

10. Analyse statistique des résultats

1. Introduction

Cet article complète les présentations et analyses qualitatives présentées dans ce dossier ; de ce fait, il n'a pas semblé utile de présenter ici le contexte, ni l'organisation générale de l'étude. Le lecteur trouvera sur le serveur de l'APMEP un document plus complet de présentation et d'analyse statistique ; il trouvera de même, en téléchargement, l'ensemble des fichiers de données ainsi que des précisions sur les traitements effectués.

L'étude complète porte sur 5 950 élèves et 245 classes de sixième et de cinquième, dont environ 13 % appartiennent à des établissements français de l'étranger.

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. Un document de synthèse des particularités observées sur la population correspondante a été communiqué aux enseignants concernés et est accessible sur le serveur.

1.1. 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, la passation n'a pas été équilibrée entre les épreuves (par exemple, les résultats de l'épreuve D de cinquième portent sur 2 435 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 admettons 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.

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

Les tableaux ci-dessous présentent le suivi de quelques indicateurs, depuis 1987.

Classes de Sixième (105 classes)⁽¹⁾

| | 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 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 en mathématiques | 13,10 | 12,86 | 11,81 | 11,58 | |
| Élèves d'âge « normal » | 82,1 % | 77,7 % | 70,8 % | 60,9 % | |
| Score évaluation nationale | 65,5 | 62,0 (64,3) | | | |

Les valeurs prises par ces indicateurs restent assez stables dans le temps. Nous pouvons toutefois 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.

Classes de Cinquième (107 classes)

| | EVAPM2008 | EVAPM1990 | EVAPM1988 |
|---|-----------|-----------|-----------|
| Nombre d'heures élèves en mathématiques (moyenne) | 4,04 | 3,94 | 3,95 |
| Nombre d'heures professeur par classe (moyenne) | 4,16 | | |
| Nombre moyen d'élèves par classe | 24,47 | 24,59 | 24,58 |
| Moyenne scolaire en mathématiques | 11,99 | 11,14 | 10,58 |
| Élèves d'âge « normal » | 77,90 % | 59,4 % | 52,1 % |

3. Résultats statistiques globaux

3.1. Scores par épreuves

Le lecteur trouvera sur le site l'ensemble des épreuves et dans l'étude statistique complète les résultats question par question et épreuve par épreuve.

(1) Nombres entre parenthèses : statistiques nationales (source DEP).

L'épreuve A de sixième est composée de QCM. Concernant ce type d'épreuve, 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. Nous préférons 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.

Le score moyen à l'ensemble des questions de l'étude est de 42 % en sixième et de 46 % en cinquième, ce qui signifie déjà que nos épreuves n'étaient pas particulièrement faciles pour les élèves. Sur l'ensemble des items, on note que 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.

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.

Notons 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 fortement biaisé, il serait hasardeux de généraliser, mais il y a sans doute là une piste de réflexion intéressante.

3.2. Scores par domaines

| Résultats par domaine : taux de réussite moyen | | | | | | | | |
|--|----------------|------------|---------|--------|----------------|------------|---------|--------|
| | Sixième | | | | Cinquième | | | |
| | Nombre d'items | TOUS | GARÇONS | FILLES | Nombre d'items | TOUS | GARÇONS | FILLES |
| Tous items | 174 | 42% | 45% | 40% | 150 | 46% | 47% | 44% |
| Géométrie sauf espace | 28 | 49% | 49% | 48% | 24 | 44% | 45% | 44% |
| Géométrie de l'espace | 30 | 41% | 42% | 41% | 24 | 46% | 47% | 45% |
| Nombres | 46 | 51% | 55% | 47% | 36 | 51% | 53% | 50% |
| Grandeurs | 70 | 35% | 38% | 33% | 66 | 43% | 46% | 41% |

On observe un assez bon équilibre entre les différents domaines. Cet équilibre 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.

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

Pour tenir compte des niveaux de 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 : Créativité ; E : Jugement.**

Une autre classification des questions concerne les niveaux des compétences utilisés dans les études PISA (voir sur le serveur le document de présentation du cadre de référence de PISA) : **niveau 1 : Reproduction ; niveau 2 : Connexions ; niveau 3 : Réflexion.** 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.

Le lecteur trouvera dans l'étude complète des tableaux présentant l'ensemble des indices et des résultats correspondants. Nous proposons simplement ici un résumé des observations les plus importantes.

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. Puisque 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 redoublants 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). Le coefficient de corrélation entre les notes scolaires et le score EVAPM pour le niveau sixième est assez élevé (0,74). Cette corrélation devient carrément spectaculaire lorsque l'on regroupe les notes scolaires par intervalles d'amplitude 2.

Dans ce cas la liaison linéaire devient évidente et le coefficient de corrélation linéaire est supérieur à 0,99 ! Cela semble signifier que notre évaluation ne s'éloignerait pas trop des pratiques et des attentes des enseignants. Pour le niveau cinquième, les coefficients de corrélation correspondants sont de 0,63 et 0,98.

Bien sûr, il est plus facile (et plus rigoureux) de travailler avec l'indicateur EVAPM. Toutefois, les remarques qui précèdent sont de nature à préciser le domaine de validité de nos observations et de nos conclusions.

Voici quelques remarques que l'on peut faire de l'observation des différentes corrélations (voir tableau dans l'étude complète). 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 (0,79). 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 renseignent 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 le 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 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.

Comme pour les études précédentes, nous observons des corrélations assez faibles entre les domaines de l'évaluation (de l'ordre de 0,50). Ceci n'est qu'une mise en évidence supplémentaire de la multi-dimensionnalité des compétences mathématiques.

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'oublie 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. 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.

Au niveau sixième, la dispersion des scores à l'évaluation nationale de début d'année est comparable à 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, les observations faites sur les classes de sixième sont valables sans modification notable pour les classes de cinquième.

On trouvera dans l'étude complète des diagrammes illustrant ces observations.

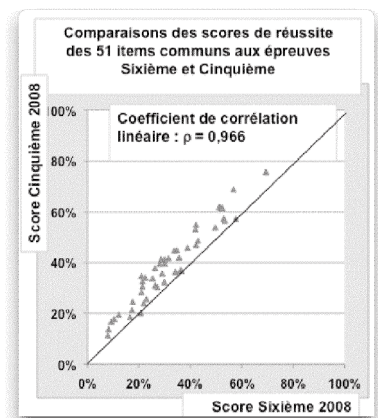
7. Évolution des acquis de la Sixième à la Cinquième

Les tableaux suivants permettent de comparer les résultats obtenus pour les questions 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 | Effectif | Score moyen ensemble de l'épreuve | Effectif | Score moyen ensemble de l'épreuve | 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.

L'accroissement des réussites à ces questions entre le niveau sixième et le niveau cinquième est de l'ordre de 8 points de pourcentage, 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 évaluations sixième et cinquième de 2005 et de 2008. Le tableau ci-contre résume les résultats observés sur ces items.

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

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

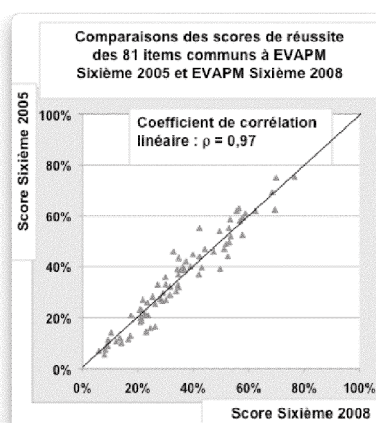
Au niveau sixième, 81 items étaient repris de l'étude 2005. Le tableau suivant montre une stabilité remarquable des résultats.

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 Sixième 2005 et Sixième 2008 (81 items) | |
|--|--------------|
| Sixième 2008 | Sixième 2005 |
| 31,9% | 31,1% |

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 points de pourcentage par rapport à l'étude de 1997 et de 5 points 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 pour lesquelles 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 d'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 cette 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.