en abscisses) est exprimé en mois à partir du premier temps de mesure.

Source : MENESR-DEPP.

Les données dont on dispose ici présentent une structure hiérarchisée : les élèves sont regroupés dans des établissements, dont les caractéristiques sont susceptibles d'avoir une influence sur leurs acquisitions scolaires, notamment en sciences. Les modèles multiniveaux, adaptés à ce type de données emboîtées, permettent de prendre en compte les différents niveaux dans l'analyse et de mieux estimer les erreurs standards des coefficients, et donc leur précision. Le fait d'avoir suivi les mêmes individus pendant plusieurs années, et de disposer de données longitudinales, apporte une information supplémentaire sur l'hétérogénéité intra-individuelle. Comme nous disposons de mesures répétées pour chaque élève de notre échantillon, il est possible d'étudier l'effet de l'EIST, non plus seulement sur le niveau des élèves en sciences à un moment donné, mais sur le taux de croissance, la progression, des performances en sciences des élèves durant la scolarité au collège. On le comprend, dans le cadre de cette évaluation du dispositif EIST, nous nous attendons à ce que le taux de croissance du score en sciences soit significativement plus important pour le groupe d'élèves ayant bénéficié du dispositif. Nous avons donc utilisé des modèles multiniveaux de croissance, en considérant les trois niveaux imbriqués suivants : temps de mesure > élève > établissement. Ces types de modèles, et leurs applications possibles dans le champ de l'éducation, sont présentés dans l'ouvrage de BRESSOUX [2010].

Du fait de l'absence de randomisation, l'étude de l'évolution du score moyen des élèves du groupe EIST et des élèves du groupe témoin a été réalisée en contrôlant les caractéristiques des élèves ainsi que celles des établissements, afin de corriger les effets de structure. Les variables introduites dans les modèles sont, en plus de la participation au dispositif EIST, le sexe, le retard scolaire à l'entrée en sixième, l'indice de position sociale de chacun des parents, calculé à partir des professions et catégories socioprofessionnelles (PCS) [LE DONNÉ et ROCHER, 2010], les deux scores conatifs, de motivation intrinsèque pour les sciences et d'intérêt pour les sciences en dehors de l'école, pour ce qui est du niveau élève, et le secteur (public/privé) ainsi que l'appartenance à l'éducation prioritaire, pour le niveau établissement.

Le **tableau 4 p. 296** présente les résultats d'une série de modèles multiniveaux dans lesquels la variable expliquée est le score cognitif. Pour les modèles 1 à 4 ces résultats ont été obtenus à partir des données de 1 894 élèves ayant participé aux cinq sessions d'évaluation et pour lesquels il a été possible de construire un score cognitif à chaque session. Le modèle 5 introduit plusieurs variables de contrôle, et a été calculé sur les 1 883 élèves pour lesquels toutes ces variables sont renseignées.

Le **modèle 1** est un modèle inconditionnel classique, c'est-à-dire sans aucune variable explicative. Il permet de décomposer la variance de la variable expliquée, le score cognitif, entre les différents niveaux et indique que 13 % de la variance se situe entre les établissements³, 40 % entre les élèves, et 47 % au niveau intra-individuel ce qui n'est pas étonnant puisqu'on se situe sur une période assez longue, presque quatre ans, au cours de laquelle le niveau de performance de chaque élève évolue beaucoup. Le score moyen varie plus d'un élève à l'autre que d'un établissement à l'autre, il est

3. Variance inter-établissements = $0,23 / (0,23 + 0,71 + 0,85) = 0,13$, variance inter-élèves = $0,71 / (0,23 + 0,71 + 0,85) = 0,40$, variance intra-élève = $0,85 / (0,23 + 0,71 + 0,85) = 0,47$

► **Tableau 4** Évaluation de l'effet de l'EIST sur les performances en science - Modèles multiniveaux de croissance

Paramètres	Modèle 1		Modèle 2	
Nombre de collèges	101		101	
Nombre d'élèves	1 894		1 894	
Nombre d'observations	9 470		9 470	
Effets fixes				
Constante	0,54	(0,05) ***	- 0,05	(0,06)
Temps			0,03	(0,001)***
EIST				
EIST*temps				
Fille				
Retard				
Indice de position sociale du père				
Indice de position sociale de la mère				
Établissement privé				
Éducation prioritaire				
Motivation intrinsèque pour les sciences				
Intérêt pour les sciences en dehors de l'école				
Motivation*temps				
Intérêt*temps				
Effets aléatoires				
Niveau 3 (inter-collèges)				
Variance des constantes	0,23	(0,04) ***	0,23	(0,04) ***
Variance des pentes				
Niveau 2 (inter-élèves)				
Variance des constantes	0,71	(0,03) ***	0,76	(0,03) ***
Variance des pentes				
Covariance constantes/pentes				
Niveau 1 (intra-élève)				
Variance des constantes	0,85	(0,01) ***	0,57	(0,01) ***
- 2LogV	28 566,74		25 557,24	
AIC	28 574,74		25 567,24	
BIC	28 585,20		25 580,32	

*** significatif au seuil de 1%, ** significatif au seuil de 5 %, * significatif au seuil de 10 %.

Lecture : le tableau présente les résultats des différents modèles multiniveaux (valeurs des paramètres, écart-type entre parenthèses et significativité indiquée par des *). Dans la partie « effets fixes » le coefficient associé au *temps* est positif et significatif : un mois supplémentaire de scolarité augmente le score en sciences de 0,02 écart-type (modèle 5).

Le coefficient de la variable indicatrice *EIST* (1 pour EIST, 0 pour témoin) indique la différence de score entre les élèves EIST et les élèves témoins au premier temps de mesure. Il n'est pas significatif. Le coefficient de l'interaction entre l'indicatrice EIST et le temps n'est pas non plus significatif : les progressions ne sont pas différentes selon que les élèves ont bénéficié ou non de l'EIST en sixième.

Source : MENESR-DEPP.

donc particulièrement important de prendre en compte les caractéristiques au niveau élève. Le **modèle 2** est un modèle inconditionnel de croissance, dans lequel la seule variable explicative est le temps (mesuré en mois à partir du premier temps de mesure). La prise en compte du temps fait diminuer la variance intra-élève de 33 % par rapport au modèle 1. La constante du modèle correspond au niveau initial moyen des élèves, au premier temps de mesure. Le coefficient de la variable temps correspond à la pente, qui représente l'évolution du niveau moyen pour une unité de temps supplémentaire (ici,

Modèle 3		Modèle 4		Modèle 5	
101		101		101	
1 894		1 894		1 883	
9 470		9 470		9 415	
- 0,05	(0,05)	- 0,06	(0,06)	- 0,07	(0,10)
0,03	(0,001) ***	0,03	(0,001) ***	0,02	(0,003) ***
		0,02	(0,10)	- 0,04	(0,08)
		- 0,001	(0,002)	- 0,001	(0,002)
				0,03	(0,04)
				- 0,83	(0,06) ***
				0,20	(0,04) ***
				0,10	(0,03) ***
				- 0,01	(0,11)
				- 0,26	(0,11) *
				0,02	(0,01) *
				0,02	(0,01) *
				0,002	(0,0003) ***
				0,0001	(0,0003)
0,17	(0,03) ***	0,17	(0,03) ***	0,10	(0,02) ***
0,0001	(0,00002) ***	0,0001	(0,00002) ***	0,0001	(0,00001) ***
0,60	(0,03) ***	0,60	(0,03) ***	0,51	(0,03) ***
0,0001	(0,00002) ***	0,0001	(0,00002) ***	0,0001	(0,00002) ***
0,004	(0,001) ***	0,004	(0,001) ***	0,002	(0,001) ***
0,53	(0,01) ***	0,53	(0,01) ***	0,52	(0,01) ***
25 271,31		25 270,80		24 666,86	
25 289,31		25 292,80		24 708,86	
25 312,84		25 321,57		24 763,78	

un mois). Dans le **modèle 3**, on autorise la pente à varier d'un établissement à l'autre, et d'un élève à l'autre. Les variances des pentes indiquent que l'évolution du score peut être significativement différente d'un élève à l'autre, et d'un établissement à l'autre. Le **modèle 4** intègre en plus l'indicatrice d'appartenance au groupe EIST, ainsi que son interaction avec le temps. Le coefficient de la variable EIST correspond à l'écart de niveau moyen entre les élèves EIST et les élèves témoins au premier temps de mesure (lorsque temps = 0), et n'est pas significativement différent de zéro. Le coefficient de l'interaction de l'EIST avec le temps (- 0,001, non significatif) nous intéresse particulièrement, puisqu'il correspond à la différence d'évolution du niveau moyen de performance en sciences des élèves EIST par rapport aux élèves témoins. Ce coefficient n'est pas significatif non plus, indiquant que le taux de croissance du score en sciences des élèves ayant bénéficié de l'EIST n'est pas différent de celui des élèves témoins. Le développement des compétences en sciences est similaire dans les deux groupes. Les indices d'ajustement montrent que ce modèle n'est d'ailleurs pas mieux ajusté que le précédent, en d'autres

termes ces variables explicatives n'apportent aucune information supplémentaire. Dans le **modèle 5**, on ajoute les caractéristiques aux niveaux élèves et établissements, ainsi que les deux scores conatifs, et leurs interactions avec le temps. On ne trouve pas d'effet, en moyenne, de l'EIST sur l'évolution des performances des élèves en sciences. Les caractéristiques qui expliquent les différences de développement des performances en sciences sont le retard scolaire et les positions sociales des parents, ainsi que l'évolution de la motivation intrinsèque pour les sciences (en moyenne, sur un mois de scolarité, une augmentation d'un point de score de motivation correspond à une augmentation de 0,17 % d'écart-type du score cognitif, toutes choses égales par ailleurs).

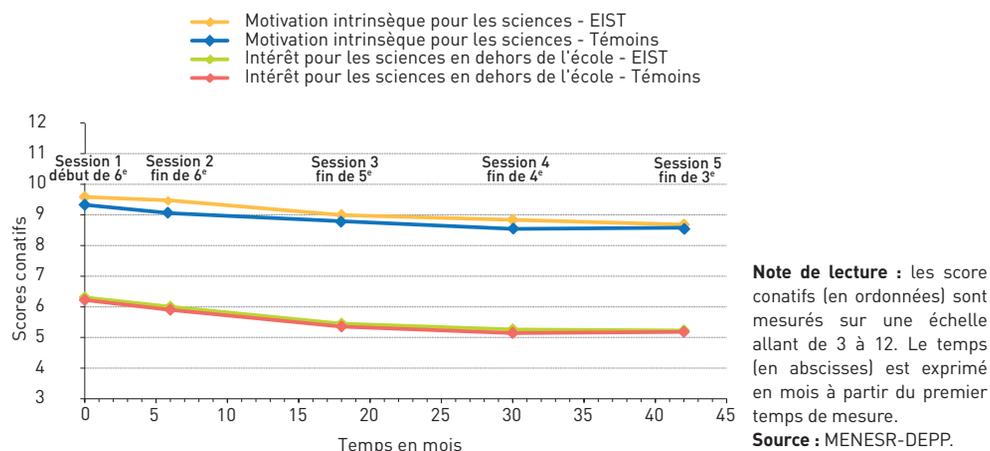
Évolution de la motivation pour les sciences, et de l'intérêt pour les sciences en dehors de l'école

Sur la base du questionnaire conatif passé à chaque temps de mesure par les élèves, nous avons identifié deux dimensions pour lesquelles la qualité de l'ajustement des données au cours du temps est satisfaisante. Nous avons donc calculé deux scores conatifs : un score mesurant la motivation intrinsèque des élèves pour les sciences, et un score mesurant l'intérêt des élèves pour les sciences en dehors de l'école. Dans la mesure où une invariance métrique a été mise en évidence au cours du temps, chacun de ces deux scores est calculé à chaque temps de mesure en sommant les réponses des élèves aux trois items retenus pour chacune des deux dimensions. Les réponses des élèves aux items étant codées de 1 à 4, ces scores conatifs peuvent varier, à chaque temps de mesure, de 3 à 12.

Les données recueillies mettent en évidence une baisse générale de l'intérêt et de la motivation pour les sciences, depuis le début de sixième et au moins jusqu'en fin de cinquième, et cela quel que soit le groupe considéré. Les élèves du groupe EIST ont en moyenne un score un peu plus élevé au départ que les élèves du groupe témoin, en particulier en termes de motivation pour les sciences. Cet écart reste stable jusqu'en fin de quatrième, et semble se réduire au cours de la troisième ► **Figure 2**.

Afin d'apprécier l'effet de l'EIST sur l'évolution des scores de motivation intrinsèque pour

► **Figure 2** Évolution des scores conatifs moyens selon le groupe (EIST vs témoin)



les sciences, et d'intérêt pour les sciences en dehors de l'école, nous avons également utilisé une modélisation multiniveaux de croissance. Seul le modèle final pour chacun de ces deux scores est reporté dans le **tableau 5**, il inclut comme variables explicatives le temps, l'indicatrice *EIST* et son interaction avec le temps, les caractéristiques au niveau établissements (secteur privé et éducation prioritaire) et au niveau élève (sexe, retard scolaire à l'entrée en sixième, indice de position sociale de chacun des parents), ainsi que le score cognitif et le deuxième score conatif à chacun des temps de mesure.

► **Tableau 5** Évaluation de l'effet de l'EIST sur les scores conatifs

Paramètres	Motivation intrinsèque pour les sciences			Intérêt pour les sciences en dehors de l'école		
Nombre de collèges	101			101		
Nombre d'élèves	1 883			1 883		
Nombre d'observations	9 415			9 415		
Effets fixes						
Constante	7,57	(0,11)	***	2,90	(0,15)	***
Temps	- 0,04	(0,004)	***	- 0,01	(0,01)	**
EIST	0,25	(0,1)	**	0,02	(0,08)	
EIST* temps	- 0,003	(0,004)		0,001	(0,002)	
Fille	- 0,06	(0,05)		0,36	(0,06)	***
Retard	- 0,12	(0,08)		- 0,08	(0,09)	
Indice de position sociale du père	0,06	(0,04)		0,03	(0,05)	
Indice de position sociale de la mère	0,00	(0,04)		- 0,13	(0,05)	***
Établissement privé	- 0,07	(0,12)		0,13	(0,10)	
Éducation prioritaire	0,20	(0,12)		- 0,03	(0,10)	
Score cognitif	0,10	(0,03)	***	0,03	(0,03)	
Score cognitif* temps	0,005	(0,001)	***	0,002	(0,001)	***
Intérêt pour les sciences en dehors de l'école	0,29	(0,01)	***			
Intérêt* temps	0,003	(0,003)	***			
Motivation intrinsèque pour les sciences				0,32	(0,01)	***
Motivation* temps				- 0,002	(0,001)	***
Effets aléatoires						
Niveau 3 (inter-collèges)						
Variance des constantes	0,12	(0,03)	***	0,02	(0,02)	
Variance des pentes	0,0002	(0,00004)	***	0,00003	(0,00002)	*
Niveau 2 (inter-élèves)						
Variance des constantes	0,79	(0,06)	***	1,58	(0,08)	***
Variance des pentes	0,001	(0,0001)	***	0,001	(0,0001)	***
Covariance constantes/pentes	- 0,01	(0,002)	***	- 0,02	(0,002)	***
Niveau 1 (intra-élève)						
Variance des constantes	1,92	(0,04)	***	1,51	(0,03)	***

*** significatif au seuil de 1 %, ** significatif au seuil de 5 %, * significatif au seuil de 10 %

Lecture : le tableau présente les résultats des différents modèles multiniveaux (valeurs des paramètres, écart-type entre parenthèses et significativité indiquée par des *). Concernant le score de motivation intrinsèque, dans la partie « effets fixes » le coefficient associé au *temps* est négatif et significatif : un mois supplémentaire de scolarité correspond à une baisse du score de motivation de 0,04 écart-type. Le coefficient de la variable indicatrice *EIST* (1 pour EIST, 0 pour témoin) indique la différence de score entre les élèves EIST et les élèves témoins au premier temps de mesure. Il est significatif, et positif, concernant la motivation, mais pas concernant l'intérêt pour les sciences en dehors de l'école. Le coefficient de l'interaction entre l'indicatrice EIST et le temps n'est pas significatif : les évolutions ne sont pas différentes selon que les élèves ont bénéficié ou non de l'EIST en sixième, aussi bien pour la motivation intrinsèque que pour l'intérêt pour les sciences en dehors de l'école.

Les modélisations confirment que le niveau moyen de motivation intrinsèque pour les sciences au premier temps de mesure est un peu plus élevé pour les élèves du groupe EIST que pour ceux du groupe témoin, de 0,25 point (sur une échelle de 3 à 12). Cette différence peut provenir de la façon dont les élèves ont été sélectionnés pour faire partie des classes qui expérimentent l'EIST, ou plus simplement du fait que faire partie d'un dispositif particulier donne aux élèves « *le sentiment qu'on leur accorde une sorte de privilège ou du moins un égard particulier et [qu']ils s'en sentent valorisés* » [PERROT, PIETRYK, ROJAT, 2009], ce qui peut impacter leur façon de répondre au questionnaire de motivation. Par ailleurs, les élèves du groupe EIST ne présentent pas un score initial d'intérêt pour les sciences en dehors de l'école significativement différent de celui des élèves du groupe témoin. Ce qui nous intéresse le plus ici concerne la comparaison des évolutions dans le temps des attitudes des élèves des deux groupes. Comme dans le cas du score cognitif, on constate que le coefficient de l'interaction avec le temps de la variable indicatrice indiquant la participation à l'EIST (mesuré en mois à partir du premier temps de mesure) n'est pas significativement différent de zéro pour les deux scores conatifs, ce qui signifie que l'évolution de ces deux scores tout au long de la scolarité au collège est tout à fait similaire, en moyenne, pour les élèves du groupe EIST et pour les élèves du groupe témoin. On ne peut donc pas conclure ici à un effet de l'EIST sur l'évolution de la motivation des élèves et de leur intérêt pour les sciences.

LIMITES

D'un point de vue méthodologique, les évaluations d'expérimentations se heurtent bien souvent à de nombreuses difficultés, qui doivent être identifiées car elles peuvent avoir des conséquences sur les résultats des analyses réalisées [FOUGÈRE, 2012].

Dans le cas de l'évaluation de l'EIST, la première difficulté réside dans le fait que l'affectation des établissements au dispositif était basée sur le volontariat. Il est probable que tous les établissements n'aient pas eu le même accès à l'information concernant l'existence du dispositif, voire que certains établissements aient été plus ou moins sollicités par les IA-IPR pour participer. De la même façon, il est possible que les enseignants engagés dans l'expérimentation aient des caractéristiques particulières. Les informations dont on dispose indiquent que les enseignants ont souvent été sollicités pour participer, plutôt que d'être eux-mêmes volontaires, mais ils ne l'ont probablement pas été au hasard. Enfin, les élèves participant peuvent également avoir des caractéristiques particulières. Cette sélection à différents niveaux peut avoir un impact sur les résultats des analyses, en particulier si elle s'est faite sur des critères fortement corrélés à nos variables d'intérêt, à savoir la performance en sciences des élèves, et leurs attitudes envers les sciences. Il s'avère que le niveau des élèves n'était pris en compte pour sélectionner les participants que dans un nombre restreint d'établissements, dans la majorité des cas la sélection étant aléatoire, ou concernant tous les élèves de sixième de l'établissement. Pour ce qui est des enseignants, on peut supposer qu'ils ont généralement été sollicités car ils étaient considérés comme susceptibles de faire bénéficier au mieux les

élèves de ce dispositif. Le risque est alors d'attribuer à l'EIST ce qui ne serait qu'un « effet enseignant ». À noter qu'il n'y a pas eu de questionnaire pour les chefs d'établissement et les enseignants du groupe témoin, ce qui ne permet pas d'évaluer dans quelle mesure les enseignants engagés dans l'EIST en sixième peuvent être différents de ceux des classes de sixième du groupe témoin. Dans la mesure où les élèves ont été suivis sur quatre années, au cours desquelles ils ont connu différents enseignants dans les matières scientifiques, on peut penser que l'effet, sur les résultats de nos modélisations, des caractéristiques des enseignants engagés dans l'EIST intervenant en sixième, s'il existe, reste limité. En tout état de cause, il ne semble pas suffisant pour faire apparaître un effet de l'EIST sur la progression des performances des élèves en sciences.

L'attrition est un autre problème méthodologique auquel nous sommes confrontés dans le cadre de ce dispositif d'évaluation, et qui peut également être la source d'un biais dans les résultats obtenus, surtout s'il touche de façon différentielle les groupes expérimentateur et témoin. Ici, l'échantillon final ne compte que 50 % des élèves sélectionnés au départ pour faire partie de la cohorte, et les causes de cette perte sont multiples (non-réponse au niveau de l'établissement, redoublements ou changements d'établissement des élèves, etc.).

Une autre difficulté réside dans le fait que les conditions d'application et la mise en œuvre de l'EIST peuvent largement varier d'un établissement et d'un enseignant à l'autre dans la mesure où l'EIST constitue un « *laboratoire de réflexion pédagogique* » [PERROT, PIETRYK, ROJAT, 2009]. Cela rend difficile l'identification de ce qui est réellement évalué.

Enfin, la mise en place du dispositif d'évaluation a dû se faire dans des délais très courts (durant l'été 2008 pour une première prise d'information à la rentrée suivante) ce qui peut se ressentir sur la qualité des instruments d'évaluation, en particulier lors des deux premiers moments de mesure. Malgré quelques items écartés, le score cognitif calculé présente de bonnes qualités psychométriques et semble pertinent pour mesurer les performances en sciences des élèves. En ce qui concerne le questionnaire conatif, la situation est moins favorable : nous avons dû nous restreindre finalement à étudier seulement deux dimensions, mesurées chacune par trois items.

CONCLUSION

L'évaluation de l'EIST ne montre pas que ce dispositif produise un effet significativement différent de celui d'un enseignement disciplinaire traditionnel sur l'évolution des performances des élèves en sciences et sur l'évolution de leur intérêt et de leur motivation pour ces dernières. Ces résultats viennent confirmer les premiers résultats, obtenus à partir des trois premiers temps de mesure [LE CAM et ROCHER, 2012]. Entre le début de sixième et la fin de troisième, les performances des élèves progressent en sciences, et leur intérêt et leur motivation à l'égard des sciences tendent à diminuer de manière générale. Les élèves ayant bénéficié du dispositif

d'enseignement intégré de sciences et technologie avaient un niveau de motivation pour les sciences en début de sixième légèrement supérieur en moyenne à celui des élèves du groupe témoin, cette différence se maintient jusqu'en fin de collège.

Des études montrent que les compétences et la qualité des enseignants sont déterminantes dans l'attitude des élèves envers les sciences, peut-être même plus que les contenus des programmes [OSBORNE, 2003]. Ainsi, un enseignement intégré de science et technologie, pour être efficace en termes d'amélioration de la performance en sciences des élèves et de leur motivation pour les sciences doit reposer sur des enseignants qui maîtrisent leur sujet. Dans le questionnaire auquel ils ont eu à répondre, les enseignants impliqués dans l'EIST soulignent particulièrement la difficulté à enseigner des contenus différents, mais également la précipitation dans laquelle s'est faite l'entrée dans le dispositif, et le manque de temps de préparation avant le début des cours. On peut penser qu'une action sur la formation des enseignants, largement en amont de la mise en œuvre d'un tel dispositif, pourrait en augmenter les effets.

Enfin, notre étude révèle un paradoxe couramment rencontré dans l'évaluation des dispositifs d'intervention, montrant que, bien que les données recueillies ne fassent pas apparaître d'effet de l'expérimentation sur les performances et attitudes des élèves, la plupart des acteurs engagés (enseignants, IA-IPR, élèves, parents, etc.) ont un ressenti plutôt positif [PERROT, PIETRYK, ROJAT, 2009]. Doit-on pour autant avancer que ce dernier aspect constitue le seul point positif du dispositif ? Au regard de la grande souplesse de mise en œuvre du dispositif expérimental et du grand nombre de facteurs confondus susceptibles d'avoir eu un impact sur les dimensions mesurées, il semble difficile de conclure avec certitude sur l'existence d'un impact positif du dispositif sur les performances et attitudes des collégiens en sciences. Rien, en tout cas, dans les résultats de cette évaluation, ne le fait apparaître.

BIBLIOGRAPHIE

ABRAHAMS I., MILLAR R., 2008, "Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science", *International Journal of Science Education*, vol. 30, No. 14, p. 1945-1969.

ABRAHAMS I., REISS M., 2012, "Practical work : Its effectiveness in primary and secondary schools in England", *Journal of Research in Science Teaching*, vol. 49, No. 8, p. 1 035-1 055.

ALFIERI L., BROOKS P., ALDRICH N., TENENBAUM H., 2011, "Does discovery-based instruction enhance learning?" *Journal of Educational Psychology*, vol. 103, No. 1, p.1-18.

BANERJEE A., DUFLO E., 2009, "The experimental approach to development economics", *Annual Review of Economics*, vol. 1, p. 151-178.

BENHAIM-GROSSE J., 2012, « L'enseignement intégré de science et de technologie (EIST) en 2008-2009 : ressenti et pratiques des enseignants », *Les dossiers*, MENJVA-DEPP, n° 200.

BRESSOUX P., 2010, *Modélisation statistique appliquée aux sciences sociales* (2^e édition), Bruxelles, De Boeck.

DELSERIEYS-PEDREGOSA A., BOILEVIN J-M., BRANDT-POMARES P., GIVRY D., MARTIN P., 2010, « Enseignement intégré de science et technologie, quels enjeux ? », *Review of Science, Mathematics and ICT Education*, vol. 4, No. 2, p. 9-28.

EURYDICE, 2006, *L'enseignement des sciences dans les établissements scolaires en Europe – État des lieux des politiques et de la recherche*, Bruxelles, 112 p.

Fondation La Main à la Pâte, 2011, *Un enseignement intégré de science et technologie au collège 6^e et 5^e*, *Guide de découverte*, 128 p.

FOUGÈRE D., 2012, « Les méthodes d'expérimentation en question », *Éducation & Formations*, n° 81, MENJVA-DEPP, p. 41-47.

GAGO J., CARO P., CONSTANTINO C., DAVIES G., PARCHMANN I., RANNIKMAE M., SJOBERG S., ZIMAN J., 2004, *Europe needs more scientists*, Bruxelles, Commission Européenne, Rapport par le groupe de haut niveau sur l'accroissement des ressources humaines pour la science et la technologie, 24 p.

GIVORD P., 2010, *Méthodes économétriques pour l'évaluation de politiques publiques*, Document de travail, G2010-08, Insee.

GURGAND M., VALDENAIRE M., 2012, « Le fonds d'expérimentation pour la jeunesse et les politiques éducatives : premier retour d'expérience », *Éducation & Formations*, n° 81, MENJVA-DEPP, p. 27-37.

KIRCHNER P., SWELLER J., CLARK R., 2006, "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching", *Educational Psychologist*, vol. 41, No. 2, p. 75-86.

OCDE, 2006, *Évolution de l'intérêt des jeunes pour les études scientifiques et technologiques – Rapport d'orientation*, Paris, 4 mai 2006, 21 p.

LE CAM M., ROCHER T., 2012, « 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 », *Éducation & Formations*, n° 81, MENJVA-DEPP, p. 79-90.

LE DONNÉ N., ROCHER T., 2010, « 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 », *Éducation & Formations*, n° 79, MENJVA-DEPP, p. 103-115.

L'HORTY Y., PETIT P., 2011, « Évaluation aléatoire et expérimentations sociales », *Revue française d'économie*, vol. 26, n°1, p. 13-48

OSBORNE J., 2003, "Attitudes towards science: a review of the literature and its implications", *International Journal of Science Education*, vol. 25, No. 9, p. 1049-1079.

OSBORNE J., DILLON J., 2008, *Science education in Europe : Critical reflections*, A report to the Nuffield Foundation, London, 32 p.

PERRON N., PIETRYK G., ROJAT D., 2009, *L'enseignement intégré de science et technologie (EIST)*, Rapport de l'inspection générale de l'éducation nationale.

ROCARD M., CSERMELY P., JORDE D., LENZEN D., WALBERG-HENRIKSSON H., HEMMO V., 2007, *L'enseignement scientifique aujourd'hui : une pédagogie renouvelée pour l'avenir de l'Europe*, Bruxelles, Commission Européenne, Rapport par le groupe de haut niveau sur l'enseignement scientifique, 28 p.

STOCKING M., LORD F., 1983, "Developing a common metric in item response theory", *Applied Psychological Measurement*, vol. 7, p. 201-210.