

En 2007, la reprise d'une enquête de 1987, portant sur la lecture, le calcul et l'orthographe en fin de CM2, permet de comparer les performances des élèves à vingt ans d'intervalle, à partir des résultats observés aux mêmes épreuves. De plus, des points intermédiaires de comparaison sont disponibles, en 1997 pour la lecture et en 1999 pour le calcul. Les résultats sont contrastés selon les domaines et les périodes. En lecture, les résultats sont stables de 1987 à 1997 ; en revanche, on observe une baisse significative du score moyen entre 1997 et 2007, plus prononcée pour les élèves les plus faibles. La situation est différente en calcul : une baisse importante des performances, touchant tous les niveaux de compétences, est observée de 1987 à 1999 ; puis, de 1999 à 2007, les résultats stagnent. Concernant l'orthographe, le nombre d'erreurs, essentiellement grammaticales, constatées à la même dictée a significativement augmenté de 1987 à 2007.

Lire, écrire, compter : les performances des élèves de CM2 à vingt ans d'intervalle 1987-2007

Le dispositif général

En 1987, l'enquête « Lire, écrire, compter » [1]* a évalué les performances d'un échantillon d'élèves de fin de CM2 en lecture, en calcul et en orthographe. La DEPP a reproduit cette enquête en 2007 auprès d'un nouvel échantillon d'élèves. Cette reprise permet de mesurer l'évolution des acquis des élèves de fin de CM2 à vingt ans d'intervalle. Les premiers résultats sont présentés dans cette Note d'Information.

Les épreuves

Les épreuves de 1987 étaient composées de 40 items (questions) pour la lecture (compréhension de textes courts et variés), de 33 items pour le calcul (opérations et petits problèmes), d'une dictée d'environ dix lignes et de dix phrases à compléter pour l'orthographe et la grammaire. En 2007, les épreuves ont été reprises à l'identique, à l'exception de certains items de calcul, hors programmes aujourd'hui, comme les divisions de nombres décimaux par exemple. Entre 1987 et 2007, des points intermédiaires de comparaison sont disponibles en lecture ainsi qu'en calcul. En effet, le « volet lecture » de l'enquête avait déjà fait l'objet d'une reprise en 1997, dont les résultats

ont été publiés par la DEPP en 1998 [2]. L'analyse avait montré une grande stabilité des performances en lecture entre 1987 et 1997. Concernant le calcul, onze items d'une évaluation menée en 1999 en fin de CM2 [3] ont été utilisés pour remplacer les items éliminés de l'épreuve de 1987. Bien qu'il n'y ait pas d'items communs entre 1987 et 1999, il est possible d'établir des comparaisons entre les trois points 1987-1999-2007, grâce à des modèles statistiques adaptés.

Les élèves

En 1987 comme en 1997, la population visée était celle des élèves de CM2 du secteur public en France métropolitaine. Par la suite, les élèves du secteur privé ont été inclus dans les échantillons des enquêtes de 1999 et 2007. Les comparaisons effectuées dans cette Note portent donc uniquement sur les élèves du secteur public. À chaque date, les échantillons sont composés de plusieurs milliers d'élèves – entre 2 500 et 4 500 selon les périodes – représentatifs de la population visée.

Les procédures

Les procédures sont standardisées : les consignes de passation et de correction des épreuves ont été reprises à l'identique aux différentes périodes. De plus, en 2007, la correction a été centralisée et fiabilisée, grâce à des procédures de doubles corrections.

Le tableau 1 résume l'économie générale du dispositif. L'encadré méthodologique détaille les différents aspects de cette enquête.

La comparaison des résultats

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). En effet, des facteurs externes peuvent 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 vingt ans pour des raisons diverses, sans rapport avec le niveau de compétences des élèves. Par exemple : le degré de familiarité avec certains types de tâches a évolué ; des apprentissages ne sont plus exigés par les programmes ; des consignes de correction n'ont pas été appliquées à l'identique, etc.

Les modèles de réponse à l'item

Afin d'assurer la comparabilité des résultats, des modèles statistiques adaptés sont employés : les modèles de réponse à l'item. Ces modèles ont pour avantage de faire la distinction entre le niveau de compétences des élèves et les caractéristiques des items, notamment leur difficulté. Cette propriété

TABEAU 1 – Le dispositif

Année	Domaines			Secteurs	Élèves
1987	Lecture (40)	Calcul (22)	Orthographe	Public	N=1994
1997	Lecture (40)			Public	N=3108
1999		Calcul (11)		Public/privé	N=2482
2007	Lecture (40)	Calcul (33)	Orthographe	Public/privé	N=4188

Lecture : en 1987, 1 994 élèves du secteur public ont été évalués en lecture, en calcul et en orthographe. En lecture, les 40 mêmes items ont été posés en 1987, 1997 et 2007. Pour le calcul, l'épreuve de 2007 comporte 33 items : 22 items de 1987 et 11 de 1999.

Source : MEN-DEPP

permet de repérer les items « biaisés » : en l'occurrence, il s'agit des items dont la difficulté n'est pas la même en 1987 et en 2007, pour des élèves de même niveau de compétence.

Ces modèles sont également utilisés pour les évaluations internationales, dont les problématiques sont similaires à celles évoquées ici, quand il s'agit d'évaluer, pour un pays donné, les items biaisés par rapport à la dimension commune, visée dans l'ensemble des pays.

Élimination des items biaisés

En lecture, avant l'analyse des biais, cinq items ont été éliminés en raison de problèmes d'édition ou de formulation. Ces items avaient déjà été repérés et retirés de l'analyse en 1997. Sur les 35 items restants, 16 d'entre eux ont révélé des fonctionnements différentiels significatifs, pour au moins une des trois comparaisons possibles (1987-1997, 1997-2007, 1987-2007). Après examen de ces items, les causes des biais semblent généralement liées à la

correction des réponses des élèves. En effet, pour certaines réponses d'élèves, dont le statut – réussite ou échec – n'est pas évident, les choix de correction pris en 1987 et en 1997 ne sont pas connus. Ces items ont été retirés de l'estimation des scores. En calcul, 5 items se sont avérés biaisés entre 1987 et 2007 et 4 autres entre 1999 et 2007. L'évolution des programmes explique en partie ces biais. Par exemple, une division dont le résultat attendu est un nombre décimal, désavantage les élèves de 2007. Ces items ont été éliminés du calcul des scores. Au final, 24 items de 2007 ont servi à l'ancrage des échelles. De plus, des items de calcul de 1987 (13 items) et de 1999 (11 items) ont été pris en compte, pour une meilleure estimation des scores des élèves de ces deux cohortes.

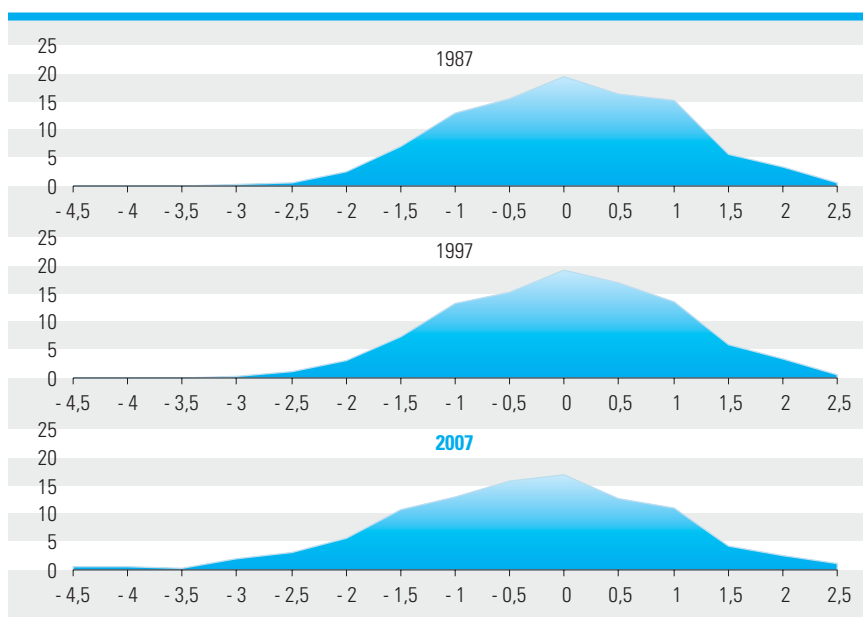
Les épreuves d'orthographe, quant à elles, n'ont pas fait l'objet d'une modélisation. Les résultats sont donnés ici de manière descriptive.

Résultats globaux

Lecture

Le graphique 1 représente les distributions des scores des élèves en lecture aux trois dates : 1987, 1997 et 2007. Les distributions

GRAPHIQUE 1 – Les performances en lecture



Source : MEN-DEPP

TABEAU 2 – Lecture

	1987	1997	2007
Moyenne	0,00	-0,03	-0,37
Écart-type	1,00	1,02	1,22
≤ 1 ^{er} décile 1987	10 %	11 %	21 %
≤ 1 ^{er} quartile 1987	25 %	26 %	37 %
≤ médiane 1987	50 %	51 %	61 %
≥ 3 ^{ème} quartile 1987	25 %	23 %	18 %
≥ 9 ^{ème} décile 1987	10 %	10 %	8 %
garçons	-0,01	-0,08	-0,39
filles	0,01	0,02	-0,31
« à l'heure »	0,25	0,13	-0,18
« en retard »	-0,50	-0,71	-1,26

N.B. L'année 1987 est prise comme référence pour les comparaisons : les scores ont pour moyenne 0 et écart-type 1 en 1987. Une valeur négative de la moyenne indique une valeur inférieure à la moyenne des scores de 1987.

Source : MEN-DEPP

de 1987 et de 1997 sont proches : cette stabilité des résultats en lecture avait déjà été constatée dans le rapport publié en 1998 [2]. En revanche, il apparaît que la distribution des scores des élèves de 2007 est décalée vers les niveaux de compétences les plus faibles, par rapport aux deux dates précédentes.

Les données synthétiques présentées dans le tableau 2 confirment ce constat : la moyenne des scores obtenus est stable de 1987 à 1997, puis baisse de 1997 à 2007 (diminution de plus d'un tiers d'écart-type). Cette baisse est plus marquée pour les élèves les plus faibles. Ainsi, deux fois plus d'élèves (21 %) se situent en 2007 au niveau de compétence des 10 % d'élèves les plus faibles de 1987. À l'opposé, les meilleurs élèves sont moins concernés par cette tendance : ils sont encore 8 % en 2007 à dépasser le niveau que les 10 % les meilleurs dépassaient en 1987. Ces observations sont à rapprocher du fait que la dispersion des résultats (l'ampleur des écarts entre élèves) est plus élevée en 2007 qu'il y a dix ou vingt ans (écart-type de 1,2 en 2007 contre 1 en 1987 et en 1997). Des analyses plus détaillées semblent montrer que la baisse moyenne s'observe quel que soit le type de compétences mises en jeu (compréhension immédiate, construction d'informations et de significations, etc.). En revanche, pour les questions

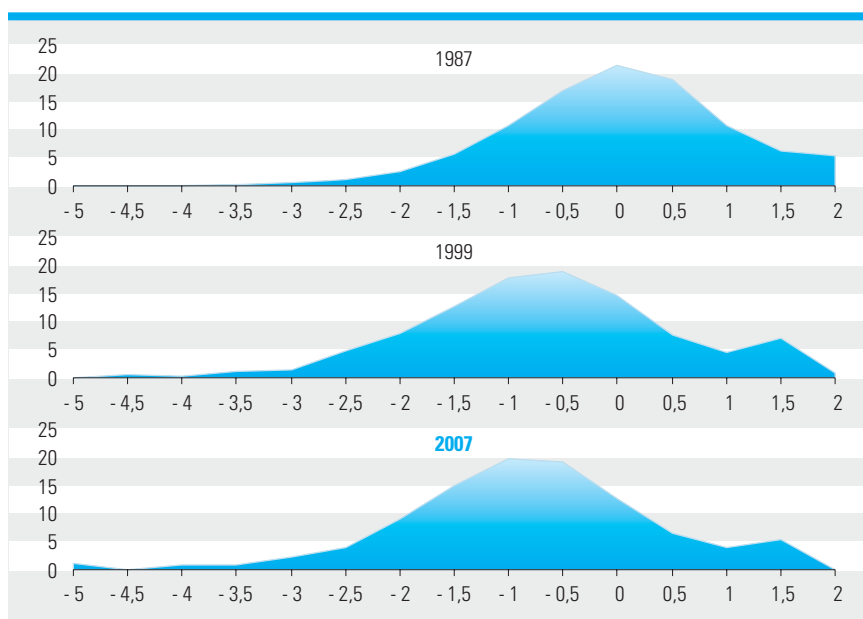
nécessitant la rédaction d'une réponse construite, le taux de non-réponses augmente de manière constante de 1987 à 2007. Ce phénomène a déjà été mis évidence, notamment grâce aux récentes évaluations internationales PISA et PIRLS. Il renvoie au fait que les élèves français ont tendance à s'abstenir de répondre aux questions exigeant un effort de rédaction.

Calcul

La situation est différente en calcul. Le graphique 2 et le tableau 3 font apparaître une baisse importante des scores obtenus en calcul entre 1987 et 1999 (le score moyen diminue d'environ deux tiers d'écart-type). Cette baisse touche tous les niveaux de compétences et s'accompagne d'un accroissement de la dispersion des scores (augmentation de l'écart-type de 1 à 1,2 entre 1987 et 1999).

De 1999 à 2007, il s'opère un « tassement » des résultats : le score moyen est en légère baisse, mais de manière peu significative, au regard des marges d'incertitude inhérentes aux enquêtes sur échantillons. Parallèlement, la dispersion se stabilise avec une légère baisse, là encore peu significative. Ce redressement est peut-être à mettre au compte de la remise à l'ordre du jour du calcul dans les programmes de 2002 avec, en particulier, l'accent mis sur le calcul mental et l'apprentissage des techniques opératoires.

GRAPHIQUE 2 – Les performances en calcul



Source : MEN-DEPP

Orthographe

La même dictée a été proposée aux élèves de 1987 et de 2007, à partir d'un texte d'une dizaine de lignes (85 mots et signes de ponctuation). Le nombre d'erreurs (nombre de mots mal orthographiés ou de ponctuations erronées) a augmenté en moyenne : de 10,7 en 1987 à 14,7 en 2007 (tableau 4). Le pourcentage d'élèves qui faisaient plus de quinze erreurs était de 26 % en 1987, il est aujourd'hui de 46 %. Ce sont principalement les erreurs grammaticales qui ont augmenté : de sept en moyenne en 1987 à onze en 2007.

Par exemple, 87 % des élèves conjuguait correctement le verbe « *tombait* » dans la phrase « *Le soir tombait.* » ; ils ne sont plus aujourd'hui que 63 % des élèves (tableau 5). En revanche, sur des conjugaisons difficiles pour les élèves de CM2, comme l'accord avec l'auxiliaire « avoir », le pourcentage de réussite n'évolue pas : environ 30 % des élèves, que ce soit en 1987 ou en 2007, écrivent correctement le verbe « *vus* » dans la phrase « *Elle les a peut-être vus !* ».

Résultats complémentaires

Inégalités sociales

Le recueil des professions des parents permet de comparer l'évolution des différences entre groupes d'élèves selon leur origine sociale. En lecture, les inégalités sociales ont tendance à se creuser : la baisse constatée entre 1997 et 2007 n'a pas touché les enfants d'origine sociale favorisée (PCS du chef de famille « cadres et professions intellectuelles supérieures ») (graphique 3).

TABLEAU 3 – Calcul

	1987	1999	2007
Moyenne	0,00	-0,65	-0,84
Écart-type	1,00	1,19	1,15
≤ 1 ^{er} décile 1987	10 %	28 %	32 %
≤ 1 ^{er} quartile 1987	25 %	51 %	57 %
≤ médiane 1987	50 %	75 %	80 %
≥ 3 ^{ème} quartile 1987	25 %	13 %	10 %
≥ 9 ^{ème} décile 1987	10 %	8 %	4 %
garçons	0,01	-0,59	-0,76
filles	-0,02	-0,71	-0,87
« à l'heure »	0,24	-0,48	-0,67
« en retard »	-0,48	-1,37	-1,59

N.B. L'année 1987 est prise comme référence pour les comparaisons : les scores ont pour moyenne 0 et écart-type 1 en 1987. Une valeur négative de la moyenne indique une valeur inférieure à la moyenne des scores de 1987.

Source : MEN-DEPP

TABLEAU 4 – Nombre et type d'erreurs obtenues à la dictée

Nombre d'erreurs	1987	2007	Type d'erreurs	1987	2007
Moyenne	10,7	14,7	Ensemble	10,7	14,7
Écart-type	8,4	9,2	Erreurs lexicales	2,1	2,6
≤ 2 erreurs	13 %	6 %	Erreurs grammaticales	7,1	10,8
≤ 5 erreurs	31 %	16 %	Erreurs de ponctuation	1,1	0,9
≥ 15 erreurs	26 %	46 %	Autres (oubli de mots, ...)	0,3	0,4
≥ 25 erreurs	6 %	12 %			

Lecture : en 1987, les élèves ont fait 10,7 erreurs en moyenne, dont 7,1 erreurs grammaticales. 31% de ces élèves ont fait moins de 5 erreurs.

Source : MEN-DEPP

TABLEAU 5 – Résultats détaillés obtenus à la dictée

1987	99	99	87	97	97	99	96	95	46	91	
2007	100	96	63	97	96	99	90	98	32	96	
1987	96	70	93	74	74	93	96	83	99	67	97
2007	93	49	92	63	77	83	96	61	99	45	98
1987	94	97	91	93	97	61	70	91	98	98	97
2007	91	97	87	87	94	50	46	94	95	93	97
1987	88	75	90	89	99	93	48	89	92	92	
2007	72	57	83	78	98	88	44	80	86	92	
1987	99	99	61	54	93	66	96	99	98	97	
2007	98	98	44	46	89	45	91	99	97	98	
1987	93	92	98	99	65	91	88	92			
2007	93	90	98	99	57	87	88	95			
1987	83	96	86	68	32	89					
2007	88	94	85	65	29	94					
1987	76	97	90	80	97	90	92	71	92	89	
2007	66	93	89	66	90	95	91	63	91	88	
1987	99	98	94	89	82	74	94				
2007	99	98	88	80	77	63	97				

N.B. Les nombres indiqués au-dessous des mots sont les taux de réussite observés en 1987 et en 2007, c'est-à-dire les pourcentages d'élèves qui écrivent correctement le mot correspondant à chaque période. Ces taux sont arrondis à l'unité la plus proche. Ainsi, environ 99 % des élèves de 1987 et près de 100 % des élèves de 2007 ont bien écrit le premier mot de la dictée « Le ».

Source : MEN-DEPP

TABLEAU 6 – Nombre d'erreurs à la dictée selon quelques caractéristiques

		1987	2007
Retard scolaire	« À l'heure »	8,5	13,2
	« En retard »	15,1	22,6
Sexe	Garçons	11,6	15,7
	Filles	9,7	13,6
PCS chef de famille	Agriculteurs	10,8	18,1
	Artisans, commerçants	10,3	15,1
	Cadres, prof. intellectuelles	6,8	9,2
	Prof. intermédiaires	8,3	13,8
	Employés	11,6	15,6
	Ouvriers	12,7	17,6
	Sans activité	13,8	19,5

Lecture : en 1987, les élèves en retard ont fait en moyenne 15,1 erreurs à la dictée ; les enfants dont le chef de famille est employé, selon la nomenclature PCS de l'INSEE, en ont fait 11,6.

Source : MEN-DEPP

Ce résultat est cohérent avec le fait que la diminution des performances en lecture concerne plus particulièrement les élèves les plus fragiles. Dans une moindre mesure, ce phénomène s'observe également en dictée : l'augmentation du nombre d'erreurs est moins marquée pour les enfants d'origine sociale favorisée (tableau 6).

En calcul, la baisse importante des résultats moyens de 1987 à 1999 est générale, elle a touché toutes les catégories sociales, ce qui laisse supposer un effet principalement lié à l'apprentissage scolaire.

Retard scolaire

La pratique du redoublement a évolué au cours des vingt dernières années.

De nombreuses études ont montré son caractère inefficace et inéquitable [4]. Pour mémoire, le taux d'élèves en retard en fin de CM2 a fortement diminué de 1987 à 1997 : il est passé de 33 % à 19 % ; il s'est ensuite stabilisé autour de 16 % en 2007. En lecture, la baisse des performances moyennes concerne les élèves « à l'heure » comme les élèves « en retard », ayant redoublé au moins une fois au primaire (tableau 2). Cependant, l'évolution des résultats mérite d'être précisée. De 1987 à 1997, on observe une légère baisse des résultats, équivalente dans les deux groupes (« à l'heure » et « en retard »), alors que la moyenne générale est stable. Cette période de réduction massive des redoublements s'est ainsi accompagnée d'une stabilité des résultats d'ensemble. En revanche, entre 1997 et 2007, le taux de retard est quasi stable et le niveau de compétence baisse de manière générale mais, plus

particulièrement, pour les élèves en retard. Ce constat est cohérent avec le fait que la baisse touche plus spécifiquement les élèves les plus faibles, en lecture. En calcul, l'évolution est différente : la baisse massive observée de 1987 à 1999 est un peu plus marquée chez les élèves « en retard » (tableau 3). Enfin, en dictée, le nombre moyen d'erreurs a plus augmenté pour les élèves « en retard » que pour les élèves « à l'heure » (tableau 6). Globalement, que ce soit en lecture, calcul ou orthographe, les différences entre élèves « à l'heure » et « en retard » ont augmenté.

Garçons, filles

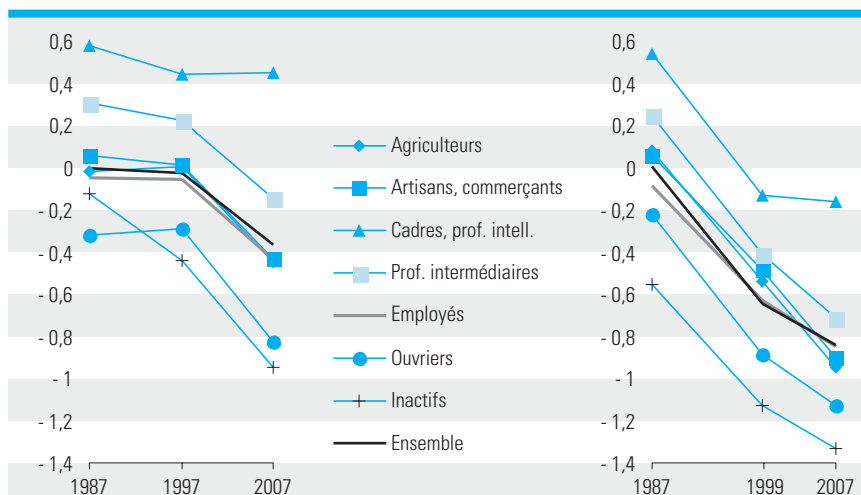
L'évolution des performances est la même chez les garçons et chez les filles. En lecture et en calcul, les différences sont ténues aujourd'hui comme il y a vingt ans, même si les garçons ont légèrement l'avantage en calcul, et les filles l'avantage en lecture (tableaux 2 et 3). C'est surtout en orthographe que les filles se distinguent : elles font en moyenne deux erreurs de moins que les garçons (tableau 6). Cependant, le nombre d'erreurs à la dictée a augmenté de façon identique entre 1987 et 2007, pour les filles comme pour les garçons.

Les résultats présentés ici sont à considérer comme de premières tendances, fondées sur des éléments objectifs qui garantissent la comparabilité dans le temps, notamment grâce à la répétition, à l'identique, de la mesure et aux précautions prises sur l'instrument et son fonctionnement.

Ces résultats sont cohérents avec ceux des récentes évaluations internationales PIRLS et PISA et doivent alerter sur l'augmentation du nombre d'élèves en difficulté dans le système éducatif français. Ainsi, l'enquête PISA fait ressortir une augmentation de la proportion de jeunes de quinze ans en difficulté de lecture. Si l'enquête PIRLS ne décèle pas de tendance significative en France ; les résultats des élèves de CM1 en lecture apparaissent médiocres par rapport aux résultats moyens des élèves européens.

Thierry Rocher, DEPP B2

GRAPHIQUE 3 – Résultats en lecture et calcul selon la PCS du chef de famille



N.B. Les points sont les scores moyens obtenus selon l'année de l'évaluation, pour chaque catégorie d'élèves. La catégorisation est celle de la profession et catégorie sociale (PCS) du chef de famille, selon la nomenclature de l'INSEE.

Source : MEN-DEPP

Exemples d'items de lecture

TEXTE A

Lorsqu'elle allait au marché, ma mère me laissait au passage dans la classe de mon père, qui apprenait à lire à des gamins de six ou sept ans. Je restais assis, bien sage, au premier rang et j'admirais la toute-puissance paternelle. Il tenait à la main une baguette de bambou : elle lui servait à montrer les lettres et les mots qu'il écrivait au tableau noir.

Un beau matin, ma mère me déposa à ma place, et sortit sans mot dire, pendant qu'il écrivait magnifiquement sur le tableau : « La maman a puni son petit garçon qui n'était pas sage. »

Tandis qu'il arrondissait un admirable point final, je criai : « Non ! Ce n'est pas vrai ! »

Mon père se retourna soudain, me regarda stupéfait, et s'écria :

« Qu'est-ce que tu dis ? »

- Maman ne m'a pas puni ! Tu n'as pas bien écrit ! ...

D'après « La gloire de mon père » de Marcel PAGNOL

	1987	1997	2007
1) Où se passe la scène qui est racontée dans le texte ?	95 %	95 %	90 %
4) Reproduis exactement ce que le père a écrit au tableau et seulement ce qui est écrit.	81 %	84 %	64 %

TEXTE G

Au XVII^e siècle, les ouvriers des villes travaillaient dans les ateliers des artisans ou dans les manufactures. C'est Louis XIV qui, avec l'aide de son ministre Colbert, mit en place le système des manufactures.

Les manufactures royales concentraient dans un même lieu un grand nombre d'ouvriers. Les machines n'existaient pas encore, et le travail se faisait à la main. Mais la quantité produite était plus importante que dans un atelier d'artisan, et la qualité facilement contrôlée.

Dans les manufactures royales, on fabriquait : des meubles, des tapisseries, à la manufacture des Gobelins ; des glaces à celle de Saint-Gobain. Ces produits étaient destinés aux Français les plus riches, ou aux pays étrangers.

	1987	1997	2007
1) Quel est dans ce texte le mot qui pourrait servir de titre ?	69 %	71 %	62 %
5) Vois-tu une différence entre les manufactures du temps de Louis XIV et les usines du XX ^e siècle ? Laquelle ?	54 %	46 %	43 %

N.B. Les taux de réussite diminuent entre 1987 et 2007. Pour la plupart des items, les taux de réussite sont proches de 1987 à 1997 et diminuent en 2007. Les exemples ci-dessus montrent que même les tâches les plus simples sont moins bien réussies, illustrant la diminution plus marquée des performances chez les élèves les plus faibles. En outre, l'item 5 du texte G illustre la réticence des élèves à produire une réponse construite : le taux de non-réponses pour cet item passe de 17 % à 28 % en vingt ans.

Pour en savoir plus

[1] Lire, écrire, compter au sortir de l'école élémentaire, *Éducation & formations*, n° 14, MEN-DEP, 1988.

[2] *La lecture en CM2 – Comparaison des résultats en lecture des élèves en fin de CM2 à dix ans d'intervalle (1987-1997)*, Les Dossiers, n° 102, MEN-DPD, novembre 1998.

[3] *Avant et après les vacances : évolution des acquis des élèves*, Les Dossiers, n° 158, MEN-DEP, septembre 2004.

[4] *Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mêmes constats*, Les Dossiers, n° 166, MEN-DEP, mai 2005.

www.education.gouv.fr

Exemples d'items de calcul

A2	S4	M2	
19 786 + 215 + 3 291	1 627 - 870 =	247 x 36	
D1	M4	P11	
74 14	1,54 x 1 000 =	30 morceaux de sucre pèsent 240 grammes 50 morceaux de sucre pèsent 400 grammes. a) 80 morceaux de sucre pèsent ... grammes	
Items	1987	1999	2007
A2	94 %		83 %
S4	92 %		81 %
M2	84 %		68 %
M4		61 %	58 %
D1		78 %	76 %
P11		70 %	71 %

N.B. les taux de réussite aux opérations sont systématiquement plus faibles en 2007 qu'en 1987. Les exemples donnés ci-dessus témoignent de cette baisse des performances pour les opérations les plus élémentaires : l'étude des réponses données par les élèves laisse apparaître de nombreuses erreurs de tables d'addition et de multiplication (A2, S4 et M2). Les différences sont plus ténues entre 1999 et 2007. Certaines difficultés, comme la multiplication d'un décimal par 1 000 (item M4), persistent. Des procédures-types continuent d'être employées. Ainsi, concernant la proportionnalité, peu d'élèves résolvent l'item P11 par linéarité ($30 + 50 = 80$), beaucoup se ramènent à l'unité, avec un risque plus élevé d'erreurs de calcul.

Méthodologie

Population

La population visée en 1987 était celle des élèves de CM2 scolarisés en France métropolitaine, dans le secteur public. Les comparaisons portent donc sur les élèves du secteur public. Il est à noter que le pourcentage d'élèves de primaire scolarisés dans le secteur privé est stable depuis vingt ans (environ 14 %). Par ailleurs, en 2007, les élèves du secteur privé obtiennent des résultats équivalents à ceux du secteur public hors éducation prioritaire, que ce soit en lecture, calcul ou orthographe.

Les échantillons

À chaque date d'enquête, les élèves ont été sélectionnés selon un plan de sondage « par grappes » : environ 150 écoles ont été tirées au sort et, dans chaque école retenue, tous les élèves de CM2 ont participé à l'évaluation (sauf pour 1999, une seule classe était évaluée dans le cas des écoles ayant plusieurs classes de CM2). En 1987, la sélection des écoles a eu lieu dans cinq académies uniquement, mais l'échantillon s'est avéré représentatif de la population-mère [1]. En 1987 et 1997, les tirages des écoles étaient stratifiés sur le caractère rural/urbain et l'éducation prioritaire. En 1999 et 2007, les tirages étaient stratifiés sur le secteur (public, éducation prioritaire, privé).

Les épreuves

L'épreuve de lecture se compose de huit textes courts et variés (littéraire, explicatif, descriptif, etc.). Pour chaque texte, les élèves doivent répondre à cinq questions de compréhension (compréhension littérale, globale, prise d'informations, etc.). Il s'agit uniquement de questions appelant une réponse rédigée, le plus souvent courte mais nécessitant parfois une construction plus élaborée. Les textes et les questions sont identiques aux trois périodes : 1987, 1997 et 2007.

L'épreuve de calcul porte sur les quatre opérations : une page pour chacune d'elles avec six à sept opérations posées, éventuellement à trous. L'épreuve comporte également quatre petits « problèmes ». Certaines questions de 1987 ont été exclues de l'épreuve en 2007 : un problème avec des pièces en francs n'a pas pu être adapté et des multiplications et des divisions de décimaux ont été écartées car elles sont aujourd'hui hors programmes. En remplacement, et pour disposer d'un point intermédiaire, onze questions de calcul ont été reprises d'une évaluation de fin de CM2 datant de 1999 [3].

La logistique

Les consignes de passation ont été reprises à l'identique. L'évaluation de 2007, comme celle de 1987, est organisée en cinq séquences (deux en lecture, deux en calcul, une en orthographe). Il est demandé de ne pas faire passer plus d'une séquence dans une demi-journée. Le taux de retour excellent (95 % en 2007) traduit le fort degré d'implication des écoles dans l'enquête.

Les consignes de correction ont également été reprises à l'identique. Mais à la différence des évaluations de 1987, 1997 et 1999 où la correction avait été assurée par les professeurs des élèves de l'échantillon, en 2007, les réponses des élèves ont été centralisées et traitées par télé-correction. Concrètement, plus de 300 000 réponses d'élèves ont été scannées et corrigées via un serveur Internet par une équipe de correcteurs recrutés par la DEPP. Ce système permet de fiabiliser les corrections et d'en mesurer la qualité. En effet, une image sur dix a été corrigée par deux correcteurs différents. La majorité des items présente un taux d'accord inter-correcteurs de plus de 95 %. Ce taux est plus variable en lecture : il tombe à 80 % pour certains items, le plus souvent à cause d'hésitations entre réussite totale et partielle, ce qui n'a que peu d'impact sur la mesure de la performance au final.