

Dans le cadre du cycle d'évaluations disciplinaires réalisées sur échantillon (CEDRE), la reprise, en 2009, de l'évaluation de 2003 portant sur les compétences générales des élèves de troisième permet de comparer les performances de ces élèves à six ans d'intervalle.

Le score moyen des élèves baisse de manière significative entre 2003 et 2009. Cette baisse se traduit par une augmentation de la proportion des élèves les plus faibles (de 15 % à 17,9 %) et par une diminution de la proportion des élèves les plus performants (de 10 % à 7,1 %). Cette tendance touche particulièrement le secteur de l'éducation prioritaire : près d'un tiers des élèves scolarisés dans les collèges de l'éducation prioritaire font partie des groupes les moins performants en 2009, contre un quart il y a six ans.

L'évolution des compétences générales des élèves en fin de collège de 2003 à 2009

Le cycle d'évaluations disciplinaires réalisées sur échantillon (CEDRE) a débuté en 2003 avec, pour objectif, de faire le point sur les compétences et connaissances des élèves à des moments-clefs du cursus scolaire, et ceci en regard des objectifs fixés par les programmes.

Ce dispositif permet ainsi une appréciation de l'évolution des acquis des élèves qui peut être mise en relation avec les caractéristiques du système et les politiques éducatives.

Cette évaluation, qui débute un nouveau cycle en reprenant en partie des exercices de 2003, permet ainsi une comparaison temporelle.

Le cadre de l'évaluation

L'évaluation présentée ici vise les compétences générales ou procédurales (savoir-faire) attendues en fin de collège, compétences qui devront ensuite être réinvesties lors de la formation ultérieure, que celle-ci soit immédiate ou différée. Cette évaluation s'appuie sur les programmes du collège, mais n'est pas une évaluation directe de ce qui est défini dans ces programmes.

Les trois grands domaines de compétences

Trois grands domaines de compétences ont été retenus, gradués en fonction d'une complexité croissante des opérations mentales nécessaires pour les mettre en œuvre :

- prélever l'information ;
- organiser l'information prélevée ;
- exploiter l'information de manière complexe.

La définition et la composition de ces domaines sont rappelées ci-après :

« **Prélever l'information** » est relatif à l'ensemble des moyens dont l'élève dispose pour accéder aux informations contenues dans les documents (textes, images, graphiques, cartes, plans...). L'information à prélever peut l'être de façon immédiate (information explicite), ou nécessiter le repérage de plusieurs éléments à combiner.

« **Organiser l'information prélevée** » consiste à trier, associer, choisir, reformuler, transposer d'un support à un autre, appliquer directement un principe, une règle, une méthode donnée à une nouvelle situation, effectuer des inférences simples à partir de vocabulaire inconnu, à partir d'indices textuels ou typographiques, ou d'accords éventuels.

« Exploiter l'information de manière complexe » est défini par le raisonnement déductif et la construction d'inférences complexes nécessaires à la compréhension du sens implicite d'un texte ou des relations de cause et de conséquence, au choix d'une argumentation, ou à la mise en œuvre d'une démarche expérimentale.

Pour chacun de ces grands domaines de compétences, on a cherché à définir une hiérarchie des différents niveaux de maîtrise, cette hiérarchie étant ensuite mise à l'épreuve, corrigée lors de l'expérimentation du protocole, puis confirmée lors de l'exploitation de l'enquête comme on le verra plus loin.

Les épreuves

Différentes composantes de ces trois domaines de compétences ont été envisagées dans quatre grands champs disciplinaires : français, mathématiques, histoire-géographie-éducation civique et sciences, en prenant en compte ce que ces composantes apportaient à la maîtrise de la compétence commune et non pas à la spécificité disciplinaire. L'évaluation est ainsi organisée en situations réparties dans les différents grands champs disciplinaires. L'évaluation de 2009 a toutefois innové en ajoutant des items pour lesquels le contexte n'est pas rattaché à l'une des disciplines citées précédemment.

Chaque situation est composée d'un document de travail—support de l'évaluation—et de plusieurs questions. Chaque support d'évaluation permet de proposer de une à neuf questions, chacune d'elles ciblant un seul niveau de compétence. Au final, l'analyse a porté sur 63 situations regroupant 303 questions ou items. Parmi ces items, 74 étaient repris à l'identique de l'épreuve passée en 2003. La diversité des situations et des supports a été recherchée et les documents proposés sont d'une nature et d'une complexité adaptée aux élèves de fin de collège.

En 2009, l'échantillon est composé de 202 collèges. Les cahiers d'évaluation comprenaient deux séquences cognitives d'une heure chacune, suivies d'un questionnaire de contexte de 30 minutes. Dans les cahiers destinés aux élèves, les situations

d'évaluation n'étaient pas regroupées selon la discipline mais s'enchaînaient en passant d'un contexte disciplinaire à un autre. Cependant, l'appartenance à une discipline spécifique était en général clairement identifiable.

Évolution des performances des élèves

Pour formaliser un constat de la maîtrise des compétences générales par les élèves en fin de collège, une échelle décrivant six niveaux de performance a été construite en 2003 (*graphique 1*). Elle sert de référence pour la comparaison des résultats à 6 ans d'intervalle.

Les performances décrites à chaque niveau de l'échelle sont identiques à celles de 2003. Une évolution apparaît dans la répartition des élèves selon les niveaux de l'échelle. En 2009, le pourcentage d'élèves des groupes de niveau faible (0 et 1) a augmenté de manière significative (1), passant de 15 % à 17,9 %. À l'autre bout de l'échelle, le pourcentage d'élèves de niveau élevé a diminué, évoluant de 10 % à 7,1 %.

L'analyse des items nouveaux de 2009 fait apparaître une similitude quant à l'acquisition des différentes compétences d'un groupe d'élèves à un autre. Le graphique 2 montre que la hiérarchie dans la maîtrise des trois domaines de compétences n'a pas changé entre 2003 et 2009. Cependant, il indique une baisse de réussite assez sensible. Plus la compétence est complexe, plus la baisse est importante.

Pour l'ensemble des groupes, on observe une gradation dans la maîtrise des compétences. La compétence « prélever l'information » est réussie en moyenne à 81 %. La réussite moyenne aux items évaluant la compétence « organiser l'information prélevée » s'élève à 67 %. Quant à la compétence « exploiter l'information de manière complexe », elle semble être la plus difficile à acquérir puisque les items qui l'évaluent ne sont réussis en moyenne qu'à 57 %.

Six groupes aux compétences différenciées

À côté de l'échelle de performance établie à partir des résultats des élèves aux

items communs aux deux enquêtes (*voir l'encadré p. 6*), les données recueillies à partir des items nouveaux de 2009 permettent d'apporter de nouvelles informations sur les connaissances et les compétences des élèves à six ans d'intervalle.

— **Groupe 0** : au vu des résultats de l'évaluation, les élèves du groupe 0 ne répondent que ponctuellement à des items dont les compétences sont relatives au prélèvement d'information.

— **Groupe 1** : les compétences relatives au prélèvement d'information sont majoritairement représentées chez les élèves du groupe 1. Ils ne sont pas encore en mesure d'organiser ni d'exploiter une ou plusieurs informations. Les élèves qui constituent les groupes 0 et 1 sont des élèves en grande difficulté. Ils arrivent en fin de scolarité au collège avec des compétences très limitées qui seront un obstacle à la poursuite d'études mais aussi dans l'adaptation à la vie quotidienne.

— **Groupe 2** : la majorité des compétences maîtrisées par les élèves du groupe 2 sont encore des compétences relatives au prélèvement d'information. Ils commencent néanmoins à savoir organiser les informations prélevées. En particulier, ils commencent à être capables d'organiser une ou plusieurs informations après avoir effectué un traitement. Les élèves du groupe 2 sont capables de justifier une réponse en identifiant les éléments d'un document nécessaire à sa compréhension et d'utiliser les codes de la discipline. S'ils savent prélever l'information et commencer à l'organiser, ils ne sont pas encore capables de l'exploiter pleinement.

— C'est à partir du **groupe 3** que commence à s'installer la compétence « exploiter l'information de manière complexe ». Les élèves deviennent capables de mettre en relation différentes informations et de comprendre des relations de cause à effet. La mise en œuvre d'un procédé ou d'une expérience à partir d'instructions devient possible à partir de ce niveau. L'exploitation des informations de manière complexe est une compétence qui est, toutefois, encore fragile chez les élèves de ce groupe.

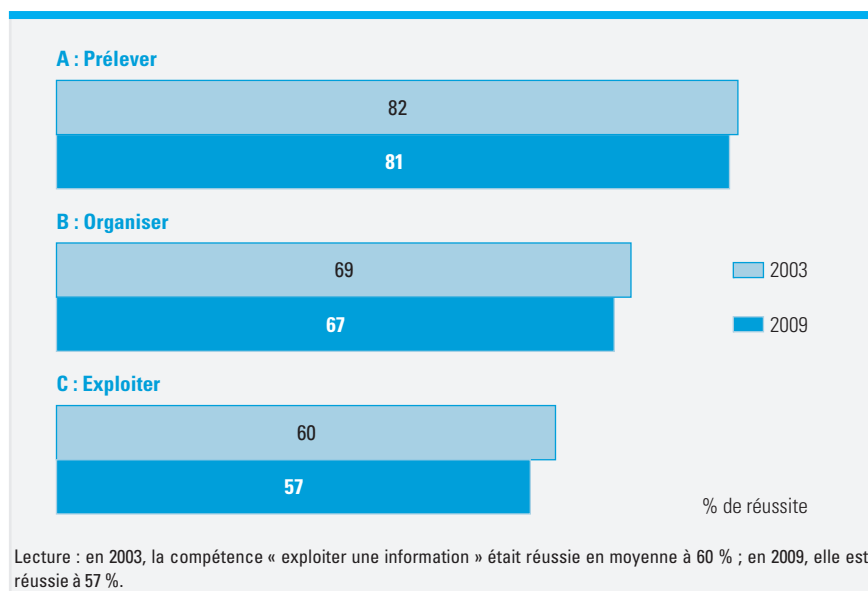
1. Pour plus de précisions sur la notion de significativité, voir l'encadré p. 6.

Évaluations en 2003 et 2009 : répartition des élèves selon l'échelle de compétences générales

Scores	% Population en 2003	% Population en 2009	échelle de compétences générales
Groupe 5 316	<div style="background-color: #90caf9; width: 100px; height: 15px; margin-bottom: 5px;">10,0 %</div> <div style="background-color: #00bcd4; width: 100px; height: 15px;">7,1 %</div>		Les élèves sont capables d'exploiter l'information contenue dans des documents écrits, en mettant en œuvre à un degré élevé l'ensemble des aspects constitutifs des compétences générales qui s'acquièrent au cours des années de collège. Ils sont capables de juger de la pertinence d'une méthode pour la résolution d'un problème.
316 Groupe 4 278	<div style="background-color: #90caf9; width: 100px; height: 15px; margin-bottom: 5px;">15,8 %</div> <div style="background-color: #00bcd4; width: 100px; height: 15px;">17,5 %</div>		Les élèves sont capables de mettre en relation des informations issues de contextes hétérogènes, de classer et de comparer ces informations. Ils sont capables de comprendre le sens général et implicite d'un texte qu'ils peuvent interpréter avec finesse. Ils sont capables de résumer et de synthétiser un texte ou un document. Ils sont capables de choisir une argumentation, de décider d'une méthode et de l'utiliser pour résoudre un problème. Ils sont capables de comprendre et de mettre en œuvre une démarche expérimentale, de mener un raisonnement complexe, d'utiliser la pensée hypothético-déductive.
278 Groupe 3 239	<div style="background-color: #90caf9; width: 100px; height: 15px; margin-bottom: 5px;">29,6 %</div> <div style="background-color: #00bcd4; width: 100px; height: 15px;">28,0 %</div>		Les élèves sont capables d'organiser une ou plusieurs informations après avoir effectué un traitement sur ces informations. Ils sont capables de justifier une réponse, en utilisant par exemple les marques anaphoriques leur permettant d'identifier les personnages d'un récit. Ils sont capables d'identifier les référents des substituts, d'identifier l'auteur ou le genre d'un texte. Ils montrent leur capacité à comprendre les relations de cause et de conséquence. Ils sont capables de choisir une conclusion en utilisant des éléments contenus dans un texte littéraire ou présentés dans un texte relatant une expérience scientifique.
239 Groupe 2 201	<div style="background-color: #90caf9; width: 100px; height: 15px; margin-bottom: 5px;">29,6 %</div> <div style="background-color: #00bcd4; width: 100px; height: 15px;">29,5 %</div>		Les élèves sont capables de prélever des informations explicites contenues dans un document, qu'elles soient immédiatement repérables ou non. Ils sont capables de combiner plusieurs éléments prélevés dans un document, d'extraire une information pertinente selon plusieurs critères. Ils sont capables de prélever des informations dans un graphique, un schéma, un texte continu, de prélever des éléments, des mots, des chiffres dans un tableau. Ils sont en passe de maîtriser l'organisation de plusieurs informations pour effectuer un traitement. Ils sont au seuil de maîtrise de la compréhension spatiale et temporelle.
201 Groupe 1 162	<div style="background-color: #90caf9; width: 100px; height: 15px; margin-bottom: 5px;">12,9 %</div> <div style="background-color: #00bcd4; width: 100px; height: 15px;">13,9 %</div>		Les élèves sont capables de prélever une information explicite lorsque celle-ci est facilement repérable. Ils sont en difficulté devant un texte complexe ou comprenant un vocabulaire peu courant ou spécifique d'une discipline.
162 Groupe 0	<div style="background-color: #90caf9; width: 100px; height: 15px; margin-bottom: 5px;">2,1 %</div> <div style="background-color: #00bcd4; width: 100px; height: 15px;">4,0 %</div>		Bien que capables de répondre ponctuellement à quelques questions, les élèves ne maîtrisent aucune des compétences attendues en fin de collège.

Lecture : les élèves du groupe 3 ont entre 239 et 278 points. Ils sont capables de réaliser des tâches de niveaux de difficulté correspondants aux groupes 0, 1, 2 et 3. Ils ont une probabilité faible de réussir les tâches spécifiques aux groupes 4 et 5. Les valeurs sont arrondies à la décimale la plus proche, et de manière à obtenir des sommes égales à 100 %.

GRAPHIQUE 2 – Évolution de la réussite dans les trois compétences



Source : MEN - DEPP

– Les élèves du **groupe 4** sont capables de situer des événements dans le temps, ils commencent à comprendre le cadre spatial et temporel. La capacité à raisonner, mettre en œuvre une démarche commencent vraiment à s’installer à partir de ce groupe. L’exploitation de l’information de manière complexe est beaucoup plus aisée que dans le groupe précédent. Dès le groupe 4, les élèves se montrent très compétents lorsqu’il faut organiser l’information.

– Ce qui caractérise les élèves du **groupe 5**, c’est une maîtrise relativement équilibrée des différentes compétences évaluées. Les élèves sont capables de mettre en œuvre des démarches scientifiques, comprennent le sens implicite d’un texte et mènent des raisonnements complexes.

Les items nouveaux qui ne sont pas ancrés dans une discipline en particulier, n’ont pas posé plus de difficulté aux élèves que ceux clairement rattachés à une discipline. Cependant, la principale compétence évaluée à travers ces items est celle qui est en général la mieux maîtrisée par les élèves : le prélèvement d’information.

L’enquête de 2009 révèle une baisse du pourcentage d’élèves de niveau élevé (10 % en 2003 contre 7,1 % en 2009) et, dans le même temps, une augmentation du pourcentage d’élèves de niveau faible (15 % en 2003 contre 17,9 % en 2009). L’évolution dans la maîtrise des compétences reste la même. Les compétences des élèves de niveau faible (groupes 0 et 1)

se limitent au prélèvement d’informations. L’organisation des informations, qui commence à s’observer chez les élèves de niveau 2, n’est vraiment maîtrisée que par un peu plus d’un élève sur deux. Seuls les élèves de niveau élevé (groupe 4) et très élevé (groupe 5) – soit environ un quart des élèves – savent exploiter une information de manière complexe.

TABLEAU 1 – Répartition et score moyen selon le sexe des élèves et répartition selon les groupes de niveaux en 2003 et en 2009

		Répartition (en %)	Score moyen	Groupe 0	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Ensemble	2003	100,0	250	2,1	12,9	29,6	29,6	15,8	10,0
	2009	100,0	245	4,0	13,9	29,5	28,0	17,5	7,1
Garçons	2003	48,2	249	2,3	14,0	29,3	29,1	15,5	9,8
	2009	49,3	244	5,1	15,3	27,4	26,6	17,9	7,7
Filles	2003	51,8	253	1,3	10,9	30,0	30,6	16,6	10,6
	2009	50,7	246	2,9	12,5	31,4	29,4	17,3	6,5

Lecture : les garçons représentent 48,2 % des élèves enquêtés en 2003 et 49,3 % en 2009. Leur score a significativement baissé (- 5 pts) entre les deux cycles d’évaluation ; 5,1 % d’entre eux appartiennent au groupe de niveau 0 en 2009 contre 2,3 % en 2003. Les évolutions significatives sont marquées en gras et en italique lorsqu’il s’agit d’une baisse et en gras et en bleu dans le cas d’une augmentation entre 2003 et 2009. Les valeurs sont arrondies à la décimale la plus proche et de manière à obtenir des sommes égales à 100 %.

Source : MEN - DEPP

TABLEAU 2 – Répartition et score moyen selon le cursus scolaire des élèves et répartition selon les groupes de niveaux en 2003 et en 2009

		Répartition (en %)	Score moyen	Groupe 0	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Ensemble	2003	100,0	250	2,1	12,9	29,6	29,6	15,8	10,0
	2009	100,0	245	4,0	13,9	29,5	28,0	17,5	7,1
Élèves « à l’heure »	2003	66,0	264	1,0	7,0	24,4	32,7	20,6	14,3
	2009	72,2	256	2,0	9,1	27,4	30,0	22,0	9,5
Élèves en retard	2003	34,0	223	4,5	24,2	39,9	23,5	6,5	1,4
	2009	27,8	217	9,3	26,2	35,0	22,8	6,0	0,7

Lecture : Les élèves « à l’heure » représentent 66 % des élèves enquêtés en 2003 et 72,2 % en 2009. Leur score a significativement baissé (- 8 pts) entre les deux cycles d’évaluation ; 2 % d’entre eux appartiennent au groupe de niveau 0 en 2009 contre 1 % en 2003. Les évolutions significatives sont marquées en gras et en bleu lorsqu’il s’agit d’une baisse et en gras et en bleu dans le cas d’une augmentation entre 2003 et 2009. Les valeurs sont arrondies à la décimale la plus proche et de manière à obtenir des sommes égales à 100 %.

Source : MEN - DEPP

Des évolutions différenciées selon les caractéristiques des élèves

Les résultats mettent en évidence une évolution différenciée en fonction des caractéristiques sociodémographiques et scolaires des élèves.

Plus de garçons que de filles aux faibles performances

Le tableau 1 montre que le score moyen des filles a davantage baissé que celui des garçons entre les deux cycles d’évaluation (- 7 points contre - 5 points). Néanmoins, les filles demeurent moins nombreuses aux plus bas niveaux : 15,4 % d’entre elles font partie des groupes 0 et 1 en 2009 contre 20,4 % des garçons, et le différentiel filles/garçons s’accroît, passant de 4,1 % en 2003 à 5 % en 2009.

Les élèves en retard scolaire toujours plus nombreux dans les bas niveaux

La part des élèves en retard par rapport à leur cursus scolaire a diminué entre les deux cycles d’évaluation : ils sont 27,8 % en 2009 contre 34 % en 2003 (tableau 2).

TABLEAU 3 – Répartition et score moyen selon les conditions de scolarisation des élèves et répartition selon les groupes de niveaux en 2003 et en 2009

		Répartition (en %)	Score moyen	Groupe 0	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Ensemble	2003	100,0	250	2,1	12,9	29,6	29,6	15,8	10,0
	2009	100,0	245	4,0	13,9	29,5	28,0	17,5	7,1
Élèves inscrits dans des collèges publics (hors EP)	2003	64,8	250	2,1	12,7	30,2	29,9	15,2	9,9
	2009	64,0	245	3,8	13,9	29,8	27,2	18,2	7,1
Élèves inscrits dans des collèges privés	2003	20,7	262	1,1	8,0	25,6	31,3	20,9	13,1
	2009	21,5	258	1,5	7,0	25,6	34,0	22,6	9,3
Élèves inscrits dans des établissements de l'éducation prioritaire	2003	14,5	235	3,9	21,0	32,9	25,3	11,0	5,9
	2009	14,5	223	8,7	23,9	33,9	22,5	7,5	3,5

Lecture : les élèves de l'enseignement prioritaire représentent 14,5 % des élèves enquêtés, en 2003 comme en 2009. Leur score a significativement baissé (- 12 pts) entre les deux cycles d'évaluation ; 8,7 % d'entre eux appartiennent au groupe de niveau 0 en 2009 contre 3,9 % en 2003. Les évolutions significatives sont marquées en gras et en italique lorsqu'il s'agit d'une baisse et en gras et en bleu dans le cas d'une augmentation entre 2003 et 2009. Les valeurs sont arrondies à la décimale la plus proche et de manière à obtenir des sommes égales à 100 %.

Source : MEN - DEPP

TABLEAU 4 – Répartition et score moyen selon la situation des élèves par rapport à l'immigration et répartition selon les groupes de niveaux en 2003 et en 2009

		Répartition (en %)	Score moyen	Groupe 0	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Ensemble	2003	100,0	250	2,1	12,9	29,6	29,6	15,8	10,0
	2009	100,0	245	4,0	13,9	29,5	28,0	17,5	7,1
Élèves non issus de l'immigration	2003	86,9	254	1,4	10,8	28,7	30,9	17,2	11,0
	2009	87,0	248	3,1	12,2	29,0	29,4	18,8	7,5
Élèves issus de l'immigration, de 1 ^{re} génération	2003	3,3	228	7,1	24,1	31,1	21,4	10,8	5,5
	2009	3,8	211	16,5	27,8	31,9	14,9	4,6	4,3
Élèves issus de l'immigration, de 2 ^e génération	2003	9,8	229	3,9	22,7	37,3	23,6	8,0	4,5
	2009	9,2	227	6,5	24,0	32,8	21,2	11,3	4,2

Lecture : Les élèves non issus de l'immigration représentent 87 % des élèves enquêtés en 2003 comme en 2009. Leur score a significativement baissé (- 6 pts) entre les deux cycles d'évaluation ; 3,1 % d'entre eux appartiennent au groupe de niveau 0 en 2009 contre 1,4 % en 2003. Les évolutions significatives sont marquées en gras et en italique lorsqu'il s'agit d'une baisse et en gras et en bleu dans le cas d'une augmentation entre 2003 et 2009. Les valeurs sont arrondies à la décimale la plus proche, et de manière à obtenir des sommes égales à 100 %.

Source : MEN - DEPP

Même si le score moyen des élèves en retard a diminué relativement moins que celui des élèves « à l'heure » (- 6 points contre - 8 points), ils continuent d'être près de cinq fois plus nombreux que ces derniers au plus bas niveau (9,3 % d'entre eux contre 2 % parmi les élèves « à l'heure » font partie du groupe 0 en 2009).

Un écart qui s'accroît en défaveur de l'éducation prioritaire

Le tableau 3 indique que les résultats sont également en baisse dans le secteur de l'éducation prioritaire où le score moyen a diminué de 12 points, et où la part des élèves les plus faibles (groupe 0) a plus que doublé entre 2003 et 2009 (de 3,9 %

2. Le secteur de l'éducation prioritaire est ici appréhendé par rapport aux informations de la base centrale des établissements à la rentrée 2008, qui avaient conservé l'information « ZEP/REP », malgré l'introduction, en 2006, des RAR et RRS. L'ensemble des collèges considérés ici en éducation prioritaire (EP) est très stable : 98 % des collèges EP en 2009 étaient déjà en EP en 2003 et 93 % des collèges EP en 2003 étaient toujours EP en 2009.

3. Les élèves issus de l'immigration de première génération sont les élèves nés à l'étranger, de parents nés à l'étranger ; les élèves issus de l'immigration de deuxième génération sont nés en France, de parents nés à l'étranger.

à 8,7 %). La différence de score avec le secteur public (hors éducation prioritaire) s'accroît quant à elle de 7 points, passant de 15 à 22 points (2).

L'immigration de première génération particulièrement touchée par l'augmentation des bas niveaux

On note enfin (tableau 4) une baisse importante de 16 points du score moyen des élèves issus de l'immigration, de première génération (3) ; 16,5 % d'entre eux font partie du groupe 0 en 2009 alors qu'ils n'étaient que 7,1 % en 2003. Par ailleurs, ces élèves qui, en 2003, obtenaient un score moyen comparable à celui des élèves issus de l'immigration de deuxième génération, réussissent, en 2009, nettement moins bien que ces derniers, dont les résultats n'ont pas significativement changé.

**Ginette Bourny,
Pascal Bessonneau,
Jeanne-Marie Daussin
et Saskia Keskpaik, DEPP B2**

Pour en savoir plus

« L'évolution des acquis des élèves de 15 ans en compréhension de l'écrit. Premiers résultats de l'évaluation internationale PISA 2009 », *Note d'Information*, MEN-DEPP, (à paraître).

« Les compétences générales des élèves en fin de collège », *Note Évaluation* 04.09, MEN-DEP, octobre 2004.

www.education.gouv.fr

depp.documentation@education.gouv.fr

Méthodologie

Le CEDRE

Le cycle des évaluations disciplinaires réalisées sur échantillon (CEDRE) établit des bilans nationaux des acquis des élèves en fin d'école et en fin de collège. Il couvre les compétences des élèves dans la plupart des domaines disciplinaires en référence aux programmes. La présentation des résultats permet de situer les performances des élèves sur des échelles de niveaux allant de la maîtrise pratiquement complète de ces compétences à une maîtrise bien moins assurée, voire très faible, de celles-ci. Renouvelées tous les six ans, ces évaluations permettent de répondre à la question de l'évolution du « niveau des élèves » au fil du temps.

Le calendrier des évaluations-bilans

2003 : compétences générales (publiée)
2004 : langues étrangères (publiée)
2005 : attitudes à l'égard de la vie en société (publiée)
2006 : histoire-géographie, éducation civique (publiée)
2007 : sciences expérimentales (en cours de publication)
2008 : mathématiques (publiée)
2009 : reprise du cycle, compétences générales

L'échantillonnage

La population visée est celle des élèves de troisième générale des collèges publics et privés sous contrat de France métropolitaine.

– L'échantillon 2003

En 2003, deux échantillons ont été tirés : l'un représentatif des élèves au niveau national et l'autre assurant une représentation de chaque académie. Au niveau national, l'échantillon est stratifié sur la taille et comporte 112 établissements. Il est constitué de classes entières sélectionnées aléatoirement dans ces établissements. Au niveau académique, trente élèves ont été tirés aléatoirement dans dix-sept établissements par académie.

Au total, le rassemblement de ces deux échantillons constitue un ensemble de 23 500 élèves, 552 collèges et 411 classes entières. En raison, notamment, des mouvements sociaux du printemps 2003, seules les réponses de 18 222 élèves ont pu être analysées, ce qui a nécessité un redressement de l'échantillon.

– L'échantillon 2009

En 2009, 202 collèges ont été sélectionnés aléatoirement en vue d'une représentativité nationale seulement.

Une stratification a été effectuée sur la taille (collèges de plus de 14 élèves et collèges de moins de 14 élèves) et le secteur de l'établissement (public hors zone d'éducation prioritaire, zone d'éducation prioritaire et le secteur privé sous contrat). Dans chaque établissement, 30 élèves ont été évalués. Cela représente environ 6 000 élèves. Les réponses de 5 142 élèves ont pu être analysées dans les 193 collèges répondants.

Pour tenir compte de la non-réponse, les échantillons ont été redressés.

Contraintes de la situation d'évaluation

Pour l'enquête de 2003, en raison de contraintes techniques fortes (la taille de l'échantillon de départ interdisant une possibilité de correction par des experts), les 204 items sont du format « questions à choix multiples » (QCM). Cette forme de questionnement a pour avantages d'éviter un travail de correction et d'automatiser la saisie des réponses des élèves, mais elle a pour inconvénient de ne pas permettre l'évaluation des compétences des élèves en production d'écrit, en situation d'autonomie.

En 2009, des questions ouvertes, c'est-à-dire appelant une réponse rédigée ont été introduites. Le test comportait 61 questions ouvertes et 112 QCM.

Par ailleurs, si chaque élève avait dû passer l'ensemble des situations proposées, au moins six heures d'évaluation par élève auraient été nécessaires. Pour limiter la passation à deux heures pour chaque élève,

les items ont été regroupés en treize « blocs », répartis ensuite dans seize cahiers (2003) ou treize cahiers (2009) différents selon un plan expérimental, organisant un tuilage des blocs. Ce dispositif, couramment utilisé dans les évaluations CEDRE et internationales, permet d'estimer la probabilité de réussite à chaque item sans que chaque élève passe l'ensemble des items.

La construction de l'échelle de performance

L'échelle de performance a été élaborée en utilisant les modèles de réponse à l'item. Le score moyen de compétences générales, correspondant à la performance moyenne des élèves de l'échantillon de 2003, a été fixé par construction à 250 et l'écart-type à 50. Cela implique qu'environ deux tiers des élèves ont un score compris entre 200 et 300. Mais cette échelle, comme celle de l'enquête PISA, n'a aucune valeur normative et, en particulier, la moyenne de 250 ne constitue en rien un seuil qui correspondrait à des compétences minimales à atteindre.

Sur la base de constats fréquemment établis dans les différentes évaluations antérieures de la DEPP, qui montrent que 15 % des élèves peuvent être considérés en difficultés en fin de scolarité obligatoire, la partie la plus basse de l'échelle est constituée des scores obtenus par les 15 % d'élèves ayant les résultats les plus faibles (groupes 0 et 1). À l'opposé, la partie supérieure, constituée des scores les plus élevés, rassemble 10 % des élèves. Entre ces deux niveaux, l'échelle a été scindée en trois parties d'amplitudes de scores égales, correspondant à trois groupes intermédiaires.

Les modèles de réponse à l'item ont l'avantage de positionner sur la même échelle les scores des élèves et les difficultés des items. Cette correspondance permet de caractériser les compétences maîtrisées par chacun des groupes d'élèves.

La comparabilité 2003-2009

Afin de pouvoir comparer les résultats des enquêtes 2003 et 2009, 74 items d'ancrage (c'est-à-dire des items de 2003) ont été repris à l'identique dans l'évaluation de 2009.

Lors de l'analyse des résultats en 2010, les modèles de réponse à l'item ont été réutilisés et appliqués cette fois à l'ensemble des résultats 2003 et 2009.

L'estimation conjointe des modèles de réponse à l'item, à partir des données de 2003 et de 2009, et la présence d'items communs entre les deux évaluations permet la comparaison directe à la fois des scores des individus et des difficultés des items entre les deux passations.

Notons qu'un soin particulier a été apporté à l'analyse de ces items communs entre 2003 et 2009. Ainsi, les items retenus dans l'analyse finale devaient, d'une part, ne pas présenter de fonctionnement différentiel (par exemple un écart de taux de réussite entre les deux passations anormalement élevé), d'autre part, devaient avoir des propriétés psychométriques satisfaisantes.

Les données de 2003 ont donc été réanalysées dans cette perspective de comparaison. C'est pourquoi certaines différences – notamment s'agissant des valeurs des scores seuils – peuvent apparaître par rapport à la publication initiale.

Significativité

Les évaluations du CEDRE sont réalisées sur échantillons. De ce fait, les résultats sont soumis à une variabilité qui dépend des erreurs d'échantillonnage. Il est possible d'estimer statistiquement ces erreurs d'échantillonnage et de produire des intervalles de confiance. À titre d'exemple, le score moyen des élèves sur l'échelle de compétences générales est de 245 en 2009 mais le vrai score, tel qu'il serait calculé pour l'ensemble des élèves de troisième, se situe, avec une probabilité de 95 %, entre 242 et 248 (c'est-à-dire ± 3 points). Par conséquent, le score moyen des élèves de 2009 est significativement différent de celui des élèves de 2003 qui est de 250 (± 2 points).