Présentation générale de l'enquête

OBJECTIFS DE L'ENQUÊTE

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 au regard des objectifs fixés par les programmes officiels. La présentation des résultats permet de situer les performances des élèves sur des échelles de niveau allant de la maîtrise pratiquement complète de ces compétences à une maîtrise bien moins assurée, voire très faible, de celles-ci. Renouvelées régulièrement, ces enquêtes permettent de répondre à la question de l'évolution du niveau des élèves.

Ces évaluations n'ont pas valeur de délivrance de diplômes, ni d'examen de passage ou d'attestation de niveau ; elles donnent une photographie instantanée de ce que savent et savent faire les élèves à la fin d'un cursus scolaire. En ce sens, il s'agit bien d'un bilan. Destinées à être renouvelées périodiquement, ces évaluations-bilans permettent également de disposer d'un suivi de l'évolution des acquis des élèves dans le temps. Pour cette raison, le contenu ne peut pas être rendu public car, devant être en grande partie repris lors des cycles d'évaluation suivants, il ne doit pas servir pour réaliser des exercices dans les classes.

Les enquêtes CEDRE apportent un éclairage qui intéresse tous les niveaux du système éducatif, des décideurs aux enseignants sur le terrain, en passant par les formateurs d'enseignants : elles informent sur les compétences et les connaissances des élèves à la fin d'un cursus ; elles renseignent sur l'attitude et la représentation des élèves à l'égard de la discipline ; elles interrogent les pratiques d'enseignement au regard des programmes ; elles contribuent à enrichir la réflexion générale sur l'efficacité et la performance de notre système éducatif.

CEDRE a été initié en 2003 avec l'évaluation des compétences générales. Afin d'assurer une comparabilité dans le temps, l'évaluation est reprise pour chaque discipline selon un cycle de six ans jusqu'en 2012 et de cinq ans depuis 2012 (tableau 1).

TABLEAU 1 : Cycle des évaluations disciplinaires réalisées sur échantillons (CEDRE) depuis 2003

Discipline évaluée	Début de cycle	Reprises	
Maîtrise de la langue et compétences générales	2003	2009	2015
Langues étrangères	2004	2010	2016
Attitude à l'égard de la vie en société	2005	_	_
Histoire, géographie et éducation civique	2006	2012	2017
Sciences	2007	2013	2018
Mathématiques	2008	2014	2019

MISE EN ŒUVRE DE L'EVALUATION DES ACQUIS DES ELEVES.

Les programmes de référence

L'évaluation CEDRE en fin de collège en histoire, géographie et enseignement moral et civique a pour objectif de faire le point des connaissances et des compétences des élèves tant sur le plan des savoirs que des savoir-faire d'une part, et, d'autre part, de mesurer l'évolution de ces connaissances et compétences entre 2006, 2012 et 2017.

Les connaissances et compétences telles qu'elles sont définies dans les programmes officiels constituent le cadre de cette évaluation.

En mai 2006 comme en mai 2012, les élèves de troisième ont été évalués à partir des programmes de 1995 (classes de sixième, cinquième et quatrième) et de 1998 (classes de troisième).

En mai 2017 en revanche, les élèves de troisième ont été interrogés en référence aux programmes de 2008 pour leurs classes de la sixième à la quatrième incluse, et aux programmes de 2015 entrés en vigueur à la rentrée 2016 pour leur classe de troisième.

Les compétences évaluées

En 2017, comme en 2012 et en 2006, une même grille de compétences est utilisée pour les trois champs de cette évaluation : histoire, géographie et enseignement moral et civique **TABLEAU 2**. La plupart des intitulés sont restés identiques entre les trois points de mesure, ce qui permet d'assurer la comparaison entre les prises d'informations.

TABLEAU 2 : Tableau de compétences

	Histoire	Géographie	Enseignement moral et civique
Identifier – Localiser	HL	GL	EL
Identifier – Décrire	HD	GD	ED
Identifier – Sélectionner	HS	GS	ES
Traiter l'information - Classer	HC	GC	EC
Traiter l'information - Mettre en relation	HM	GM	EM
Interpréter - Généraliser	HG	GG	EG
Interpréter - Argumenter	HA	GA	EA
Interpréter - Réaliser	HR	GR	ER

Note de lecture : chaque compétence est définie par une lettre (L pour localiser, D pour décrire...). Chaque matière est définie de façon identique (H pour histoire, G pour géographie et E pour enseignement moral et civique). Le croisement de la matière et de la compétence est matérialisé par un couple de lettres, associé à l'item produit.

Cette grille reste pertinente dans le cadre du renouvellement des programmes de 2015 et des nouvelles compétences qu'ils redéfinissent et qui, pour rappel, ont concerné les élèves évalués en 2017 pour leur année de troisième seulement. Par exemple, les compétences « se repérer dans le temps » et « se repérer dans l'espace » des programmes de 2015 renvoient à la compétence CEDRE « identifier-localiser » ; « raisonner, justifier une démarche et les choix effectués » se superpose à « interpréter » ; « analyser et comprendre un document » correspond, selon le cas, à « identifier-sélectionner l'information » ou « traiter l'information », voire « interpréter ». Enfin, « pratiquer différents langages en histoire-géographie » renvoie à « interpréter-réaliser ».

Seules deux nouvelles compétences de 2015, à savoir « s'informer dans le monde du numérique » et « coopérer et mutualiser », n'ont pas pu être évaluées dans le cadre de CEDRE 2017. Il faut cependant préciser que ces compétences de 2015 sont destinées à prendre la place de la grille des compétences Cedre pour la prochaine échéance de 2022, du fait notamment des nouvelles opportunités d'une évaluation qui sera intégralement opérée sur support numérique, et sous réserve que les programmes de 2015, ou du moins les compétences qu'ils définissent, restent en vigueur jusqu'à ce quatrième temps de mesure.

Par ailleurs, pour mieux prendre en compte la dimension déclarative constitutive de la notion même de compétence, chacun des items du corpus de l'évaluation de 2017 a aussi été indexé selon une seconde grille **TABLEAU 3**. Elle définit les connaissances évaluées selon une ou plusieurs de ces composantes : concepts ou notions, autre vocabulaire, repères (historiques et géographiques), acteurs ou actrices (historiques ou spatiaux).

Tableau 3 : Grille des connaissances associées aux compétences (pour deux exemples d'items)

items	Concepts et notions	Autre vocabulaire	Repères	Acteurs ou actrices
C3HGG2102	mondialisation	métropole	\	\
C3HHD850101	Monarchie (absolue)	\	\	Louis XIV/le Roi-Soleil

Le calendrier et les supports de l'évaluation

La passation de l'évaluation CEDRE histoire, géographie et enseignement moral et civique en fin de collège a eu lieu en mai 2017. La méthodologie mise en œuvre par la DEPP en la matière, de la conception de l'évaluation à la passation des épreuves puis à l'analyse et à la publication des résultats, s'étale sur trois années selon le schéma suivant :

Tableau 4 : Les étapes de la réalisation de l'évaluation

1ère année	2ème année	3ème année
2015-2016	2016-2017	2017-2018
Étape 1	Étape 2	Étape 3
Expérimentation Passation mai 2016	**Évaluation Passation mai 2017**	**Analyse et publication des résultats**
Objectif: Prise en compte du cahier des charges. Création d'items • Items sur support papier • Items sur support numérique Expérimentation effectuée auprès d'un échantillon représentatif d'élèves.	Objectif: Analyse de l'expérimentation. Construction de l'évaluation à partir d'une sélection d'items • Items validés après l'expérimentation ; • Items d'ancrage (repris des évaluations précédentes). Evaluation effectuée auprès d'un échantillon représentatif d'élèves.	Objectif : Analyse des premiers résultats Publications : • Note d'information • Repères et références statistiques (RERS) • Etat de l'école

La conception et la sélection des items

La préparation des items fait intervenir des concepteurs qui sont des enseignants et/ou formateurs, ainsi qu'un représentant des corps d'inspection. Chaque item est le produit d'un travail collectif des concepteurs encadré par le chargé d'études. Un item proposé par un concepteur, pédagogue de terrain ayant une bonne connaissance des pratiques de classe, fait l'objet d'une discussion jusqu'à aboutir à un consensus. L'item est alors validé. Ce questionnement fait ensuite l'objet d'un premier test, anonyme et empirique, auprès d'une ou plusieurs classe(s) pour estimer sa difficulté, les durées de passation minimale, maximale et moyenne, et recueillir les réactions éventuelles des élèves. Tous les nouveaux items font l'objet d'une expérimentation plus précise sur le plan statistique, auprès d'un échantillon représentatif d'élèves (environ 2 000), l'année précédant la passation.

Un équilibre de proportion est recherché entre les items de niveaux de difficulté variés ("faciles", "moyennement faciles" ou "difficiles").

La constitution des cahiers et des modules

Afin de pouvoir évaluer un nombre important d'items sans allonger le temps de passation pour l'élève, CEDRE utilise la méthodologie des cahiers tournants. Les items sont ainsi répartis dans des blocs d'une durée de 25 minutes et les blocs sont ensuite distribués dans les cahiers tout en respectant certaines contraintes : chaque bloc doit se retrouver un même nombre de fois au total et chaque association de blocs doit figurer au moins une fois dans un cahier. Ce dispositif, couramment utilisé dans les évaluations bilans, notamment les évaluations internationales, permet d'estimer la probabilité de réussite de chaque élève à chaque item sans qu'il ait à répondre à l'ensemble de ceux-ci.

Au final, pour l'évaluation CEDRE histoire-géographie 2017, chaque cahier comprend deux séquences de 50 minutes, obligatoirement espacées d'une pause (au moins le temps d'une récréation). La seconde séance se termine par un questionnaire de contexte, d'une durée de 10 minutes environ, identique dans tous les cahiers, dans lequel l'élève doit répondre à des questions concernant notamment l'environnement scolaire, son intérêt et sa motivation pour l'histoire, la géographie et l'enseignement moral et civique.

L'évaluation CEDRE 2017 est constituée de 13 cahiers tournants intégrant un ensemble de 13 blocs d'évaluations contenant (tableau 5) des items de 2006 et 2012 repris à l'identique pour assurer une comparaison diachronique et de nouveaux items qui ont fait l'objet d'une expérimentation en 2016.

Pour garantir la qualité de la comparaison avec 2006 et 2012, notamment en termes de passation des épreuves, l'évaluation de 2017 s'appuie sur 195 items dont 108 d'ancrage soit 55 %.

En 2017, pour la première fois dans le cadre de CEDRE histoire-géographie, un second test, à titre d'expérimentation, a été opéré sur supports numériques (sur ordinateurs) auprès d'un second échantillon représentatif d'élèves. La passation a eu lieu au même moment que pour le test principal, papier-crayon, en mai 2017. La durée de passation était également identique : deux séquences d'une heure chacune, espacées d'une pause obligatoire, à savoir une heure de test sur ordinateurs et une heure sur papier. Ces items papier-crayon étaient repris du corpus de l'évaluation principale. Les résultats des élèves de cet échantillon spécifique, numérique, n'ont pas été pris en compte pour le calcul des scores de Cedre 2017 ni pour l'établissement de l'échelle de performances.

Tableau 5 : Exemple de répartition des blocs dans les cahiers

	Séquence 1		Séquence 2	
N° du cahier	1er bloc	2ème bloc	3ème bloc	4ème bloc
1	B3	B5	B2	B11
2	B4	B2	B6	B1
3	B5	B6	B9	B7
4	B2	B9	B10	B8
5	B6	B10	B11	B12
6	B9	B11	B1	B13
7	B10	B1	B7	B3
8	B11	B7	B8	B4
9	B1	B8	B12	B5
10	B7	B12	B13	B2
11	B8	B13	B3	B6
12	B12	B3	B4	B9
13	B13	B4	B5	B10

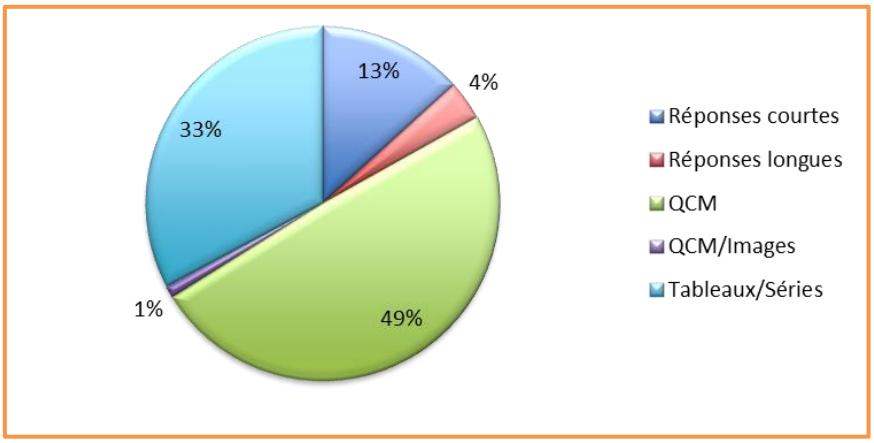
Les caractéristiques des items

Les items dits papier-crayon

L'enquête dite papier-crayon repose sur un corpus de 195 items réparti dans 13 cahiers. Pour simplifier, la réponse à une question constitue un item. Chaque élève répond aux questions d'un cahier, autrement dit durant deux séquences d'une heure chacune, espacées d'une pause obligatoire - au moins le temps d'une récréation.
Les élèves sont interrogés au plus près du terme de leur scolarité en collège (en mai). L'évaluation porte sur des attendus des programmes en vigueur, depuis leur classe de sixième et jusqu'à leur classe de troisième incluse. Mais les exigences des questionnaires sont bien sûr plus importantes pour des contenus correspondant au programme de troisième que pour ceux relevant des classes antérieures.

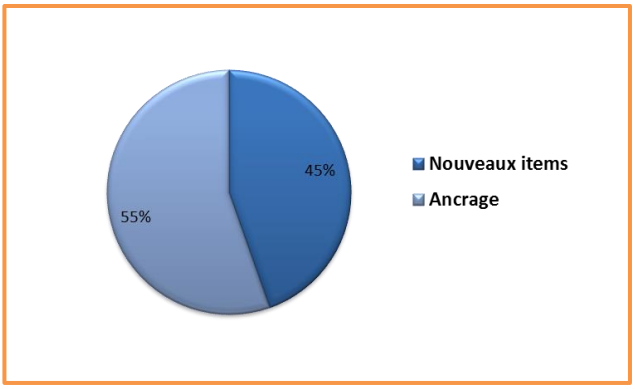
Quant aux formats d'items **FIGURE 1**, on note la part prépondérante des questions à choix multiple (QCM). Ils permettent de ménager la possibilité de coder rigoureusement les réponses, à des fins de standardisation, d'automatisation et, partant d'objectivation de la correction et de la mesure des résultats. Il faut cependant relever la part non négligeable des questions ouvertes, dont 4 % appellent des réponses longues, autrement dit une rédaction. Ces réponses aux questions ouvertes sont aussi traitées sous forme d'items, et donnent également lieu à un codage que des professeurs télécorrecteurs opèrent à partir d'un guide de correction le plus précis possible. Le logiciel de télécorrection mis au point par la DEPP permet également de prendre la mesure d'éventuels désaccords et écarts entre correcteurs, au moyen d'une double correction aléatoire, et d'arbitrer le cas échéant.

FIGURE 1 : Diagramme – les formats d'items



Pour garantir une comparabilité diachronique des résultats aux divers temps de la mesure, 55 % du corpus est constitué d'items dits d'ancrage **FIGURE 2**, autrement dit repris à l'identique des passations antérieures de 2012 voire de 2006. Cet ancrage de 108 items permet

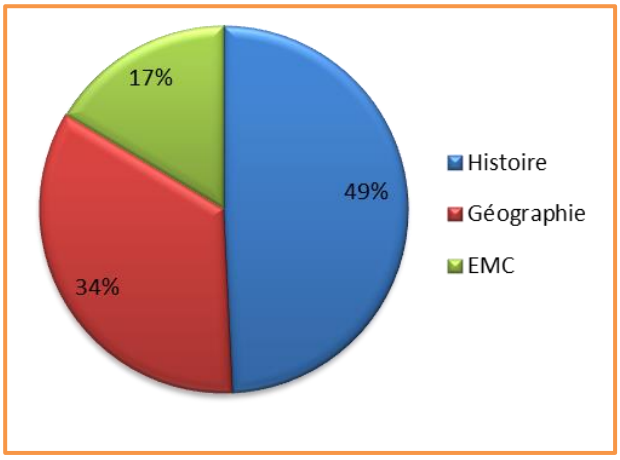
FIGURE 2 : Diagramme - l'ancrage



une comparaison particulièrement robuste des résultats pour le cycle 2012-2017 (108 items sur les 195 de l'évaluation au total), 73 items autorisant en outre une comparaison entre les résultats de 2006-2012-2017. L'enquête compte aussi 87 nouveaux items, créés spécifiquement pour faire place aux renouvellements des programmes et des approches didactiques depuis 2012.

En ce qui concerne la distribution des items entre les trois disciplines **FIGURE 3**, la part moindre de la géographie s'éclaire à la lumière des évolutions des programmes de référence depuis 2006. Au premier temps de la mesure comme en 2012 en effet, les programmes de référence étaient ceux de 1995-1998. En 2017 en revanche, les élèves ont été évalués à partir des programmes de 2008 pour leurs classes de sixième, cinquième et de quatrième, et des programmes de 2015, entrés en vigueur à la rentrée de 2016, pour leur année de troisième. Or, des « inflexions majeures », pour reprendre les termes de Catherine Biaggi[1], ont marqué les programmes de géographie de 2008, tant au niveau des approches que des thématiques. Ces changements importants ont conduit à périmer une part importante de l'ancrage (des anciens items) ce qui explique que la parité n'ait pas pu être atteinte avec l'histoire, malgré un renouvellement important du corpus en géographie - 55 % de nouveaux items -, autant qu'en enseignement moral et civique (EMC) qui renouvelle en profondeur l'ancienne éducation civique, depuis leur rentrée de quatrième (en 2015) pour les collégiens évalués en mai 2017.

Figure 3 : Diagramme - la distribution des items selon les disciplines



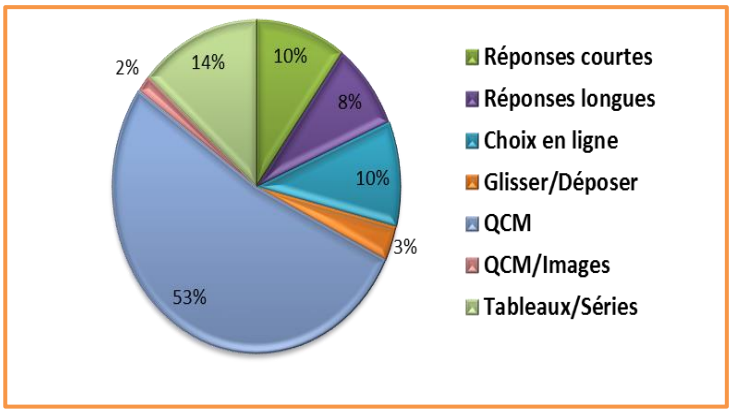
Les items numériques

En mai 2017, pour rappel, un second échantillon représentatif de collèges et de classes de troisième a été interrogé sur ordinateurs. Cet autre test, purement expérimental, n'a pas été pris en compte dans les résultats de Cedre 2017, pour ne pas biaiser la comparaison avec les précédentes enquêtes de 2006 et de 2012. L'effectif, pour ce test numérique, était de 80 classes dans autant de collèges, soit 2 000 élèves environ. Ces élèves ont étalement dû répondre à des items papier crayon repris du corpus de l'évaluation principale. Ainsi, après publication de la note d'information, lors de la phase d'analyses complémentaires, nous avons pu projeter les items numériques sur l'échelle de performances Cedre 2017, pour enrichir la description des différents niveaux de compétence. Les 51 items

[1] Inspectrice générale de l'Education nationale (IGEN), groupe histoire et géographie, « Habiter, concept novateur dans la géographie scolaire ? », dans *Annales de géographie*, 2015/4.

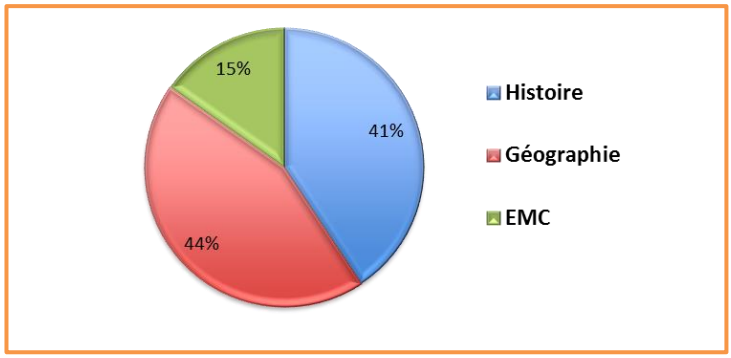
de cette enquête spécifique sont par définition nouveaux.

Figure 4 : Diagramme - les formats d'items numériques



La répartition est identique entre QCM et questions ouvertes **FIGURE 4**, même si on relève la palette nécessairement élargie des formats de questions fermées, du fait des opportunités du numérique. On atteint pour ces nouveaux items un équilibre entre histoire et géographie, et même légère une faveur pour celle-ci **FIGURE 5**.

Figure 5 : Diagramme - la répartition des items numériques entre les disciplines.



Le recours aux modèles de réponse à l'item (MRI).

La reprise d'anciens items (l'ancrage) ne garantit pas à elle seule la comparabilité des résultats dans le temps. Il s'agit aussi de s'assurer que le niveau de difficulté objective du test reste stable à chaque temps de mesure. Ainsi, le test CEDRE est inscrit dans un cadre conceptuel et méthodologique utilisant les modèles de réponse à l'item (MRI). Il s'agit de modèles statistiques et de leurs traductions graphiques mis en œuvre par les statisticiens de la DEPP, dans le cadre de CEDRE mais également d'autres évaluations ou enquêtes nationales et internationales comme PISA (Programme international pour le suivi des acquis des élèves) ou TIMSS (*Trends in Mathematics and Science Study*)… Ces modélisations révèlent la métrique de chaque item et, en particulier, ses paramètres de difficulté calculés à partir des taux de réussite effectifs de différents groupes de niveau, selon leurs scores globaux sur l'ensemble du test. La prise en compte de cette métrique et de ces paramètres permet d'observer la distribution des items sur une échelle de difficulté et, ainsi, de vérifier et de garantir une répartition rigoureusement équivalente aux divers temps de mesure.

L'échantillonnage

L'enquête principale – et la seule prise en compte pour les résultats publiés dans le cadre de CEDRE 2017 – a porté sur un échantillon représentatif de 195 collèges de France métropolitaine pour autant de classes de troisième. Il s'agit donc d'un tirage d'échantillon à deux degrés : les statisticiens ont d'abord veillé à la représentativité des collèges sélectionnés, puis à celle des classes au sein de ces établissements, selon trois strates **TABLEAU 5** : collèges publics hors éducation prioritaire, collèges relevant de l'éducation prioritaire puis établissements privés sous contrat… On veille également à l'équilibre de l'échantillon pour ces composantes : professions et catégories socioprofessionnelles (PCS), effectifs

Tableau 5 : l'échantillon CEDRE 2017 et les taux de participation selon ses trois states

Strate	Nb élèves attendus	Nb élèves répondants	% d'élèves répondants
1. Public hors EP	3 247	2 967	91.4 %
2. EP	699	565	80.8 %
3. Privé	1 117	1 009	90.3 %
Total	**5 063**	**4 541**	**89.7 %**

des garçons et des filles, des élèves doublants ou à l'heure dans leur scolarité.

Le taux de participation a été élevé en mai 2017, de l'ordre de 90 %, soit 189 collèges répondants sur les 195 visés par l'échantillon, comme c'est le cas traditionnellement pour les enquêtes CEDRE.

Pour achever de bien garantir la représentativité de l'effectif évalué, les statisticiens procèdent également à un redressement de la non-réponse après la passation. Il s'agit d'une pondération, au sens statistique, de l'effectif de certains groupes (élèves scolarisés en éducation prioritaire, élèves doublants…) dont on observe la propension à moins participer au test ou à moins répondre à ses questions. En d'autres termes, et très concrètement, les statisticiens attribuent un poids supérieur à ces groupes pour qu'ils soient dûment représentés dans l'effectif total des réponses.

Il faut rappeler qu'un second échantillon, indépendant, a été tiré suivant la même méthode pour le test numérique. Cette seconde enquête, expérimentale, a été opérée dans 80 autres collèges pour autant de classes, soit 2 000 élèves environ, sans différence notable en termes de participation par rapport à l'enquête principale, dite papier-crayon.

UN POINT SUR L'EVOLUTION DES ACQUIS DES ELEVES

Le dispositif d'évaluation Cedre histoire-géographie et éducation morale et civique existe depuis 2006. C'est donc la troisième passation (2006, 2012 et 2017). Ce dispositif permet une étude inédite de l'évolution des acquis des élèves en fin de collège entre 2006 et 2017.

La construction de l'échelle de performances

L'échelle de performances a été élaborée en utilisant les modèles de réponse à l'item. Le score moyen en histoire-géographie et éducation civique, correspondant à la performance moyenne des élèves de l'échantillon de 2006, 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 difficulté en fin de scolarité obligatoire, la partie la plus basse de l'échelle a été constituée, lors de l'évaluation de 2006, des scores obtenus par les 15 % d'élèves ayant les résultats les plus faibles. 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. On classe les items par ordre de difficulté croissante pour les répartir entre les différents groupes. Pour qu'un item soit attribué à l'un de ceux-ci, il faut que son élève le plus faible ait une probabilité supérieure ou égale à 50 % de le réussir. C'est ainsi que les élèves des groupes supérieurs réussissent les items associés aux groupes plus faibles. Les compétences de chacun des groupes d'élèves ont pu ainsi être caractérisées, puis analysées.

En 2006, en appliquant la méthode du modèle de réponse à l'item aux données issues de l'évaluation, on a attribué un score à chaque élève, ce qui a permis de les classer dans différents groupes de niveaux. Le groupe 5 réunit

arbitrairement 10 % des élèves ; les scores sont ainsi compris entre 317 et 473 points. À l'opposé, les groupes < 1 et 1 représentent 15 % des élèves, et les scores sont inférieurs à 199 points. Entre ces groupes, on observe un intervalle de 118 points. Les trois groupes intermédiaires ont ainsi des amplitudes de scores quasiment égales à 40 points. Pour distinguer le groupe 1 du groupe < 1, on reporte cette même amplitude de score sur le groupe 1, et les élèves du groupe < 1 sont donc ceux qui ont un score inférieur à 160 points (199 - 40). Une fois les groupes d'élèves déterminés, on attribue à chacun d'entre eux les items qui lui sont associés. Dire que des items sont associés à un groupe signifie que c'est à partir de ce groupe qu'on peut affirmer que les élèves maîtrisent les connaissances ou compétences requises pour répondre correctement à ces items. Ainsi, en utilisant la même méthodologie en 2017 (comme en 2012), à l'aide des données collectées auprès d'un nouvel échantillon d'élèves, une nouvelle échelle a été construite. Elle relève des mêmes compétences et des mêmes niveaux de difficulté qu'en 2006 et permet donc de comparer en particulier la taille des différents groupes.

L'évolution des compétences des élèves entre 2006, 2012 et 2017.

Alors que le score moyen des élèves a connu une baisse de 10 points entre 2006 et 2012, il progresse en 2017 et atteint 245 point. Cette hausse du score entre 2012 et 2017 doit être nuancée ; il reste inférieur à celui de 2006.

L'analyse de la répartition des élèves dans les différents groupes de compétences montre une augmentation de la proportion des élèves appartenant aux groupes les plus performants (groupe 3 et 4) malgré un groupe 5 qui reste stable, trois points en dessous de la part en 2006. Cette évolution est concomitante à une baisse de la proportion d'élèves dans les groupes les plus fragiles. Cette variation positive depuis 2012 révèle que davantage d'élèves maitrisent les compétences fondamentales en histoire et géographie.

La description des compétences de chaque groupe est précisée dans la **FIGURE 6** et permet de mieux appréhender le contraste entre les compétences extrêmement limitées du groupe 1, et l'ampleur de la culture historique, géographique et civique accumulée par les groupes 4 et 5. Toutefois, les groupes 2 et 3, qui rassemblent la majorité des élèves, ont déjà acquis des compétences essentielles à la fin de leur scolarité au collège.

Un glissement général des performances vers le haut, sans atteindre le niveau de 2006.

Pour analyser l'évolution des acquis des élèves, suivant la même méthodologie qu'en 2006 et 2012, une échelle décrivant six niveaux de performance a été construite. On y observe pour le collège une hausse de 5 points du score moyen qui passe de 240 à 245 points. Il reste cependant en dessous de 5 points du score de 2006.

On remarque sur la **FIGURE 7** que les groupes les moins en réussite voient leur proportion baisser (groupe <1, groupe 1 et groupe 2). La part des élèves appartenant au groupe 3 a, quant à elle nettement progressé, exprimant un glissement des élèves des groupes les plus faibles vers des groupes plus en réussite.

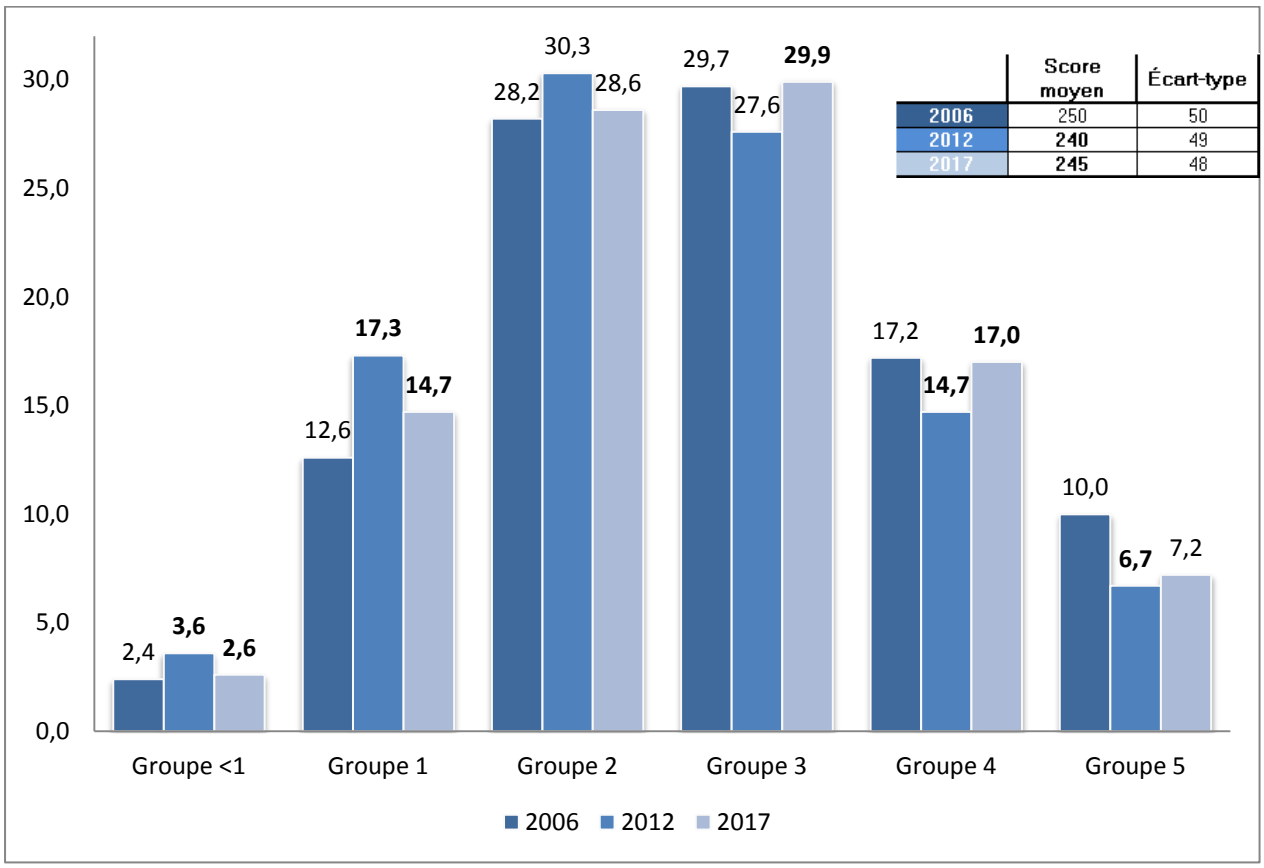
FIGURE 6 : Échelle de performances 2017 en histoire-géographie et enseignement moral et civique

% Population	
Groupe 5 **7,2 %**	317 — 511 Les élèves du groupe 5 ont acquis de manière très approfondie l'ensemble des connaissances et des compétences construites par l'enseignement de l'histoire-géographie et une culture civique déjà solide pendant leurs quatre années de scolarité au collège. Ces élèves vont directement au sens, à l'analyse critique, y compris à partir de leur lecture de textes complexes. De plus, ils répondent plutôt bien aux questions ouvertes : le travail de rédaction est plus abouti en histoire, de même que les tâches cartographiques.
Groupe 4 **17 %**	278 — 317 Grâce à leurs connaissances approfondies, y compris sur les programmes des classes antérieures à la troisième, les élèves du groupe 4 font preuve d'une très large maîtrise des compétences évaluées, au sein desquelles l'équilibre devient vraiment probant entre tâches ou capacités et connaissances. Ils identifient ou définissent le vocabulaire spécifique et notionnel plus complexe en étant capables d'en appréhender et d'en distinguer largement les composantes et les implications. Ils font preuve d'une très large maîtrise des repères historiques du programme de 3e. Les élèves de ce groupe synthétisent des documents variés, longs et complexes… Ils donnent du sens et ils comprennent la portée d'un document ou d'une situation en repérant souvent ce qui relève de l'implicite. Ils confrontent ces documents entre eux. Ils peuvent justifier, argumenter ou critiquer des affirmations ou un point de vue. Ces élèves commencent à appréhender la démarche des disciplines et à l'intégrer dans leur analyse. Ils passent d'un langage à l'autre.
Groupe 3 **29,9 %**	239 — 278 Ce groupe d'élèves maîtrise l'ensemble des compétences (capacités ou tâches et connaissances) fondamentales en histoire, géographie et en enseignement moral et civique. Ces élèves font preuve d'une bien meilleure maîtrise notionnelle et lexicale que les groupes de plus faible niveau. Ils savent identifier les notions usuelles et commencent à les caractériser . Leurs connaissances s'exercent non seulement sur le programme de troisième mais aussi sur certains contenus relevant d'enseignements suivis depuis la classe de cinquième au moins. Cependant leur maîtrise des repères historiques et géographiques est encore incomplète. Ces élèves parviennent non seulement à identifier des informations, explicites ou implicites, mais aussi à traiter ou interpréter le matériau prélevé. La maîtrise de la compétence "identifier" est en effet très bonne, étayée par les connaissances lexicales, notionnelles et thématiques de ces élèves. Par ailleurs, la lecture de textes longs ou de propositions de réponses longues et complexes n'est plus un frein à la réussite. Pour traiter l'information, les élèves de ce groupe gèrent des documents et des tâches complexes. Ils donnent du sens aux documents et parviennent à généraliser : intituler, dégager l'idée principale d'un texte... Ils perçoivent les différences de points de vue. Ils peuvent reconnaître un récit d'historien dont ils distinguent certaines caractéristiques (argumentation qui prend appui sur des sources, effort d'objectivité...).
Groupe 2 **28,6 %**	199 — 239 Les élèves du groupe 2 ont un taux moyen de réussite à l'ensemble des items sensiblement plus élevé que celui du groupe 1. Ils restituent des connaissances correspondant surtout au programme de troisième. Ils en reconnaissent le lexique spécifique ou en identifient les bonnes définitions parmi plusieurs propositions, de manière encore restreinte mais nettement plus étendue que le groupe 1. Leurs connaissances apparaissent en effet plus développées en enseignement moral et civique. Ils ont des acquis parcellaires sur des enseignements antérieurs à la troisième. La réussite de ces élèves s'exerce moins nettement dans le champ procédural (des capacités ou tâches) que sur des connaissances plus spécifiquement. Ils savent cependant exercer leur compréhension ou mobiliser des connaissances sur tous les supports documentaires. Ils commencent à traiter voire à interpréter (très rarement) les informations prélevées de ce matériau documentaire.
Groupe 1 **14,7 %**	160 — 199 Les élèves du groupe 1 ont des connaissances et capacités très fragmentaires et restreintes. Leur compréhension du vocabulaire simple s'affirme cependant un peu par rapport au groupe <1. Ils sont capables de prélever quelques informations très explicites sur des supports très simples.
Groupe <1 **2,6 %**	160 Les élèves du groupe < 1 ne sont capables que de réponses ponctuelles et dispersées. Leurs problèmes de compréhension de l'écrit sur tous les supports accentuent encore leur difficulté à identifier des informations auxquelles ils ne peuvent le plus souvent donner sens.

Lecture : la barre grisée symbolise l'étendue croissante de la maîtrise des compétences du groupe < 1 au groupe 5 et la partie bleue de la barre traduit la plage de score du groupe. Les élèves du groupe 3 représentent 29,9 % des élèves. L'élève le plus faible de ce groupe a un score de 239 points et le score du plus fort est de 278 points. Les élèves de ce groupe sont capables de réaliser les tâches du niveau des groupes < 1, 1, 2 et 3 (partie grisée) mais ils ont une probabilité faible de réussir les tâches spécifiques aux groupes 4 et 5.

Champ : élèves de 3ème générale scolarisés dans des établissements publics et privés sous contrat en France métropolitaine.
Source : CEDRE, MEN-DEPP

FIGURE 7 : Score moyen et répartition (en %) selon les groupes de niveaux en 2006, 2012 et 2017



Six groupes d'élèves aux compétences encore très inégales

Les élèves ont été interrogés à partir de trois compétences générales : « Identifier », « Traiter l'information » et « Interpréter » **TABLEAU 6**. Il est impossible de comparer la réussite des élèves selon chaque compétence. En effet, l'étude ne peut certifier le même niveau de difficulté pour les items mobilisant la compétence « identifier » et ceux mobilisant la compétence « traiter l'information ».

Néanmoins, on observe, d'une part, une progression sensiblement régulière des performances entre les six groupes. D'autre part, on remarque que ce n'est qu'à partir du groupe 3 que les élèves maitrisent une majorité d'items, et ce, quelle que soit la compétence en lien avec l'exercice.

Tableau 6 : Pourcentages de réussite par compétences pour chaque groupe en 2017

Compétences	Groupe <1	Groupe 1	Groupe 2	Groupe 3	Groupe 4	Groupe 5
Identifier	0,153	0,27	0,404	0,543	0,674	0,801
Traiter l'information	0,173	0,318	0,458	0,612	0,722	0,834
Interpréter	0,154	0,265	0,383	0,526	0,673	0,802

Une progression plus importante des filles.

Malgré une progression du score des filles, les garçons restent plus performants. Ainsi, le score moyen des garçons est, en 2017, de 246 tandis que celui des filles est de 244 **TABLEAU 7**. Cependant, ce différentiel tend à se résorber. En effet, le score des filles s'est particulièrement amélioré depuis 2012. Il est passé de 238 à 244 points. La proportion des filles dans les groupes « en réussite », du groupe 3 au groupe 5 a progressé, avec une évolution particulière dans le groupe 4 où elle passe de 13.4% en 2012 à 16.6% en 2017. Dans le groupe 3, elle dépasse les chiffres de 2006.

En 2006, la différence de score entre filles et garçons était de 6 points, en 2012 il était de 4 points, et il est aujourd'hui de seulement 2 points. Dans les groupes les plus en réussite, du groupe 3 au groupe 5, les garçons étaient 59% en 2006 et les filles 54,9. En 2017, 55.2% des garçons et 52.9% des filles appartiennent aux groupes les plus en réussite. On passe alors de 4.1 points d'écart à 2.3 points.

Des écarts de performances toujours élevés entre les élèves en retard et à l'heure.

En 2017, les élèves « en retard » représentent 14,9 % des élèves enquêtés et 36,1 % d'entre eux sont dans les groupes en difficulté (< 1 et 1) **TABLEAU 6**. Le score moyen de ces élèves est resté stable entre 2012 et 2017, tout comme celui des élèves à l'heure.

Les écarts de score moyen entre les élèves « en retard » et les élèves « à l'heure » restent très importants (38 points).

TABLEAU 7 : Score moyen en 2006, 2012 et en 2017 selon l'indice de position sociale moyen de l'établissement

	Année	Répartition (en %)	Score moyen	Écart-type	Groupes					
					<1	1	2	3	4	5
Garçons	2006	49,0	253	52	2,9	12,2	26,0	29,6	17,8	11,6
Garçons	2012	49,8	**242**	50	3,7	**16,4**	28,9	27,7	16,0	**7,4**
Garçons	2017	50,0	246	50	2,8	15,0	**27,0**	29,8	17,5	7,9
Filles	2006	51,0	247	48	1,8	13,0	30,3	29,7	16,7	8,5
Filles	2012	50,2	**238**	48	**3,4**	**18,1**	31,7	27,5	**13,4**	**5,9**
Filles	2017	50,0	**244**	47	**2,4**	**14,4**	30,3	29,9	**16,6**	6,4
Elèves en retard	2006	31,9	224	40	5,2	22,7	37,8	24,9	7,7	1,6
Elèves en retard	2012	23,4	**213**	28	7,0	**30,2**	39,2	**18,6**	**4,1**	1,0
Elèves en retard	2017	14,9	213	37	6,2	29,9	39,8	19,9	3,7	0,5
Elèves à l'heure	2006	68,1	262	49	1,0	7,9	23,7	31,9	21,6	13,9
Elèves à l'heure	2012	76,6	**248**	49	**2,5**	**13,3**	**27,6**	30,3	**17,9**	**8,4**
Elèves à l'heure	2017	85,1	**251**	48	2,0	12,1	26,7	31,6	19,4	8,3

Lecture : les filles représentent 50 % de l'échantillon en 2017 contre 50,2 % en 2012. Leur score a augmenté de 6 points entre les deux derniers cycles d'évaluation, passant de 238 à 244 points. 14,4 % d'entre elles appartiennent au groupe de niveau 1 en 2017 contre 18,1 % en 2012. Les évolutions significatives entre deux évaluations successives sont indiquées en **gras**. Par le jeu des arrondis, les totaux des pourcentages en ligne peuvent être légèrement différents de 100 %.
Champ : élèves de 3ème générale scolarisés dans des établissements publics et privés sous contrat en France métropolitaine.
Source : CEDRE, MEN-DEPP

Un déterminisme social encore marqué.

À partir des professions de leurs parents, les élèves peuvent être caractérisés par un indice de position sociale moyen, qui permet d'étudier l'évolution des performances des élèves selon le niveau social des écoles et des collèges. Pour chaque échantillon de 2006, 2012 et 2017, la moyenne de cet indice permet d'appréhender de manière simple la composition sociale de l'établissement. Quatre groupes sont alors constitués selon l'indice moyen, des collèges les moins favorisés (premier quartile), où figure l'éducation prioritaire aux plus favorisés (quatrième quartile).

L'analyse des scores moyens en histoire, géographie et éducation civique selon ces quatre groupes permet de relever quelques évolutions significatives entre 2012 et 2006 **TABLEAU 8**.

D'une part, le score moyen des élèves a progressé dans chacun des quartiles, même si cette progression est moins importante dans le premier quartile (4 points contre 6 points dans les autres quartiles). Dans ce quartile, l'évolution la plus nette est celle du groupe 3. Ainsi, on remarque un glissement des groupes les moins performants vers le groupe 3 sans qu'ils n'atteignent le groupe 4 (en baisse) et le groupe 5 (stable).

Ensuite, dans le quatrième quartile, c'est le groupe 4 qui est en net progrès. Il passe de 19.6% en 2012 à 18.5% en 2017.

Enfin, il est important de souligner l'importance des groupe les plus faibles <1 et 1 et 2 dans les collèges les moins favorisés. Ils sont en effet nettement plus représentés dans les collèges du premier quartile que dans ceux du quatrième quartile (59,8 % contre 30,3 % en 2017). Le niveau scolaire des élèves reste encore profondément, et de plus en plus, marqué par leur origine sociale. Ainsi, le score moyen progresse à mesure que le niveau social augmente et la différence de score moyen entre les élèves du premier quartile et ceux du dernier quartile reste toujours aussi importante. Cet écart a même tendance à s'accentuer (32 points en 2006, 34 points en 2012 contre 36 points en 2017).

Tableau 8 : Score moyen en 2006, 2012 et en 2017 selon l'indice de position sociale moyen de l'établissement

Indice moyen de l'établissement	Année	Score moyen	Écart-type	Groupes					
				<1	1	2	3	4	5
Premier quartile	2006	233	48	4,5	19,7	33,9	24,3	11,8	5,8
	2012	**224**	46	6,6	**26,4**	32,5	21,2	10,2	3,1
	2017	228	47	4,8	24,5	30,5	**26,5**	9,8	3,9
Deuxième quartile	2006	246	46	1,3	12,9	33,3	28,6	15,5	8,4
	2012	**236**	46	**3,3**	17,5	32,9	27,2	14,4	**4,6**
	2017	242	45	2,3	14,9	32,2	30,3	15,3	5,1
Troisième quartile	2006	255	50	2,2	11,2	24,2	32,7	19,0	10,7
	2012	**241**	49	3,3	15,6	**32,6**	27,6	**14,2**	6,7
	2017	247	45	2,0	12,8	29,9	30,6	18,5	6,4
Quatrième quartile	2006	265	50	1,5	6,9	21,6	32,9	22,3	14,8
	2012	258	48	1,1	9,9	23,2	34,2	19,6	12,1
	2017	264	49	1,4	6,8	22,1	32,1	**24,3**	13,3

Lecture : en 2017, le groupe de niveau 3 rassemble 26,5 % des élèves du premier quartile (collèges les plus défavorisés selon l'indice de position sociale), contre 21,2 % en 2012. Les évolutions significatives entre deux évaluations successives sont indiquées en **gras**.
On utilise désormais un indice de position sociale croisé qui a été appliqué sur les données des évaluations précédentes.
Par le jeu des arrondis, les totaux des pourcentages en ligne peuvent être légèrement différents de 100 %.
Champ : élèves de 3ème générale scolarisés dans des établissements publics et privés sous contrat en France métropolitaine. Source : CEDRE MEN-DEPP