

L'évolution des performances en calcul des élèves de CM2 à trente ans d'intervalle (1987-2017)

- ▶ En 2017, la reprise d'une enquête initiée en 1987, portant sur le calcul en fin de CM2, permet de comparer les performances des élèves à trente ans d'intervalle. Cette enquête donne des résultats à quatre temps de mesure : 1987, 1999, 2007 et 2017. Entre 1987 et 1999, les performances des élèves avaient fortement baissé. Cette tendance s'était atténuée entre 1999 et 2007. Dix ans plus tard, on assiste à une baisse des performances des élèves moins marquée cependant qu'entre 1987 et 1999. Les écarts restent les mêmes selon l'origine sociale. Les garçons réussissent mieux que les filles sur toute la période, mais l'écart de performances se réduit en 2017. La baisse des performances des élèves en retard scolaire, moins nombreux en 2017 qu'en 1987, est plus marquée que celle des élèves « à l'heure ».

Léa Chabanon, Jean-Marc Pastor, DEPP-B2

▶ En 2017, la reprise du volet consacré au calcul dans le cadre de l'enquête « Lire, écrire, compter » fait apparaître une baisse significative des scores obtenus depuis trente ans. En 1987, l'enquête « Lire, écrire, compter » a évalué les performances d'un échantillon d'élèves de fin de CM2 en lecture, orthographe et calcul. La direction de l'évaluation, de la prospective et de la performance (DEPP) a reproduit ces enquêtes en 1997, 2007 et 2015 pour la partie « lire et écrire », et en 1999, 2007 et 2017 pour la partie « compter ». Pour chaque point de mesure, de nouveaux échantillons représentatifs ont été tirés au sort. Ces reprises permettent de mesurer l'évolution des acquis des élèves en fin de CM2 à douze, vingt et trente ans d'intervalle. Les procédures utilisées sont standardisées : les consignes de passation et de correction des épreuves ont été répercutées à l'identique pour les quatre moments d'évaluation.

d'établir des comparaisons entre les quatre temps de mesure, grâce à des modèles statistiques adaptés.

Les évaluations 2007 et 2017 sont constituées des items communs suivants :

- sept additions présentant les variables didactiques suivantes : utilisation de nombres entiers (6) et de nombres décimaux (1), nécessitant les retenues (6), comportant plus de deux lignes (3), comportant des trous (2). Cinq additions sont proposées, dont deux sont à poser et effectuer ;
- sept soustractions : utilisation de nombres entiers (6) et de nombres décimaux (1), nécessitant les retenues (6), comportant des trous (2). Cinq soustractions sont proposées, dont deux sont à poser et effectuer ;
- cinq multiplications : à un, deux et trois chiffres au multiplicateur (les multiplicandes étant des nombres entiers) deux opérations en ligne mettant en jeu la multiplication par 100 et 1 000 avec en multipliant des nombres décimaux ;

- six divisions : deux à un chiffre au diviseur et deux à deux chiffres au diviseur, deux opérations en ligne mettant en jeu la division par 10 et 100 avec en multipliant un nombre entier pour la première et un nombre décimal pour la seconde ;
- quatre problèmes : trois ayant trait à la proportionnalité et un quatrième basé sur des calculs d'horaires.

En 2017, sont ajoutés 8 items complémentaires qui permettent de couvrir les exigences du programme en cours. Ils visent à apporter des informations supplémentaires sur la multiplication ou la division par les puissances de dix, les additions ou les soustractions utilisant systématiquement des retenues ou des nombres décimaux. Les analyses de ces items viennent compléter l'état des lieux de la maîtrise du calcul des élèves en fin de CM2.

Les épreuves

En 1987, les épreuves étaient constituées de 35 items pour le calcul couvrant les quatre opérations et proposant aux élèves trois problèmes de proportionnalité et un problème mettant en jeu des calculs sur les durées.

En 2007, 11 items d'une évaluation menée en fin de CM2 en 1999 ont été utilisés pour remplacer les items éliminés de l'épreuve de 1987. En 2007 et 2017, les épreuves ont été reprises des évaluations de 1987 et 1999 à l'exception de certains items de calcul « hors programme » comme la division de nombres décimaux ▶ **figure 1**. Il est ainsi possible

▶ 1 Dispositif d'évaluation

	Nombres d'items de l'évaluation	Nombre d'items créés à chaque point de mesure	Recouvrement des épreuves (Ancrage)
Évaluation 1987	35	35	22 items repris en 2007 et 2017
Évaluation 1999	30	11	11 items repris en 2007 10 items repris en 2017
Évaluation 2007	33	aucun	22 items de 1987 11 items de 1999
Évaluation 2017	40	8	22 items de 1987 10 items de 1999

Source : MENJ-DEPP, enquête « Lire, écrire, compter ».

Pour cette étude, la population de référence est celle des élèves de CM2 du secteur public scolarisés en France métropolitaine.

Des résultats comparables dans le temps

Des procédures et des épreuves identiques ne sont pas des conditions suffisantes pour garantir la comparabilité des résultats. Il convient également de distinguer ce qui est observé (la performance) de ce qui est visé (la compétence), des facteurs externes pouvant agir sur le niveau de performance indépendamment du niveau de compétence. Ainsi, il est possible que certains items s'avèrent plus difficiles ou plus faciles qu'il y a trente ans pour des raisons qui ne sont pas directement liées au niveau de compétences des élèves. Par exemple : le degré de familiarité avec certains types de tâches a évolué ; quelques

apprentissages ne sont plus exigés par les programmes ; des consignes de correction n'ont pas été appliquées à l'identique, etc.

Afin d'assurer la comparabilité temporelle des résultats, le modèle statistique dit « de réponse à l'item » a été employé. Ces types de modèle ont pour avantage de faire la distinction entre le niveau de compétences des élèves et les caractéristiques des items, notamment leur difficulté. Cette méthodologie est couramment employée dans les évaluations nationales, comme Cedre, ou internationales comme TIMSS et PISA.

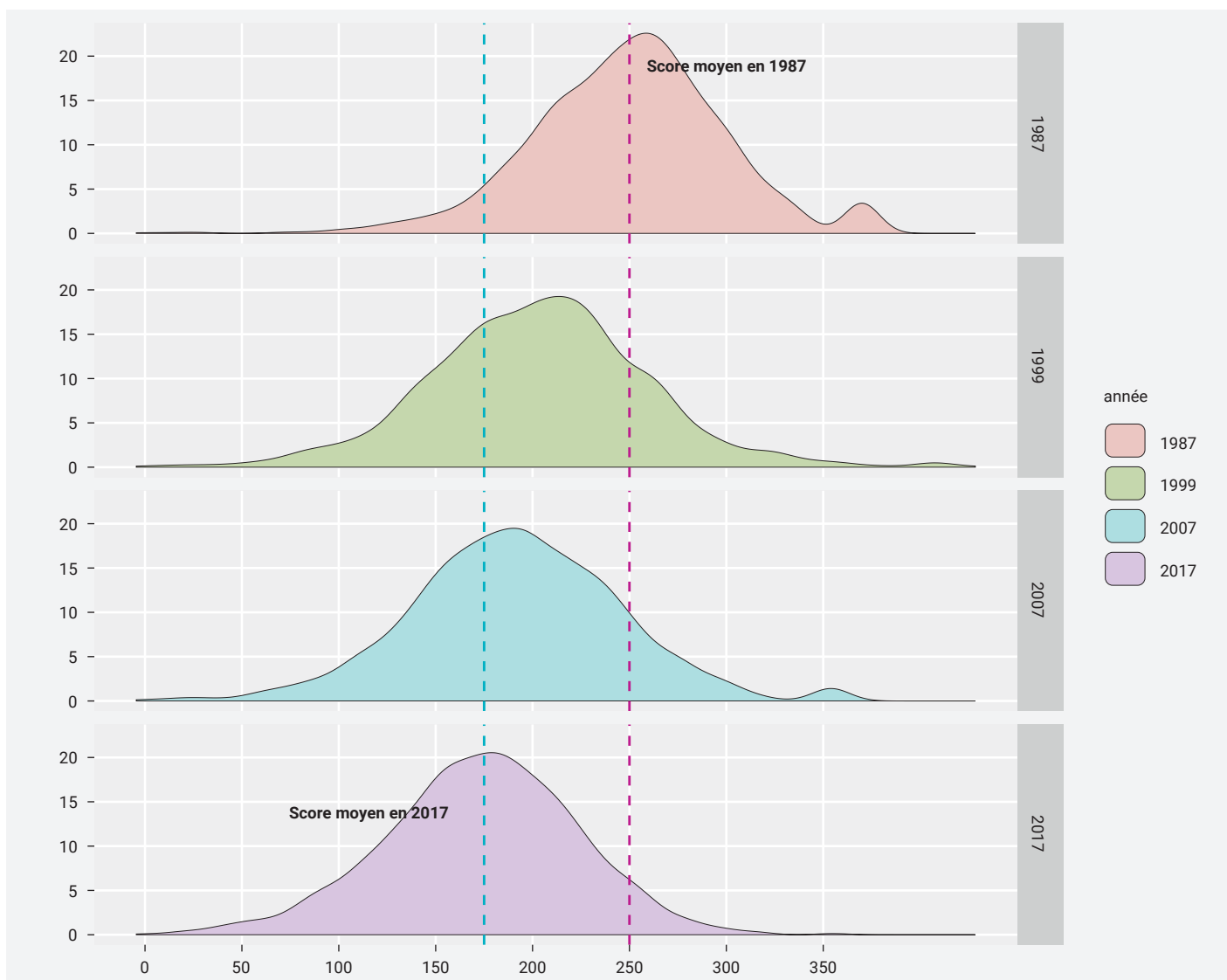
Sur les 22 items communs à l'évaluation de 1987, 7 items se sont avérés biaisés en 2017. L'évolution des programmes ou l'utilisation d'exercices datés et peu utilisés (additions ou soustractions à trous) expliquent en partie ces biais. L'étude comparative se base donc sur 15 items d'ancrage repris en 2007 et 2017 et 10 communs à 1999, 2007 et 2017. Elle a été complétée par 8 items exclusifs à 2017

permettant notamment une meilleure estimation des scores des élèves.

Des performances en baisse

L'année 1987 est utilisée comme référence caractérisée par une moyenne de 250 points et une dispersion autour de la moyenne de 50 points. Les figures 2 et 3 font apparaître une baisse significative des scores obtenus entre 1987 et 2017. Un décalage de la distribution des scores vers la gauche – de plus deux tiers d'écart-type soit 33 points de score – avait déjà été constaté en 1987 et 1999. La baisse des performances était beaucoup moins marquée entre 1999 et 2007 : le score moyen était en légère baisse, mais de manière peu significative, au regard des marges d'incertitude inhérentes aux enquêtes sur échantillons. Dix ans plus tard, on assiste à nouveau à une baisse des performances

► 2 Performances en calcul



Source : MENJ-DEPP, enquête « Lire, écrire, compter ».

Réf. : Note d'Information, n° 19.08. © DEPP

des élèves, de près d'un demi écart-type par rapport à 2007, soit 26 points de score.

La baisse des performances de 1987 à 2017 est caractérisée par un accroissement continu du pourcentage de la population dans le premier quartile de l'année de référence, constituant les scores les plus faibles et une diminution dans le dernier quartile (scores les plus élevés) ► **figure 3**. Ainsi, en 2017, plus de neuf élèves sur dix ont un niveau inférieur ou égal à la médiane des performances de 1987 et près de six élèves sur dix ont un niveau inférieur au premier décile.

L'observation des pourcentages de réussite aux items d'ancrage indique aussi une baisse générale des performances. Ainsi, les taux de réussite moyens pour les additions s'élevaient à 90 % en 1987 et à 77 % en 2007. En 2017, ils sont à 69 %. Pour les soustractions, réussies autour de 83 % en 2017, elles ne le sont qu'à hauteur de 64 % en 2007 et de 55 % en 2017. La baisse est encore plus marquée pour les divisions : 74 % en 1987, 43 % en 2007 et 37 % en 2017. Concernant les problèmes, les taux de réussite moyens sont de 32 % en 2017. En 1987 et en 2007, ils s'élevaient respectivement à 52 % et 40 %.

La non-réponse en hausse

L'analyse détaillée des opérations révèle une augmentation de la non-réponse. Or la non-réponse est comptée comme une réponse fautive. Ainsi, parmi les problèmes qui étaient présentés aux élèves, le taux moyen de non-réponse se situait autour de 4 % en 1987 ; il est de plus de 14 % en 2017. Le taux de réussite moyen était de 52 % en 1987 ; il est de 32 % en 2017. Ces résultats traduisent sans doute le fait que les élèves ne rentrent pas dans une tâche complexe nécessitant plusieurs étapes.

Par ailleurs, quatre opérations à trous étaient proposées aux élèves (deux additions et deux soustractions). Les taux de non-réponse constatés sont de 2 % en 1987 contre 15 % en 2017 pour les additions et de 6 % en 1987 contre 25 % en 2017 pour les soustractions. Par ailleurs, les taux de réussite de ces opérations sont très faibles en 2017 : autour de 50 % pour les additions et de 30 % pour les soustractions contre 85 % et 68 % en 1987.

Rapprochement des performances entre les filles et les garçons

L'évolution des performances suit la même tendance à la baisse chez les filles et les

garçons ► **figure 4**. Aux quatre temps de mesure, les garçons présentent de meilleures performances que les filles. Cependant, l'écart se resserre en 2017 (3 points). La baisse des performances constatée pour les filles en 2017 est donc moins importante que celle des garçons (voir « Pour en savoir plus »).

L'écart se creuse pour les élèves en retard scolaire

La pratique du redoublement a évolué au cours des trente dernières années. Celui-ci a fortement diminué et les maintiens dans un cycle deviennent exception.

► 3 Distribution et scores

Année	1987	1999	2007	2017
Score moyen	250	210	202	176
Écart-type	50	56	58	62
≤ 1 ^{er} décile 1987	10	34	39	56
≤ 1 ^{er} quartile 1987	25	57	62	79
≤ médiane 1987	50	79	84	93
≥ 3 ^e quartile 1987	25	8	6	2
≥ 9 ^e décile 1987	10	4	3	1

Source : MENJ-DEPP, enquête « Lire, écrire, compter ».

Réf. : Note d'Information, n° 19.08. © DEPP

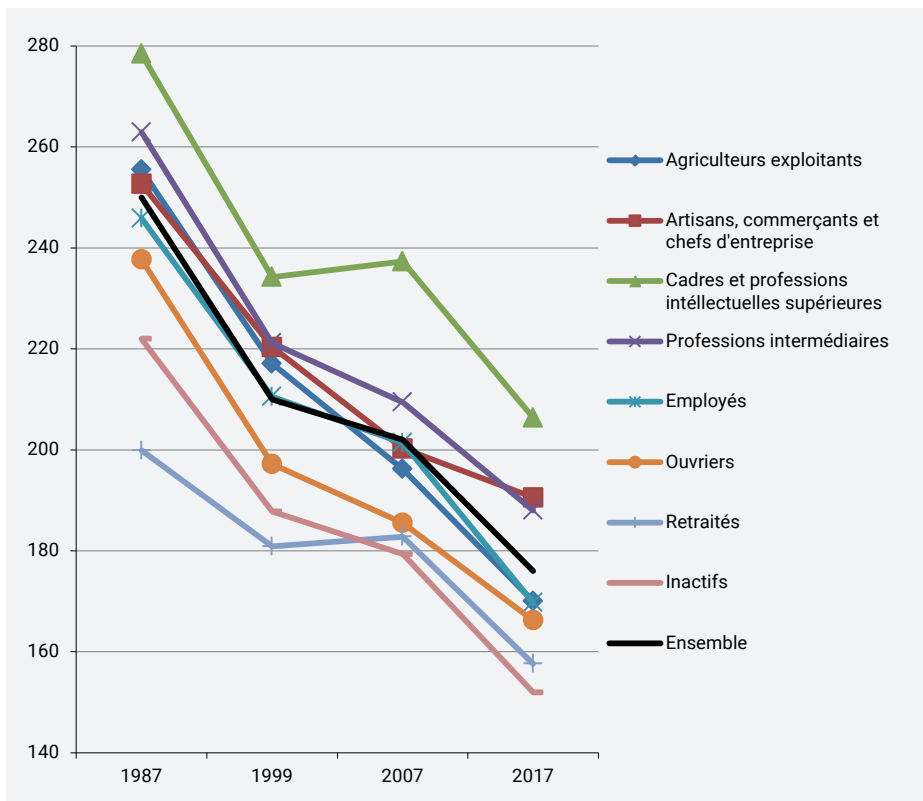
► 4 Scores moyens selon le sexe et le retard scolaire

Année	1987	1999	2007	2017
Garçons	251	213	208	182
Filles	248	206	199	179
à l'heure	262	218	210	184
En retard	226	174	166	130

Source : MENJ-DEPP, enquête « Lire, écrire, compter ».

Réf. : Note d'Information, n° 19.08. © DEPP

► 5 Scores moyens selon la PCS du chef de famille



Source : MENJ-DEPP, enquête « Lire, écrire, compter ».

Réf. : Note d'Information, n° 19.08. © DEPP

Le pourcentage de la population « en retard » était respectivement de 33 % et 16 % en 1987 et 2007 ; en 2017, il est de 9 %. Entre 1987 et 2007, on observe une baisse plus importante des performances pour les élèves « en retard » par rapport aux élèves « à l'heure » ce qui est à relier à la baisse des redoublements qui touche plus spécifiquement les élèves fragiles scolairement. En 2017, l'écart se creuse encore entre ces deux profils (54 points) avec une baisse plus marquée des scores des élèves en retard ► **figure 4**.

Des résultats en baisse quelle que soit l'origine sociale des élèves

Le recueil de la profession des parents permet de comparer l'évolution des différences entre groupes d'élèves selon leur origine sociale ► **figure 5**. Quelle que soit la catégorie observée, les résultats sont en baisse entre 1987 et 2017 et les inégalités sociales restent du même ordre de grandeur.

En effet, depuis trente ans, les scores sont les plus élevés parmi les élèves d'origine sociale favorisée (PCS du chef de famille « cadres et professions intellectuelles supérieures »). À l'opposé, les enfants issus des milieux les plus défavorisés (PCS du chef de famille « inactifs » ou « ouvriers ») présentent les performances les plus faibles. En lien avec le profil social des élèves, on observe, en 2017, un écart de 21 points entre les performances des élèves scolarisés hors éducation prioritaire (score moyen de 179 points) et ceux scolarisés en éducation prioritaire (score moyen de 158 points).

Les résultats présentés dans cette étude montrent que la tendance à la baisse des performances reprend après un léger « tassement » entre 1999 et 2007. La baisse est constatée à tous les niveaux de performance et pour l'ensemble des items présentés. Ces résultats sont en cohérence avec ceux issus de l'enquête internationale TIMSS qui montrent

notamment que les performances en mathématiques des élèves français en fin de CM1 en 2015 se situent en dessous de la moyenne européenne. ■

POUR EN SAVOIR PLUS

la *Note d'Information* 19.08 est en ligne sur education.gouv.fr/statistiques

Pour accéder aux **figures**, aux encadrés « **Méthodologie** » et « **Bibliographie** » ou à des informations complémentaires, voir la rubrique « Télécharger les données : tableaux et graphiques au format XLS ».