

Présentation des données statistiques

1. Introduction

1.1 Informations sur l'évaluation

L'étude EVAPM conduite au niveau des classes de sixième et de cinquième en 2008 a été conçue pour prolonger et compléter l'étude menée en sixième en 2005. L'étude de 2005 avait elle-même été conçue pour servir de point de repère, avant les modifications de programmes qui devaient avoir lieu à la rentrée 2006. Il s'agissait de rendre possible une évaluation des effets de ce changement sur les acquis des élèves.

Les études EVAPM sont nécessairement limitées dans leur ambition. Il n'est d'ailleurs jamais possible d'évaluer sur toutes les dimensions qui permettraient de rendre compte de l'ensemble des acquis des élèves. En 2005, nous avions laissé de côté le calcul mental et les démarches mentales ; dimensions que nous regroupons sous l'appellation de «gestion mentale d'informations mathématiques».

L'étude 2008 a fait une place particulière et importante à cette gestion mentale en faisant passer un questionnaire totalement oral (questions lues par l'enseignant) et un questionnaire dit « visuel », pour lequel les questions étaient rétroprojetées une par une. Dans les deux cas un minutage strict limitait le temps d'exposition de l'élève à chacune des questions et rendait difficile les retours en arrière.

Une autre particularité de l'étude 2008 consiste en un prolongement de l'étude Sixième à la classe de Cinquième. Une partie importante des questions ont ainsi été posées simultanément en classe de sixième et en classe de cinquième. Autant dire qu'EVAPM 2008 ne prétend pas avoir fait une étude des acquis des élèves en cinquième. Plus modestement, l'étude a été pensée pour pouvoir étudier les progrès faits entre la classe de sixième et la classe de cinquième sur des objectifs qui étaient déjà évaluables en sixième (compétences que les programmes supposent acquises ou en voie d'acquisition en fin de sixième).

L'objet du présent chapitre est de présenter les résultats de cette étude sur le plan quantitatif. L'organisation de l'étude telle qu'elle est rappelée ci-dessus justifie que soient présentés, ensemble, les résultats observés en Sixième et les résultats observés en Cinquième. Notre souci de relier les diverses études justifie aussi que, chaque fois que possible, l'on ait cherché à comparer les résultats 2008 avec les résultats des études précédentes.

L'étude porte sur 5 950 élèves répartis selon le tableau suivant :

ENSEMBLE	SIXIÈME		CINQUIÈME	
	Nombre de classes	Nombre d'élèves	Nombre de classes	Nombre d'élèves
ENSEMBLE	125	3 011	120	2 932
FRANCE	105	2528	107	2615
Établissements français de l'étranger	20	483	13	317

Rappelons que l'évaluation proprement dite était constituée de 4 épreuves en sixième et de 3 épreuves en cinquième. Chaque élève a passé deux épreuves d'une durée de 45 minutes. En sixième, pour l'épreuve écrite, les enseignants avaient la possibilité de choisir les épreuves, ce qui a introduit un biais, difficilement mesurable, mais qu'il convient de garder présent à l'esprit.

Pour des raisons techniques, nous avons préféré faire une analyse différenciée des résultats des élèves des lycées français de l'étranger. Un document de synthèse des particularités observées sur la population correspondante a été communiqué aux enseignants concernés et est accessible sur le serveur de l'APMEP. Sauf mention contraire, les classes des lycées français de l'étranger ne sont pas prises en compte dans les analyses présentées dans le présent chapitre.

1.2 Présentation des analyses statistiques

Notre plan d'évaluation a été conçu de façon à permettre des analyses de divers types : par domaine, en fonction des niveaux de complexité, des types de compétences, selon le type de passation,... Il permet aussi de différencier l'étude selon divers critères : âge, sexe, orientation, taille de la classe...

Compte tenu des conditions de l'étude, rappelées ci-dessus, la passation n'a pas été équilibrée entre les épreuves (par exemple, les résultats de l'épreuve D de cinquième portent sur 2435 élèves, tandis que ceux de l'épreuve C de sixième ne portent que sur 377 élèves). Ce fait, ajouté au caractère volontaire de l'inscription à l'étude et de l'affectation non aléatoire des épreuves aux élèves, fait que l'on ne peut qu'estimer un intervalle de confiance pour les résultats calculés. Pour étendre ces résultats à l'ensemble de la population des élèves de sixième, nous admettrons que l'intervalle de confiance des taux présentés est de $\pm 3\%$, au seuil de confiance de 95% ; mais cela est davantage basé sur l'expérience acquise en vingt ans d'études EVAPM que sur un calcul rigoureux.

Rappelons que notre souci n'est pas d'avoir des taux précis à 1 ou 2% près. Pour les conclusions que nous souhaitons pouvoir tirer de nos études, des valeurs approchées à 3 ou 4% près sont largement suffisantes. Nous cherchons en effet à obtenir des indicateurs et non des mesures. Si un exercice est réussi par 60% des élèves de l'échantillon, cela signifie qu'il serait réussi, dans l'ensemble de la population des élèves de sixième, par un pourcentage d'élèves compris entre 55% et 65% (en étendant encore un peu notre intervalle de confiance). Cela nous renseigne assez bien sur la capacité des élèves à traiter la question comme sur la difficulté de la question pour les élèves, tels qu'ils sont. On tirerait des conclusions de même type pour une question pour laquelle le taux de réussite observé serait, par exemple, de 20%.

En fait, toutes les études, qu'elles soient internationales ou nationales, génèrent des biais. Le biais le plus important étant d'ailleurs, en général, passé sous silence : il s'agit du biais introduit par la qualité des questions elles-mêmes. Biais difficile à contrôler et auquel nous n'échappons évidemment pas. De ce fait, les pourcentages de réussite observés pour une question renseignent autant sur la question elle-même que sur les compétences des élèves. Seule une analyse qualitative du type de celle qui est faite dans les autres chapitres de cette brochure permet de faire la part des choses.

Pour éviter de laisser croire que nous donnons dans le "tout mesure", rappelons encore que l'ensemble de nos analyses fait une large place à l'étude qualitative des résultats (examen systématique d'échantillons de copies d'élèves).

2. Le contexte et son évolution

Le tableau ci-dessous présente le suivi de quelques indicateurs, d'EVAPM Sixième 1987 à EVAPM Sixième 2005.

Classes de Sixième (105 classes)					
	EVAPM 2008	EVAPM 2005	EVAPM 1997	EVAPM 1989	EVAPM 1987
Nombre d'heures élèves en mathématiques (moyenne)	4,04	3,92	3,66	3,92	3,99
Nombre d'heures professeur, en mathématiques et par classe (moyenne)	4,21	4,07	4,04		
Nombre moyen d'élèves par classe	25,00	24,68	24,8 (24,6)	24,6	24,3
Moyenne scolaire annuelle en mathématiques (moyenne des moyennes)	13,10	12,86	11,81	11,58	
Élèves d'âge "normal" (nés en 1996 ou après pour EVAPM2008)	82,1%	77,7%	70,8%	60,9%	
Score évaluation nationale	65,5	62,0 (64,3)			

* Nombres entre parenthèses : *statistiques nationales (source DEP)*

Classes de Cinquième (105 classes)					
	EVAPM 2008	EVAPM 1990	EVAPM 1988		
Nombre d'heures élèves en mathématiques (moyenne)	4,04	3,94	3,95		
Nombre d'heures professeur, en mathématiques et par classe (moyenne)	4,16				
Nombre moyen d'élèves par classe	24,47	24,59	24,58		
Moyenne scolaire annuelle en mathématiques (moyenne des moyennes)	11,99	11,14	10,58		
Élèves d'âge "normal" (nés en 1995 ou après pour EVAPM2008)	77,90%	59,4%	52,1%		

Les valeurs prises par ces indicateurs restent assez stables dans le temps. On notera cependant un abaissement sensible de l'âge moyen des élèves, dû à la diminution des redoublements à l'école élémentaire et à la diminution des redoublements de la classe de Sixième.

On notera enfin que les notes scolaires continuent à augmenter (il s'agit, rappelons-le de la moyenne des notes de mathématiques communiquée par les enseignants).

Nous ne pouvons que reprendre et confirmer le constat fait lors de l'étude 2005 : par rapport aux études précédentes, les élèves de sixième sont plus jeunes, moins souvent redoublants ou menacés de l'être, et ont de meilleures notes de mathématiques. Ce dernier point montre qu'il est plus faux que jamais de penser que les mathématiques seraient une discipline sélective mettant en échec, dès le début du collège, une partie importante des élèves. Comme nous le verrons plus loin, cela ne signifie pas pour autant que les acquis des élèves soient supérieurs à ce qu'ils étaient précédemment.

En plus des données rassemblées dans le tableau précédent, les enseignants étaient invités à nous donner quelques informations complémentaires concernant leurs classes.

Ainsi, en plus des heures destinées aux classes entières, des heures de soutien en mathématiques sont affectées aux classes de sixième dans la moitié des cas et aux classes de cinquième dans le quart des cas (environ).

les enseignants considèrent que plus du quart des classes

Il y a des heures de soutien dans cette classe :		
	Sixième	Cinquième
OUI	51%	28%
NON	49%	72%

Le professeur estime que sa classe est :		
	Sixième	Cinquième
Plutôt homogène	15%	30%
Plutôt hétérogène	58%	40%
Très hétérogène	27%	29%

sont très hétérogènes ; seules 15% d'entre-elles en sixième et 30% en cinquième étant présentées comme plutôt homogènes.

Le professeur estime que sa classe est :		
	Sixième	Cinquième
Très bonne	14%	21%
Moyenne	75%	63%
Médiocre	11%	17%

Toutefois, les mêmes enseignants décrivent leurs classes comme étant « très bonnes », ou au moins « moyennes », tandis que la moitié des classes, en sixième comme en cinquième sont considérées comme travaillant « suffisamment », ou même, mais assez rarement, « beaucoup ».

Les divers tableaux présentés ici concernent les classes, et non les élèves. Ils dénotent qu'une grande majorité des professeurs portent un regard positif sur leurs classes. Très peu en effet déclarent que leur classe est « médiocre » et, même si dans environ la moitié des cas ils estiment que le travail est « insuffisant », ils n'estiment le travail inexistant que pour 3% des classes de sixième et pour à peu près 0% des classes de cinquième.

Les trois-quart des professeurs de sixième estiment avoir suffisamment de temps pour traiter le programme, mais il est symptomatique que cette proportion passe à un-tiers pour les professeurs de cinquième (qui sont parfois les mêmes que ceux de sixième).

Enfin, les deux tiers des groupes classes voient les cours de mathématiques se dérouler dans la même salle, ce qui semble dénoter une légère amélioration des conditions de travail des enseignants. Malheureusement, il reste environ 5% des salles de classes dans lesquelles le professeur n'a pas la place de circuler entre les bancs.

Le professeur estime que ses élèves travaillent :		
	Sixième	Cinquième
Beaucoup	0%	5%
Suffisamment	55%	44%
Insuffisamment	42%	51%
Pas du tout	3%	0%

Les cours de mathématiques en classe entière se déroulent dans :		
	Sixième	Cinquième
La même salle	62%	73%
Deux salles	17%	10%
Plus de deux salles	21%	17%

3. Résultats statistiques globaux

3.1 Scores par épreuves

Les tableaux suivants présentent les résultats épreuve par épreuve. En premier lieu, ces tableaux sont destinés à mettre en évidence la difficulté relative des épreuves. Pour ne pas alourdir la présentation, nous n'indiquons pas les écarts-types. Dans tous les cas, la dispersion est importante ; nous la retrouverons au niveau des classes.

Les tableaux présentent aussi les résultats selon le sexe. Sur l'ensemble des items, l'écart entre les filles et les garçons est d'environ 5 points de pourcentage en valeur absolue et de 10% en valeur relative (44,5% pour les garçons contre 39,9% pour les filles – la différence étant très significative). Ce résultat n'est pas nouveau, mais il mérite notre attention. D'une part, il est confirmé, pour la France, par de nombreuses études nationales et internationales (ce n'est pas le cas dans tous les pays), et, d'autre part, il est contredit par les notes scolaires de mathématiques qui placent systématiquement les filles au dessus des garçons.

Notons que si l'on se ramène à l'échelle normée réduite utilisée pour présenter les résultats de PISA, cette différence garçons-filles est à peu près égale à la différence observée pour PISA 2003 entre les résultats français et les résultats finlandais. Cela à la fois pour relativiser ce qui est couramment dit sur l'excellence de la Finlande et pour signaler que la question mérite d'être prise au sérieux.

Outre, et peut-être de façon plus utile que la variable sexe, d'autre variables pourraient être considérées. On sait déjà, par exemple, que les résultats sont très sensibles à l'âge des élèves : dans tous les cas, un retard scolaire d'un an fait baisser les résultats moyens de plus de 10%. Nous n'avons pas la possibilité de nous intéresser aux variables socio économiques ou culturelles, mais d'autres études mettent bien en évidence leur influence. Notons ici que les élèves des lycées français de l'étranger obtiennent des résultats supérieurs de 8 points de pourcentage à ceux des élèves de France et que, du moins au niveau Sixième, on n'observe pas, chez eux, de différence entre les garçons et les filles. L'échantillon étudié pouvant être biaisé, il serait hasardeux de généraliser, mais il y a sans doute là une piste de réflexion intéressante.

L'épreuve A est composée de QCM ; les épreuves B et C de sixième sont identiques aux épreuves B et C de 2005.

L'épreuve D est l'épreuve à passation orale (oral sans support visuel et visuel pur).

En ce qui concerne l'épreuve A, le score moyen de réussite aux items élémentaires (de nature dichotomiques) est un indicateur peu satisfaisant. En effet, il prend en compte de la même façon les résultats exacts obtenus par choix forcé, ou par hasard, et ceux correspondant à une vraie maîtrise de la question (bien que l'on sache que nos élèves ne répondent que rarement au hasard). Nous préférerons donc nous fier aux scores moyens des réussites conjointes. Chaque réussite conjointe à une QCM est en effet un signe de maîtrise de l'ensemble de la question. Cela explique le score relativement bas de cette épreuve. Ce score, qui est de 34%, serait de 63% si nous nous limitions aux réponses dichotomiques exactes.

L'épreuve B de cinquième est légèrement différente de l'épreuve B de sixième, tandis que les épreuves D de sixième et de cinquième ont en commun la moitié de leurs questions (20 questions sur 40). Le tableau suivant présente les résultats des épreuves ou parties d'épreuves passées par les seuls élèves de cinquième.

En ce qui concerne les questions du domaine « gestion mentale d'informations mathématiques » (épreuve D), les questions simplement rétroprojetées sont moins bien réussies en moyenne que les questions énoncées oralement. Cela est peut-être dû au fait que ce mode de présentation est assez peu fréquent. Pourtant la tendance, y compris pour les études internationales, est de privilégier ce mode de présentation sur ordinateur. Il serait donc judicieux d'y habituer les élèves, d'autant que cela permet d'obtenir des informations sur les conceptions des élèves et sur la disponibilité de leurs connaissances qu'il est difficile d'obtenir autrement.

L'épreuve C était commune aux niveaux Sixième et Cinquième, on la retrouvera plus loin.

PASSATION SIXIÈME SEULE						
	Épreuve A		Épreuve B_6ème		Épreuve D_6ème	
	Effectif	Score moyen ensemble de l'épreuve	Effectif	Score moyen ensemble de l'épreuve	Effectif	Score moyen ensemble de l'épreuve
TOUS	1182	34%	1186	44%	2363	52%
MASCULIN	571	37%	548	44%	1206	56%
FEMININ	598	31%	626	44%	1131	49%

PASSATION CINQUIÈME SEULE					
	Épreuve B_5ème		Épreuve D_5ème		
	Effectif	Score moyen ensemble de l'épreuve	Effectif	Score moyen ensemble de l'épreuve	
TOUS	1364	44%	2310	50%	
MASCULIN	679	45%	1125	53%	
FEMININ	685	44%	1185	48%	

Questionnaires Oraux : comparaison Oral sans support visuel et Visuel (attention, seule une partie des questions est commune aux deux niveaux)						
	SIXIÈME			CINQUIÈME		
	Effectif	ORAL	VISUEL	Effectif	ORAL	VISUEL
TOUS	2 363	59%	47%	2 310	48%	55%
MASCULIN	1 206	63%	50%	1 125	51%	56%
FEMININ	1 131	54%	45%	1 185	45%	55%

3.2 Scores par domaines

Les tableaux précédents mettaient en évidence que les difficultés de notre évaluation étaient assez bien équilibrées entre les diverses épreuves ; l'épreuve C, centrée sur les grandeurs étant toutefois plus difficile que les autres. On retrouve cet équilibre entre les différents domaines, sauf, encore, en ce qui concerne le domaine des grandeurs. L'équilibre entre épreuves ou domaines ne dit certes rien sur les acquis des élèves, mais il renseigne sur l'écart entre les attentes des concepteurs de l'évaluation, lesquelles reflètent les attentes des programmes, et les acquis des élèves. Le résultat faible dans le domaine des grandeurs traduit, on le sait, une difficulté spécifique à notre pays (difficulté qui se manifeste, en particulier, dans les études internationales). Il y a là une difficulté d'enseignement que le simple constat ne peut suffire à résoudre.

Résultats par domaine : taux de réussite moyen					
	Sixième			Cinquième	
	Nombre d'items	TOUS	GARÇONS	FILLES	Nombre d'items
Tous items	174	42%	45%	40%	150
Géométrie sauf espace	28	49%	49%	48%	24
Géométrie de l'espace	30	41%	42%	41%	24
Nombres	46	51%	55%	47%	36
Grandeurs	70	35%	38%	33%	66

4. Comparaisons internes à l'étude

Pour permettre des comparaisons, en particulier entre les scores d'élèves n'ayant pas passé les mêmes épreuves, les scores obtenus à EVAPM, ainsi que les valeurs prises par d'autres variables associées, ont été normalisés. Les distributions de scores sont donc ramenées à des distributions de moyenne 0 et d'écart-type 1. L'importance des échantillons étudiés autorise en général cette manipulation. On fait cependant l'hypothèse que les élèves seraient placés de la même façon sur l'échelle normée réduite quelles que soient les épreuves qu'ils auraient passées (et l'on sait que cette hypothèse n'est qu'imparfaitement vérifiée).

Ces scores réduits ont permis de calculer des indices, eux mêmes réduits, pour chaque variable étudiée.

Pour tenir compte de la complexité des traitements sollicités (complexité dite cognitive), nous utilisons une taxonomie dont les grandes catégories sont les suivantes (on trouvera sur le site de l'APMEP une présentation complète de la taxonomie) :

- A : Connaissance et reconnaissance.
- B : Compréhension.
- C : Application.
- D : Creativité.
- E : Jugement.

Une autre classification des questions concerne les niveaux des compétences tel qu'ils sont définis dans les études PISA, et que précisent le tableau suivant :

Niveaux de compétence : définitions du cadre de référence des évaluations internationales PISA		
1	Reproduction	Les compétences classées dans ce groupe impliquent essentiellement la reproduction de connaissances déjà bien exercées

2	Connexions (Mathématisation simple)	Les compétences du groupe connexions sont dans le prolongement de celles du groupe reproduction, dans la mesure où elles servent à résoudre des problèmes qui ne sont plus de simples routines, mais qui impliquent à nouveau un cadre familier ou quasi-familier.
3	Réflexion (Mathématisation complexe)	Les activités cognitives associées à ce groupe demandent aux élèves de faire preuve d'une démarche mentale réfléchie lors du choix et de l'utilisation de processus pour résoudre un problème. Elles sont en rapport avec les capacités auxquelles les élèves font appel pour planifier des stratégies de solution et les appliquer dans des situations-problèmes qui contiennent plus d'éléments que celles du groupe connexions, et qui sont plus « originales » (ou peu familières).

Compte tenu de l'absence dans notre évaluation de questions relevant du niveau 3, seuls les niveaux 1 et 2 apparaissent dans le tableau des résultats.

On trouve enfin dans le tableau les variables :

- Pour le niveau sixième, le score à l'évaluation nationale de début d'année, communiqué par les enseignants.
- Pour les deux niveaux, la moyenne scolaire. Il s'agit de la moyenne de mathématiques des deux premiers trimestres, communiquée par les enseignants..

EVAPM2008 - Synthèse SIXIÈME														
		ENSEMBLE des questions	Domaine numérique	Géométrie plane	Géométrie espace	Domaine grandeurs	Taxonomie : A	Taxonomie : B	Taxonomie : C	Taxonomie : D	Classe compétence : 1	Classe compétence : 2	Score évaluation début d'année	Moyenne deux trimestres
Nombre d'items		174	28	30	46	70	24	38	54	6	60	63		
Tous 100%		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	65,5	13,1
Garçons 51,5%		0,13	0,16	0,04	0,03	0,14	0,16	0,08	0,16	0,01	0,12	0,09	68,5	13,0
Filles 48,5%		-0,13	-0,16	-0,03	-0,03	-0,15	-0,17	-0,07	-0,16	-0,01	-0,11	-0,09	62,3	13,1
nés en 1994 0,8%		-1,19	-1,15	-1,04	-0,79	-0,64	-1,25	-1,10	-0,73	-0,54	-1,27	-0,77	38,9	7,7
nés en 1995 17,0%		-0,71	-0,65	-0,57	-0,45	-0,46	-0,67	-0,62	-0,58	-0,31	-0,69	-0,56	52,8	10,2
nés en 1996 79,9%		0,14	0,13	0,11	0,08	0,09	0,14	0,12	0,11	0,03	0,14	0,11	67,9	13,6
nés en 1997 2,2%		1,01	0,95	0,83	0,84	0,69	0,81	0,99	0,81	0,52	0,87	0,73	83,0	16,4
Redoublants 4,7%		-0,52	-0,49	-0,33	-0,36	-0,33	-0,52	-0,47	-0,43	-0,24	-0,45	-0,37	57,5	10,9
Redoublement prévu 6,4%		-1,00	-0,85	-0,93	-0,55	-0,77	-0,92	-0,82	-0,86	-0,67	-1,00	-0,78	46,1	8,1

Lecture des tableaux : par exemple : synthèse sixième, colonne ensemble des questions (i.e. tous les items) :

- La moyenne des scores de tous les élèves est 0 (ce qui traduit simplement le fait que les scores ont été normalisés).
- Les garçons ont alors un score moyen de 0,13, c'est à dire 13 centièmes d'écart-type au dessus de la moyenne dans une distribution de moyenne 0 et d'écart-type 1.

- Les filles ont, elles, un score moyen de - 0,13, c'est à dire 13 centièmes d'écart-type au dessous de la moyenne dans une distribution de moyenne 0 et d'écart-type 1.

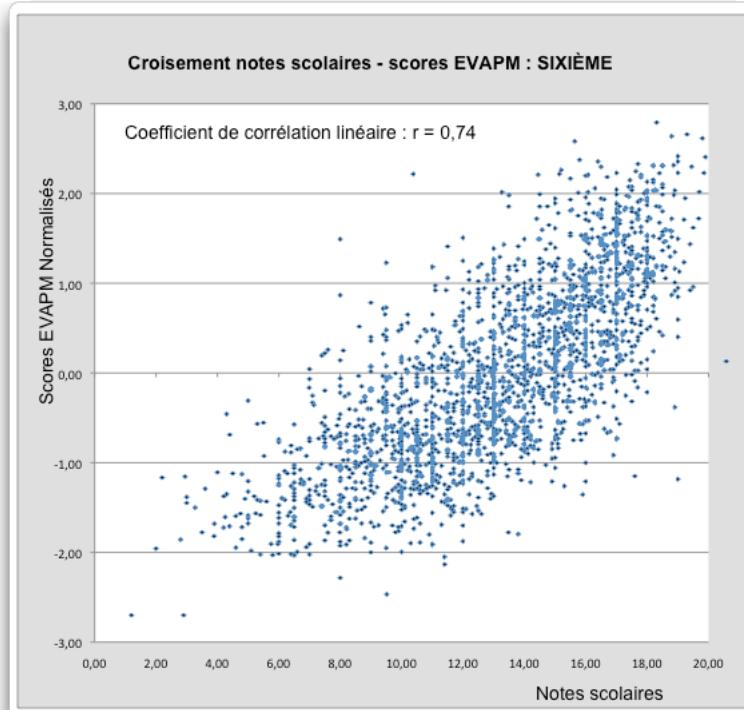
EVAPM2008 - Synthèse CINQUIÈME												
	ENSEMBLE des questions	Domaine numérique	Géométrie plane	Géométrie espace	Domaine grandeurs	Taxonomie : A	Taxonomie : B	Taxonomie : C	Taxonomie : D	Classe compétence : 1	Classe compétence : 2	Moyenne deux trimestres
Nombre d'items	150	24	24	36	66	14	35	46	9	48	57	
Tous	100%	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	12,0
Garçons	49,1%	0,11	0,12	0,04	0,03	0,06	0,05	0,05	0,15	0,03	0,09	11,8
Filles	50,9%	-0,11	-0,11	-0,04	-0,03	-0,06	-0,05	-0,05	-0,14	-0,02	-0,09	12,2
nés en 1993	1,8%	-0,68	-0,64	-0,46	-0,68	-0,34	-0,81	-0,55	-0,61	-0,26	-0,70	-0,46
nés en 1994	20,2%	-0,63	-0,60	-0,44	-0,36	-0,44	-0,51	-0,47	-0,61	-0,30	-0,56	-0,52
nés en 1995	74,6%	0,13	0,12	0,08	0,09	0,09	0,12	0,11	0,13	0,05	0,13	0,11
nés en 1996	3,3%	0,82	0,83	0,65	0,53	0,68	0,66	0,53	0,84	0,51	0,68	0,77
Redoublants	2,4%	-0,25	-0,19	-0,27	-0,22	-0,21	-0,16	-0,25	-0,20	-0,06	-0,15	-0,24
Redoublement prévu	3,1%	-0,67	-0,65	-0,54	-0,42	-0,42	-0,60	-0,42	-0,62	-0,27	-0,56	-0,62
												6,9

On retrouve le caractère plus discriminant, dans notre évaluation, du domaine numérique et, au moins en sixième, du domaine grandeurs, par rapport aux domaines géométriques. De même, on retrouve l'écart très important entre les élèves d'âge normal et les élèves ayant un an de retard : presque un écart-type. Puis que nous avons parlé des études internationales, c'est la différence pour PISA 2003 entre la France et la Thaïlande (pays particulièrement mal placé) !

Les domaines numérique et grandeurs creusent davantage l'écart entre les diverses catégories d'élèves que le domaine géométrique : écart garçons-filles, écart entre les élèves d'âge normal et les élèves en retard, entre les non redoublant et les redoublants... Cela confirme l'impression de difficultés spécifiques dans le domaine numérique, sans doute accentuées par notre souci de prendre largement en compte la question des grandeurs.

5. Relation avec les notes scolaires

Les professeurs des classes participant à l'étude nous ont communiqué les notes scolaires de leurs élèves (moyennes des deux premiers trimestres en mathématiques). On sait que tous



les enseignants n'ont pas la même échelle de notation et il est clair que cet indicateur est délicat à utiliser lorsqu'il s'agit de comparer des élèves ayant des professeurs différents.

Toutefois, dans la mesure où nous travaillons avec les notes d'un assez grand nombre de professeurs, on peut considérer que des compensations s'effectuent et que l'indicateur peut être pris en compte au moins pour traduire la façon dont les enseignants jugent leurs élèves. La figure illustre la corrélation entre la note annuelle (sur 20) et le score global aux épreuves EVAPM de sixième (score réduit). Le coefficient de corrélation, 0,74 pour le niveau sixième est assez élevé. Cela semble signifier que notre évaluation ne s'éloignerait pas trop des pratiques des enseignants. Pour le niveau cinquième ce coefficient de corrélation est de 0,63 ; ce qui est plus faible que ce que nous obtenons habituellement. Ce dernier point n'a pas de quoi surprendre dans la mesure où notre évaluation ne porte pas, pour l'essentiel, sur les apprentissages en cours en cinquième

Avec EVAPM et les notes scolaires, nous avons affaire à deux "mesures" indépendantes ; chacune des deux a vocation à rendre compte du "niveau" des élèves, mais chacune des deux laisse échapper des éléments de compétence et, sans doute, en valorise exagérément d'autres. Aucune des deux mesures ne peut prétendre être meilleure que l'autre. L'indicateur EVAPM a l'avantage d'une relative objectivité (un instrument unique est utilisé), l'indicateur "note du professeur" a l'avantage de la prise en compte, tout au long de l'année, d'éléments de jugement qu'EVAPM, par son caractère ponctuel et extérieur à la vie de la classe (au contrat didactique), n'est pas en mesure de prendre en compte.

Bien sûr, il est plus facile (et plus rigoureux) de travailler avec l'indicateur EVAPM. C'est ce que nous allons faire essentiellement dans la suite de ce chapitre. Toutefois, les remarques qui précèdent sont de nature à préciser le domaine de validité de nos observations et de nos conclusions.

Le tableau suivant présente, pour les niveaux sixième et cinquième, et pour quelques critères, les corrélations observées entre les scores EVAPM et la moyenne des notes scolaires, ainsi que, pour le niveau sixième, avec le score de l'évaluation de début d'année.

Corrélations des scores EVAPM avec les résultats scolaires				
		Sixième		Cinquième
Corrélation des scores EVAPM avec :		Notes scolaires	score évaluation de début d'année	Notes scolaires
ENSEMBLE (tous items)		0,74	0,79	0,63
DOMAINES	NAL : Numérique	0,68	0,77	0,58
	GES : Géométrie plane	0,61	0,59	0,48
	GEE : Géométrie espace	0,49	0,49	0,46
	GRA : Grandeur	0,51	0,59	0,47
NIVEAUX TAXONOMIQUES	A : Connaissance et reconnaissance	0,69	0,76	0,51
	B : Compréhension	0,64	0,65	0,48
	C : Application	0,63	0,72	0,60
	D : Creativité	0,40	0,42	0,31
Classes de compétences	C_1 : Reproduction	0,71	0,74	0,55
	C_2 : Connexions	0,61	0,63	0,56

Au niveau sixième, l'évaluation EVAPM est fortement corrélée avec l'évaluation nationale. C'est même là que la plus forte des corrélations est observée. Cela traduit le fait que, outre sa valeur diagnostique, l'évaluation de début d'année a aussi une valeur pronostique. Mais cela peut aussi révéler une mise en défaut de l'usage qui est fait du diagnostique. Nous laissons la question ouverte, mais ce peut être une piste intéressante de réflexion.

Les valeurs des corrélations entre certains critères de l'évaluation et les notes attribuées par les enseignants renseigne sur l'importance que les enseignants accordent, implicitement à ces critères. Ainsi, la corrélation nettement plus forte entre les notes scolaires (et l'évaluation nationale) et le domaine numérique d'EVAPM, qu'entre ces mêmes variables et le domaine des grandeurs indique que les notes scolaires prennent davantage en compte les compétences du domaine numérique que celles relatives aux grandeurs ; cela aussi bien en sixième qu'en cinquième. De même, comme pour les études précédentes, on observe que niveau D de la taxonomie (créativité, mise en œuvre d'idées personnelles, adaptation) est peu pris en compte.

Implicitement, les enseignants accordent donc, dans leur propre évaluation, plus d'importance aux compétences du domaine numérique qu'à celles du domaine géométrique (bien, et cela n'est pas contradictoire, que les élèves maîtrisent mieux les compétences du domaine géométrique). Par rapport à notre étude de 1997, la remarque faite en 2005 d'une meilleure prise en compte du niveau compréhension semble se confirmer.

Le tableau suivant présente d'autres types de corrélations (la légende est incluse dans le tableau précédent).

Comme pour les études précédentes, le tableau met en évidence des corrélations assez faibles entre les domaines de

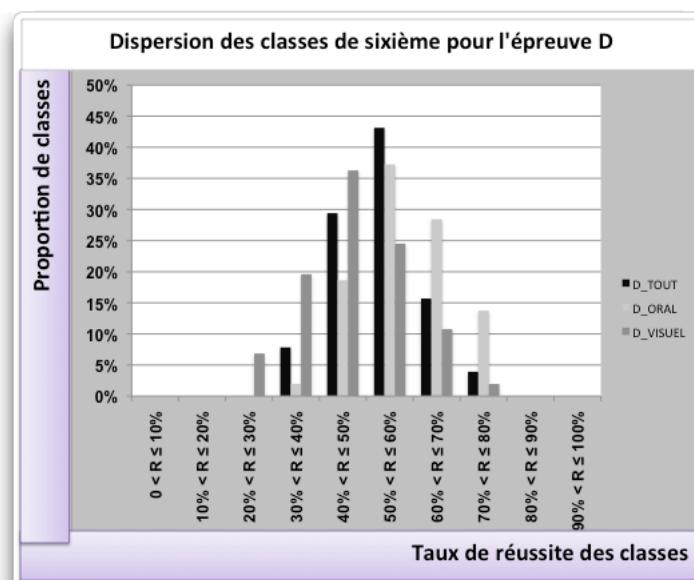
l'évaluation. Ceci n'est qu'une mise en évidence supplémentaire de la multi dimensionnalité des compétences mathématiques. Il est dès lors à la fois réducteur et trompeur de prétendre rendre compte de ces compétences en utilisant une échelle unique

Corrélations internes à l'étude							
SIXIÈME							
	GES	GEE	GRA	B	C	D	C_2
NAL	0,57	0,50	0,59				
GES		0,50	0,46				
GEE			0,39				
A				0,65	0,70	0,37	
B					0,63	0,35	
C						0,45	
C_1							0,78

CINQUIÈME							
	GES	GEE	GRA	B	C	D	C_2
NAL	0,44	0,45	0,68				
GES		0,42	0,29				
GEE			0,33				
A					0,56	0,66	0,33
B						0,60	0,37
C							0,47
C_1							0,72

6. Distribution des résultats des classes

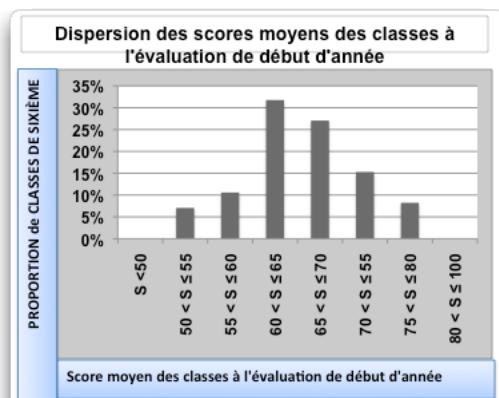
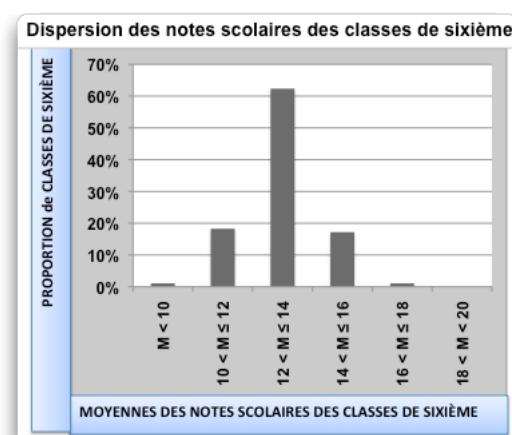
L'hétérogénéité intra-classes est bien connue des enseignants. L'hétérogénéité inter-classes l'est peut-être un peu moins. La présente étude ne fait que confirmer ce que nous observons depuis longtemps : la dispersion entre les classes est importante. Pour avoir une idée de l'importance de cette dispersion, on ne peut pas utiliser les scores normalisés : en effet, la normalisation des scores crée artificiellement de la dispersion. Il est donc nécessaire de revenir aux scores bruts ; ce qu'oublient bien souvent de faire les commentateurs des études internationales. Ces dernières ont cependant le mérite d'avoir mis en lumière le fait que la dispersion entre classes est plus grande chez nous que dans la plupart des pays comparables.



Prenons comme exemple l'épreuve D de sixième. Le diagramme montre la distribution des scores des classes. Environ 10% des classes ont un score moyen inférieur à 40%, tandis qu'environ 20% des classes enregistrent un score supérieur à 60%. La dispersion est encore plus grande si l'on considère séparément les parties orales et à support visuel de ce questionnaire.

Chacune des autres épreuves, que ce soit au niveau sixième ou au niveau cinquième, génère une dispersion de même ampleur que celle illustrée ici.

Au niveau sixième, la dispersion des scores à l'évaluation nationale de début d'année est du même type que celle enregistrée par EVAPM : 10% des classes obtiennent un score moyen inférieur à 60% tandis qu'un peu plus de 20% des classes obtiennent un score moyen supérieur à 70%.



Par contre, on note que l'ampleur de la dispersion des notes scolaires est moins importante que dans les deux autres cas. De plus, les moyennes de classe sont très rarement inférieures à 10. Cela illustre le fait, bien connu, de l'adaptation de la notation des enseignants au niveau réel de leur classe.

Évaluation de début d'année mise à part, ce qui est présenté ci-dessus à propos des classes de sixième se transpose sans modification notable pour les classes de cinquième.

7. Évolution des acquis de la Sixième à la Cinquième

Le tableau suivant permet de comparer directement les résultats obtenus selon le niveau ; il présente aussi les résultats comparés des questions de l'épreuve D qui étaient communes aux niveaux sixième et cinquième.

QUESTIONS COMMUNES AUX NIVEAUX SIXIÈME ET CINQUIÈME								
	Épreuve C				Épreuve D			
	Épreuve C_6ème		Épreuve C_5ème		Épreuve D_6ème		Épreuve D_5ème	
	Effectif	Score moyen ensemble de l'épreuve						
TOUS	213	30%	1210	37%	2363	49%	2310	59%
MASCULIN	117	32%	578	39%	1206	53%	1125	61%
FEMININ	96	27%	632	35%	1131	46%	1185	57%

Sauf exception, les questions des épreuves C et D portent sur des points qui font partie du socle commun de connaissances et de compétences. Les pourcentages de réussite observés en fin de cinquième montrent qu'il reste du chemin à faire pour que les objectifs du socle soient atteints à la fin de la scolarité obligatoire. Faut-il se réjouir de l'accroissement des réussites à ces questions entre le niveau sixième et le niveau cinquième ? Cet accroissement est de l'ordre de 8%, ce qui n'est pas

négligeable, surtout si l'on garde à l'esprit que l'enseignement en cinquième n'est plus focalisé sur les mêmes objets d'enseignement qu'en sixième. Cet accroissement témoigne du fait que les compétences développées en sixième sont entretenues en cinquième et qu'elles mûrissent, ce qui est plutôt rassurant. Toutefois, s'agissant de compétences que l'on pourrait penser bien installées en cinquième on peut exprimer une certaine inquiétude.

En prenant en compte les questions de l'épreuve B qui étaient communes aux niveaux sixième et cinquième, il y avait en tout 47 items communs aux évaluation sixième et cinquième de 2005 et de 2008. Le tableau suivant résume les résultats observés sur ces items

L'évolution concernant le calcul mental et la gestion mentale d'informations mathématiques ne se distingue pas de l'évolution observée ailleurs. De même, toujours concernant l'évolution sixième-cinquième, on n'observe pas de différence entre la partie orale du questionnaire D et la partie visuelle de ce questionnaire.

Par contre, on observe une diminution significative des écarts garçons-filles sur le questionnaire D et,

spécialement sur les questions présentées visuellement ; ces différences disparaissant même totalement en cinquième. Ce fait est assez surprenant et demande confirmation dans d'autres situations avant de pouvoir être interprété. S'il était confirmé, il pourrait résulter d'une différence statistique de maturité à 11 ans, entre les garçons et les filles, différence qu'il pourrait alors être bon de prendre en compte dans l'enseignement comme dans l'évaluation.

De la sixième à la cinquième, on aurait pu s'attendre à ce que certaines questions progressent davantage que d'autres, compte tenu de l'importance différente qui aurait pu être donnée à certains domaines selon les niveaux. Au lieu de cela, on observe une corrélation très importante entre les scores observés, sur les items communs ($\rho = 0,97$). De la sixième à la cinquième, les élèves progressent un peu mais à

peu près de la même façon partout.

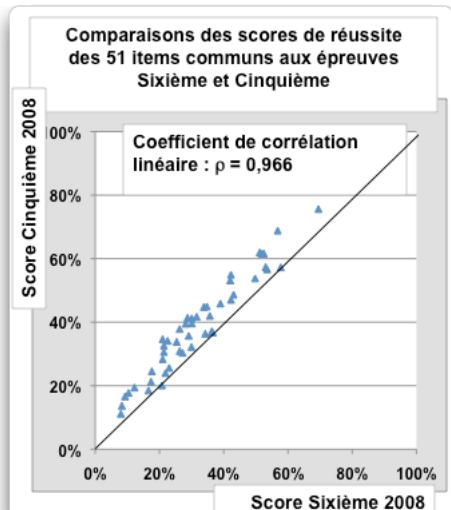
8. Comparaisons avec les études antérieures

Au niveau sixième, 81 items étaient repris de l'étude 2005. Le tableau suivant montre une stabilité remarquable des résultats.

La corrélation entre les scores de 2005 et de 2008 est tout aussi remarquable ($\rho = 0,97 !$). C'est dire que les élèves sont à l'aise aux mêmes endroits en 2008 qu'en 2005 et qu'ils éprouvent des difficultés de même ampleur d'une étude à l'autre. Il ne semble donc pas que les modifications apportées au programme de sixième entre 2005 et 2008 aient eu beaucoup d'effets.

ITEMS COMMUNS EVAPM 2005 et 2008 Sixième et cinquième (47 items)		
Sixième 2008	Cinquième 2008	Sixième 2005
32%	39%	31%

Partie des questionnaires orales et visuelles communes aux niveaux Sixième et Cinquième				
	ORAL		VISUEL	
	Sixième	Cinquième	Sixième	Cinquième
TOUS	55%	64%	46%	55%
MASCULIN	62%	68%	48%	56%
FEMININ	49%	60%	44%	55%



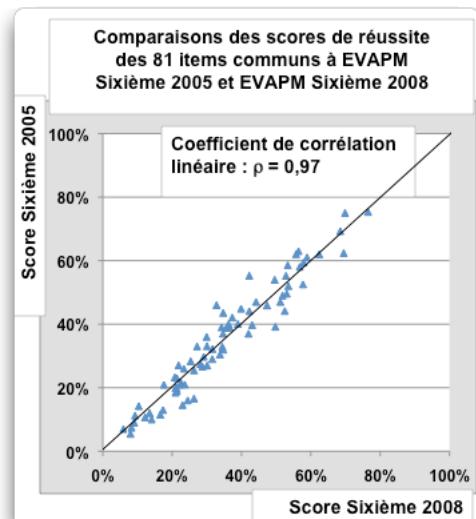
ITEMS COMMUNS EVAPM Sixième 2005 et Sixième 2008 (81items)	
Sixième 2008	Sixième 2005
31,9%	31,1%

En ce qui concerne les comparaisons avec les études antérieures à 2005, rappelons que, jusqu'en 2003, les études EVAPM se déroulaient en fin d'année scolaire. En 2005 et en 2008 elles se sont déroulées à la fin du second trimestre ; cela biaise à l'évidence les comparaisons avec des études antérieures à 2005. Il conviendra de tenir compte de cette remarque dans l'analyse des tableaux présentés. Rappelons que cette passation précoce est justifiée par notre souci de fournir aux enseignants des résultats utilisables, pour eux-mêmes et avec leurs élèves, dès le troisième trimestre de l'année de l'évaluation.

Les spécificités de l'étude 2008, à savoir, la place faite au calcul mental et la gestion mentale d'informations mathématiques, ainsi que la mise en relation de compétences observées simultanément en sixième et en cinquième, ne permettent pas de faire des comparaisons directes avec les études menées depuis 1987 en sixième et en cinquième.

Pour le niveau sixième, compte tenu de la stabilité des résultats observés entre 2005 et 2008, les conclusions de l'étude 2005 restent valables : dans la mesure où les comparaisons sont possibles, on observe une baisse moyenne de 8% par rapport à l'étude de 1997 et de 5% par rapport aux études de 1987 et de 1989. Ces baisses sont significatives et si on les rapporte aux taux moyens des scores observés, lesquels ne dépassent jamais 40%, ils sont, évidemment, très importants.

Le fait est que les résultats des élèves ne sont conformes ni aux attentes des enseignants, ni aux attentes des programmes. Toutefois l'interprétation de la baisse observée au fil du temps n'est pas aisée. D'une part, par rapport aux années 80 ou 90, de nouvelles compétences peuvent avoir été développées par rapport auxquelles les comparaisons ne sont pas possibles. D'autre part, et nous avons déjà signalé ce point, les élèves de 2008 sont en moyenne 3 mois plus jeunes qu'au cours des années 80. Pendant cette période, le taux des élèves ayant au moins un an de retard a en effet été divisé par 2 (et même un peu plus, passant de 40% à moins de 20%). Il n'est pas certain que la réduction des taux de redoublements imposée par le ministère se soit accompagnée d'une amélioration du niveau des élèves à l'entrée en sixième. Outre le sentiment exprimé par les enseignants, plusieurs indices vont plutôt dans le sens contraire. Cela signifierait que les difficultés d'enseignement se seraient accrues et donc, qu'il serait de plus en plus difficile pour les enseignants de mener leurs élèves au niveau attendu par les programmes.



9. Conclusion

L'analyse statistique présentée dans ce chapitre doit être lue sous l'éclairage des analyses qualitatives faites dans les autres chapitres. Notre expérience de l'évaluation en mathématiques nous a appris, et cela est une nouvelle fois vérifié, qu'il y a des stabilités étonnantes dans le temps (temps du système et temps de l'élève) et donc, qu'il faut développer beaucoup d'énergie pour sortir de cette stabilité vers le haut. Elle nous apprend aussi qu'il n'est pas possible de considérer la compétence mathématique comme étant unidimensionnelle. Le désir simpliste de disposer d'un indicateur unique de niveau mathématique ne résiste pas longtemps à l'examen. Cela signifie que, malgré nos efforts nous laissons de côté des dimensions de l'activité mathématique et des compétences qu'elle suppose.

Cette étude est une étude particulière qui a ses forces et ses limites. Elle ne prétend pas apporter une vérité définitive ni être supérieure à d'autres études ou avis. Au contraire elle demande à être confrontée à d'autres études et à l'expérience des enseignants.

Redisons ici que toute évaluation doit impliquer les acteurs du système évalué (ce que nous essayons de faire), qu'elle doit être diversifiée dans ses démarches et ouverte à la confrontation avec d'autres études.