

Organisation et fonctionnement des comparaisons internationales à grande échelle, avec un accent particulier sur PISA

Version longue

Antoine Bodin

IREM de Marseille

Consultant - mathématiques et évaluation

Ancien membre des groupes d'experts mathématiques
des études internationales TIMSS et PISA

Congrès de l'Espace Mathématique Francophone – EMF 2009

Université Cheikh Anta Diop – DAKAR

6 avril 2009

Projet Spécial 2 : Évaluations internationales

Impacts politiques, curriculaires et place des pays francophones.

Préparé pour le congrès EMF2009, le présent article reprend des éléments de certains textes publiés dans diverses revues (voir références), en développe certains et en complète d'autres.

Un diaporama sur le même thème est par ailleurs disponible.

Une version courte de cet article a été publiée dans les actes du congrès.

Résumé

L'article présente les études internationales dans ce quelles ont de spécifique par rapport aux mathématiques : origine, objectifs, démarches suivies, clés de lecture des résultats, enjeux et impact sur les systèmes éducatifs.

L'intérêt de ces études est souligné, mais aussi le risque lié à l'imposition généralisée de cadres de références et d'instruments d'évaluation susceptibles de modifier considérablement les conceptions relatives aux mathématiques et à leur enseignement. L'article est, de fait, un plaidoyer pour que la communauté mathématique (en particulier francophone) s'intéresse davantage à ces études et qu'elle en contrôle davantage le déroulement et les effets.

Abstract

This paper presents international studies in mathematics education: origin of the studies, goals, methodology, keys to reading the results, challenges and impact on educational systems.

The interest of these studies is acknowledged, but so is the risk of a general imposition of frameworks and tools which may significantly change ways of looking mathematics and mathematics teaching. Also, this paper advocates for the mathematical community (especially in French speaking countries) takes more interest in these studies and act to get more control on their conduct and their effects.

Sommaire

1. Introduction	3
2. Généralités sur le études internationales	4
2.1. Évaluation des connaissances - évaluation des systèmes ?	5
2.2. La place des mathématiques dans les évaluations nationales et internationales	6
2.3. L'objet évalué : quelles mathématiques ?	6
2.4. Les sujets de l'évaluation	8
3. Objet des études, domaines évalués,...	9
3.1. Organisation du domaine	9
3.2. Les cadres théoriques	11
3.3. Les types de questions	12
4. L'étude PISA	14
4.1. Origine et buts de l'étude PISA (Qui ? Pourquoi ?)	14
4.2. L'organisation des études PISA	16
4.3. Comprendre les scores de PISA	17
4.4. Les mathématiques dans PISA	19
4.5. Réflexions d'ordre épistémologique et didactique	19
5. Principaux enseignements des évaluations internationales et nationales	25
5.1. Remarques générales.	25
5.2. Quels résultats ?	27
5.3. Les acquis des élèves en France	29
6. Retombées des études	34
6.1. En général	34
6.2. En France	35
7. Le cas français	37
7.1. Résultats globaux	37
7.2. Influence du format des questions ?	38
7.3. Classement suivant les niveaux de compétence	38
7.4. Tentatives d'explications	39
7.5. Rapport au verbal et aux autres domaines disciplinaires	40
7.6. Les réactions	41
8. Impact possible de ces études sur les pays francophones	44
9. Conclusion	45
Références	46
Annexe 1 : Les études internationales concernant les mathématiques de 1960 à 2012.	49
Annexe 2 : Taxonomie des demandes cognitives pour la construction et l'analyse de tâches mathématiques – organisée par niveaux intégrés de complexité.	51
Annexe 3 : Classes de compétences	52

1. Introduction

Le présent article cherche à préciser les origines, les buts, les démarches et les effets, des études internationales portant, en particulier, sur les mathématiques. Plus précisément, portant sur l'organisation et les effets de l'enseignement des mathématiques aux niveaux scolaires.

Préparé en particulier pour le congrès 2009 de l'espace mathématique francophone (EMF2009) Dakar), nous avons repris ici des éléments de textes publiés dans diverses revues (voir références), en avons spécialement développé certains et en avons complété d'autres.

Compte tenu de l'actualité et, en particulier, de l'importance que prennent aujourd'hui, pour de nombreux pays, les études conduites par l'OCDE (PISA), nous avons été amenés à donner une place importante à ces études. Cependant, compte tenu du fait que certains pays de l'espace francophone participent, soit uniquement, soit simultanément, aux études conduites par l'IEA (TIMSS), et que d'autres, encore, se posent la question de leur participation, Il a semblé naturel de prendre aussi en compte ces dernières études.

Par ailleurs, de nombreux pays se dotent d'instruments d'évaluation basés sur des méthodologies héritées de celles développées par l'IEA et par l'OCDE. Les démarches sont souvent voisines comme le sont les instruments utilisés. Les études nationales et internationales se complètent utilement : les premières permettant d'éclairer les secondes, parfois d'en confirmer certaines conclusions, souvent, encore, d'en nuancer d'autres.

Cela explique que, pour le cas français, nous évoquons les études menées par le ministère de l'éducation nationale (DEPP) et celles menées par l'association des professeurs de mathématiques (études EVAPM).

Pour toutes ces études, de nombreuses questions se posent : Qui organise ces études ? Pour qui ? Pour quoi faire ? Leurs résultats sont-ils fiables ? Quels en sont les effets ?....

Bien d'autres questions de nature politique, idéologique, épistémologique, didactique, ... se posent encore.

Nous laisserons ici de côté les questions qui sont relativement indépendantes du ou des domaines concernés : questions méthodologiques et logistiques, recueil et traitement des données, contrôle de la validité des questions, en particulier par les méthodes de modélisation des réponses aux items et autres techniques complexes, contrôle de la fidélité¹, construction d'indices et d'échelles...

Nous laisserons aussi largement de côté la question des relations que ces études permettent d'établir entre les acquis des élèves et les variables relatives aux contextes (contexte socio-économique, contexte éducatif), ainsi qu'avec les variables personnelles relatives aux élèves.

L'incidence de ces questions sur la qualité et l'intérêt des évaluations du domaine mathématique ne doit pas sous-estimée, mais, dans cet article, nous chercherons plutôt à nous centrer sur les questions mettant directement en jeu le domaine évalué (les mathématiques).

Une version courte de cet article a été publiée dans les actes du congrès.

¹ Fidélité : terme technique qui signifie que la répétition de l'observation (y compris, dans notre cas, en faisant intervenir des "codeurs" différents, n'est pas susceptible d'en modifier le résultat.

2. Généralités sur le études internationales

Dans le domaine de l'éducation, les études internationales ont déjà une longue histoire. En cinquante ans, les domaines d'études se sont diversifiés, le nombre de pays concernés est passé d'une petite douzaine à une centaine. Ces études portent aussi bien sur les acquis scolaires (TIMSS, PIRLS) que sur la capacité des citoyens à affronter les problèmes de la vie courante (PISA, ALL). Les domaines d'étude traditionnels sont les mathématiques, la lecture et les sciences, mais, depuis quelques années on voit émerger d'autres sujets : langues, citoyenneté, géographie,...

Dans la plupart des pays, en résonance avec ces études et souvent en s'appuyant sur des cadres de référence proches, on voit se mettre en place des évaluations nationales sur échantillons ou portant sur l'ensemble des élèves d'un ou plusieurs niveaux scolaires.

Même les pays qui ne participent pas directement à ces études sont influencés par elles. En effet, les institutions internationales (EMS, Banque Mondiale,...) leur demandent de mettre en place des évaluations nationales, ce qui les amène inévitablement à emprunter aux cadres de référence des études internationales.

Pour ces études, plusieurs organisations internationales selon les cas coopèrent ou sont en concurrence :

- l'IEA (Association internationale pour l'évaluation des résultats scolaires, basé aux Pays-Bas),
- l'ETS (*Educational Testing Service*, basé aux USA),
- l'OCDE (Organisation de Coopération et de Développement Économiques, dont le siège est à Paris).

Il est à noter que les études menées par ces organismes, en relation étroite avec les ministères de l'éducation des pays concernés, répondent à une demande sociale et politique, qui, partie des USA (théorie de l'*« accountability »*), s'est étendue à l'ensemble des pays et des organisations internationales. Les questions proprement pédagogiques ou didactiques commencent à émerger, mais, dans l'ensemble, il s'agit essentiellement de fournir des indicateurs de pilotage aux responsables des systèmes d'enseignement.

Concernant cet aspect de la question, nous renvoyons le lecteur au rapport très documenté établi par N. Bottani et P. Vrignault pour le Haut Conseil de l'Évaluation de l'École (2005, cf. Références).

On trouvera en annexe 1 un tableau rappelant les différentes études menées depuis 1960 (première étude de ce type) ainsi que les études prévues pour les prochaines années.

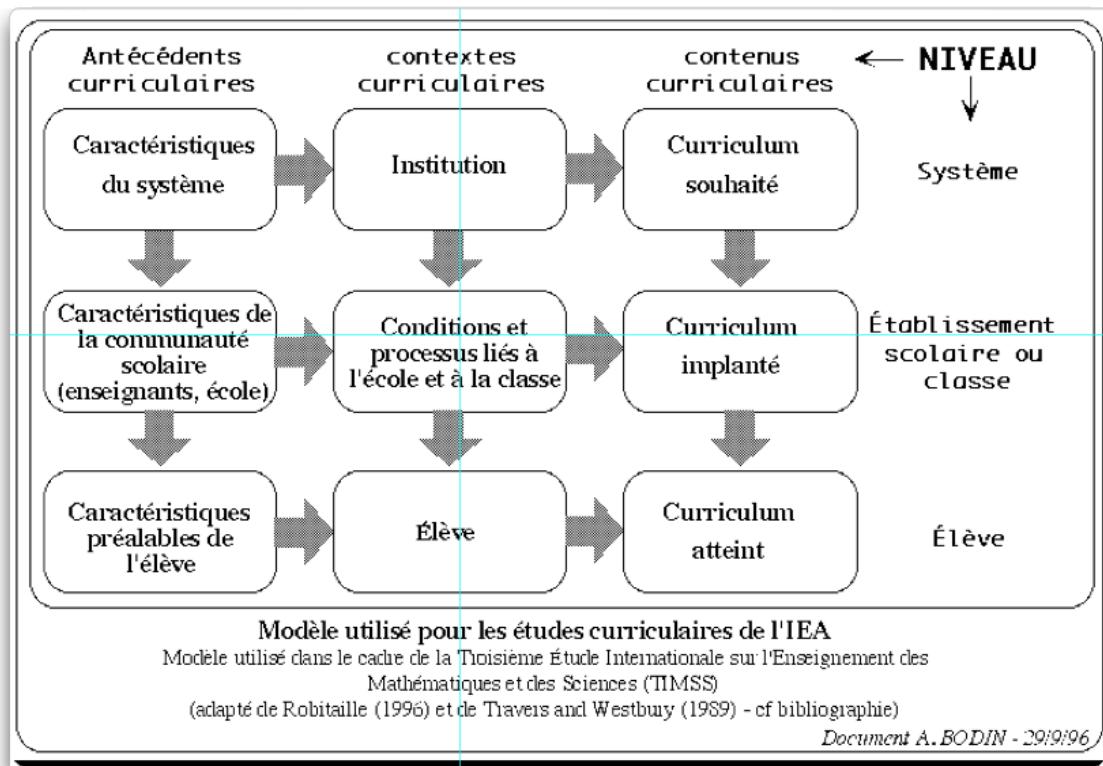
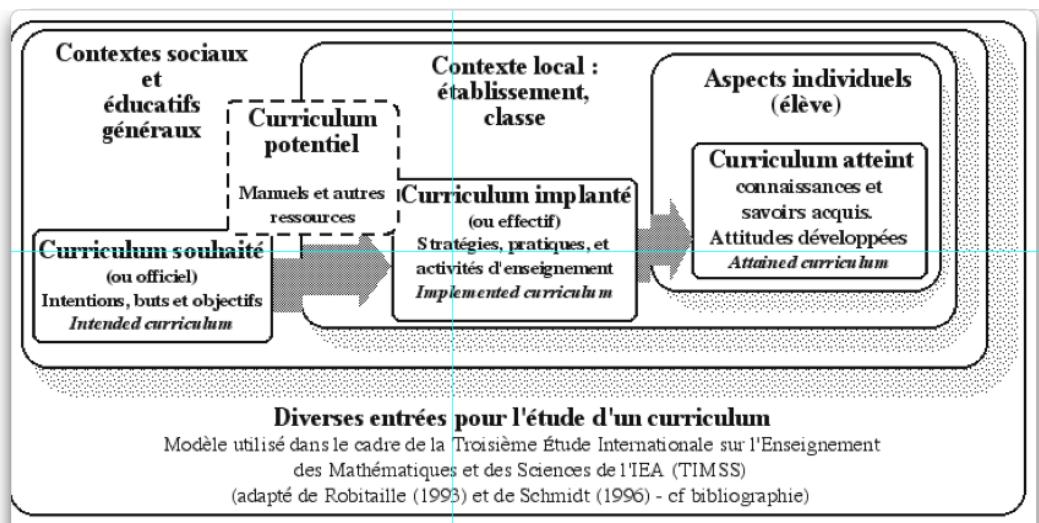
On remarquera la variabilité du vocabulaire : évaluations ou études ? Certaines opérations répondent en effet davantage à un souci de connaissance alors que d'autres émettent des jugements de valeur sur la qualité des enseignements, et donc des systèmes éducatifs. Ce dernier cas signifierait que certaines valeurs seraient reconnues d'une façon universelle, valables de la même façon pour l'ensemble des systèmes éducatifs, ce qui n'est pas sans poser quelques questions.

On constate que la France n'a participé qu'à une partie à ces études. Il s'en suit que les réflexions et les propositions françaises, lorsqu'elles existent, sont peu prises en compte. D'une façon assez logique moins la France participe et moins son influence se fait sentir, et réciproquement.

2.1. Évaluation des connaissances - évaluation des systèmes ?

Assez souvent, on ne retient des études internationales que les classements qu'elles produisent entre les pays, ou ce qu'elles disent des acquisitions cognitives des individus. Cependant ces études s'intéressent aussi, et surtout, aux rapports existant entre les acquisitions des individus en termes de connaissance et le niveau de développement des pays, le niveau économique des familles, les contextes dans lequel se déroule l'éducation, les conceptions et les représentations des enseignants et des élèves,...

Le schéma suivant, emprunté aux études TIMSS, met en évidence que, dans ce cas, on cherche à mettre en liaison les acquis des élèves, avec les programmes officiels, les moyens consentis pour l'enseignement, les pratiques d'enseignement,... Plus que les acquis des individus, c'est la totalité d'un système qui est ainsi soumis à l'analyse ou à l'évaluation.



Ces schémas mettent en évidence la complexité des systèmes qu'il s'agit d'évaluer. Les palmarès résultant de ces études, trop souvent mis en exergue, ne sont que des caricatures grossières des tentatives d'exploration et de compréhension des systèmes conduits à travers ces études.

Plus que l'établissement de constats, ces études cherchent en effet à identifier des points d'entrées susceptibles de faciliter, selon les pays et les systèmes, les améliorations jugées nécessaires.

Reste à savoir si le modèle est consistant, si les études sont valides (i.e. si elles évaluent bien ce qu'elles prétendent évaluer) et si l'usage qui en est fait est bien de nature à aider à l'amélioration des systèmes. Disons de suite que les avis sur cette question sont assez partagés.

2.2. La place des mathématiques dans les évaluations nationales et internationales

Les mathématiques semblent constituer la discipline privilégiée des évaluations nationales et internationales. Les sciences suivent de plus ou moins près selon les cas. La langue d'usage² aussi, mais il s'agit le plus souvent de lecture (plus précisément du traitement de l'information écrite) plutôt que de linguistique, d'expression orale ou écrite, ou de littérature. Le mot « mathématique », lui, est généralement utilisé comme désignant un domaine unifié, bien identifié, et dont chacun s'accorderait à penser que son étude et une certaine maîtrise seraient essentielles à toute formation.

Alors que d'autres disciplines sont éclatées en sous-domaines considérés comme distincts et plus ou moins indépendants (par exemple, pour les langues : lecture, production écrite, littérature,...), les mathématiques sont souvent supposées générer une sorte de trait latent qui permettrait de classer de façon univoque aussi bien les individus que les systèmes éducatifs. De fait, nombre d'études s'appuient implicitement sur une conception unidimensionnelle de la compétence mathématique (le niveau !). La recherche et les observations faites dans le cadre des évaluations à grande échelle, ainsi que le rappel de la grande variété des formes de l'activité mathématique fera nécessairement passer de cette conception naïve à une perspective multidimensionnelle.

Il y a encore, derrière cette situation, l'idée que les mathématiques constituaient :

- un langage universel, et qu'elles seraient, de ce fait, particulièrement adaptées aux comparaisons internationales,
- un domaine facile à évaluer,
- un bon indicateur de compétences plus générales.

Ces idées ne sont vérifiées, tout au plus, que dans des cas particuliers et sous certaines conditions ; dans bien des cas, elles conduisent au réductionnisme et à des interprétations fautives.

Il est à noter que, dans beaucoup de pays, les mathématiciens et les membres de la communauté mathématique, élargie aux enseignants de cette discipline, considèrent les évaluations nationales et internationales avec beaucoup de méfiance. Ils leur reprochent bien souvent un manque de validité (c'est-à-dire de manquer leur objet) et de donner une idée fausse de la discipline.

2.3. L'objet évalué : quelles mathématiques ?

Différents aspects des mathématiques peuvent être concernés par l'évaluation. Ces aspects entretiennent entre eux des rapports subtils, mais on ne peut pas les considérer comme hiérarchisés.

² Nous évitons dans cette communication de parler de langue maternelle. Pour beaucoup d'élèves dans le monde, et dans une certaine mesure en France, la langue d'enseignement n'est pas leur langue maternelle.

A - Les mathématiques pratiques

Il s'agit des mathématiques reconnues comme utiles à tous, pour la vie de tous les jours, ... (littératie mathématique au sens de PISA³).

Il est difficile de parler ici de bases mathématiques, et encore moins de socle, compte tenu des choix qui sont souvent faits. Base ou socle de quoi ? Pourquoi ?

B - La culture mathématique

On pourrait dire mathématiques pour tous, au sens de « l'honnête homme » moderne, qu'il ne saurait être question de réduire au citoyen-consommateur-employable.

Pour essayer de préciser les choses, les idées suivantes, par exemple, ne devraient-elles pas faire partie de cette culture mathématique commune ?

- Les mathématiques sont vivantes ; elles continuent à se construire (ou plutôt à être construites par des humains) - de nouveaux résultats sont trouvés chaque jour.
- Les nombres premiers ont de nombreuses utilisations pratiques, essentielles dans le domaine du cryptage...
- Il existe un nombre dont le carré est 2. Ce nombre ne peut pas s'écrire sous la forme d'une fraction d'entiers ; il est dit irrationnel. La découverte de ce fait, qui remonte au 6ème siècle avant notre ère, a provoqué une crise de la pensée qui a duré plusieurs siècles.
- Le théorème de Fermat : son énoncé, son histoire, sa résolution récente par Andrew Wiles.

Chacun multipliera sans peine les exemples en prenant en compte la place des mathématiques dans l'histoire de la pensée, ses relations avec la philosophie, avec les sciences, avec la technologie,...

Il n'est pas question ici de techniques ou de formalisme mathématique, mais d'une dimension culturelle trop souvent absente de nos programmes de formation.

C - Les mathématiques spécialisées

Il s'agit des connaissances, théories et méthodes mathématiques spécifiques sur lesquelles peuvent se greffer des développements ultérieurs.

Le caractère multidimensionnel de l'ensemble des connaissances et compétences relatives à ces trois catégories apparaît clairement. La maîtrise des savoir-faire pratiques n'ouvre pas nécessairement la porte à la culture au sens B ni aux compétences sollicitées en C. Il est toutefois évident, qu'une partie des connaissances et savoir-faire de la première catégorie est indispensable à l'exercice de la troisième, la difficulté est de savoir laquelle et, pour la partie concernée, de savoir jusqu'à quel niveau d'approfondissement.

Par ailleurs, à l'intérieur même de chacune de ces catégories, on peut distinguer des sous-domaines : algèbre, géométrie, analyse, statistiques... De nombreuses observations montrent que, en ce qui concerne les élèves de l'enseignement secondaire ayant en principe suivi le même

³ Le mot anglais « literacy » a été malencontreusement traduit en français par « culture », ce qui est source de nombreux malentendus.

programme d'enseignement, les réussites dans ces sous-domaines sont assez faiblement corrélées. On peut aussi identifier des types différents de processus cognitifs mis en jeu (reconnaissance et synthèse par exemple) et remarquer que les capacités à faire fonctionner ces différents types de processus sont eux-mêmes faiblement corrélés entre eux.

Quelques remarques avant d'aller plus loin :

- Aucune évaluation n'est possible si l'on ne définit pas au préalable un référentiel d'évaluation⁴, autrement dit, si l'on ne délimite pas soigneusement l'objet qu'il s'agit d'évaluer.
- Il est difficile d'évaluer les acquis des élèves sans prendre en compte leurs conditions d'apprentissage. Les compétences développées par les élèves dépendent en effet des opportunités d'apprentissage qui leur sont offertes (des curriculums suivis), comme elles dépendent de facteurs personnels (motivation, goût pour l'étude, habileté, rapport aux mathématiques, capacités générales...).
- Plus les questions deviennent spécifiques et plus sont faibles les corrélations entre les différents domaines (algèbre, géométrie,...). Il en est de même pour les corrélations entre ces domaines et la langue d'usage. Les inférences et généralisations hâtives doivent donc être évitées.

En particulier, l'idée qu'un certain niveau de maîtrise de la langue serait une condition préalable au développement des compétences mathématiques n'est que partiellement fondée, du moins à un certain niveau. Lorsque le questionnement mathématique est saturé de verbal, les corrélations observées sont évidemment fortes entre mathématiques et langue d'usage, mais ces corrélations diminuent lorsque les concepts et les outils mathématiques deviennent plus prégnants.

- L'insistance sur les « mathématiques de la vie », « mathématiques du réel », et sur « l'évaluation authentique », très en vogue aujourd'hui⁵,... ne facilite pas nécessairement la réussite des élèves. En général, cela rend même les choses plus difficiles, du moins tant que la dévolution n'est pas assurée (c'est-à-dire tant que l'élève, non seulement n'a pas compris la situation qui lui est proposée, mais encore, tant qu'il ne l'a pas faite sienne). Or cette dévolution est ce que les évaluations savent le moins bien prendre en compte.

Cela ne signifie pas pour autant qu'il faille se limiter à des questions mathématiques formelles. D'une part, les habillages des questions peuvent rester minimales et internes aux mathématiques, d'autre part il convient de s'assurer que le caractère concret des situations proposées est ressenti comme tel par les élèves. Le « faux concret » ne rendant pas nécessairement les situations proposées plus proche des intérêts des apprenants ; il peut même renforcer l'impression d'inutilité et d'artificialité des mathématiques.

2.4. Les sujets de l'évaluation

Selon les cas, les études portent soit sur les élèves d'un niveau scolaire donné (avec des difficultés de mise en correspondance des divers systèmes éducatifs), soit sur une classe d'âge.

⁴ Cadre de référence, en anglais : *framework*.

⁵ En anglais : «*realistic math* », «*authentic assessment* »

Les niveaux habituellement visés sont les 4^e, 8^e, et 12^e années d'enseignement (en France, CM1, Quatrième et classes terminales), plus, avec PISA, le niveau 15 ans (fixé sur une classe d'âge et non sur un niveau de scolarité).

La France est restée jusqu'à présent absente des études portant sur l'enseignement élémentaire. Il y aurait pourtant beaucoup de choses à apprendre de comparaisons internationales à ce niveau : la place des mathématiques y est en effet assez différente selon les systèmes éducatifs. Les notions enseignées et les pratiques d'enseignement sont aussi assez variables.

3. Objet des études, domaines évalués,...

Si la tendance, dans certains milieux, est de considérer la formation mathématique comme génératrice d'une compétence unidimensionnelle, les auteurs des diverses études se sont vite aperçus qu'il convenait de distinguer divers domaines, ainsi que différents niveaux de demande cognitive (appel aux concepts et aux processus de pensée).

La difficulté est alors de repérer les dimensions utiles ; cela nous amène naturellement à la question des cadres théoriques des évaluations.

Nous présenterons rapidement les cadres taxonomiques successifs de ces évaluations, depuis la NLSMA⁶ utilisée pour SIMS (1982) jusqu'au système de classification mis un point pour PISA. Cela permettra de mettre en évidence les progrès effectués mais aussi les limites de ces cadres.

3.1. Organisation du domaine

Les cadres de référence relatifs aux contenus ont le plus souvent constitué en une mise à plat des découpages traditionnels du corpus mathématique, obéissant en fait à une logique de syllabus. On trouve classiquement :

- Nombres et algèbre
- Géométrie
- Fonctions et Analyse
- Statistiques et probabilités

Une mise à plat de ce type a en particulier été élaborée pour les études TIMSS⁷.

En fait, chacun de ces sous-domaines se dérive de différentes façons suivant les études et des pans entiers des mathématiques scolaires n'y trouvent pas facilement leur place.

Par exemple, où placer la question des grandeurs ? les mathématiques discrètes ? Les différentes traditions scolaires s'en arrangent plus ou moins bien, mais ... les comparaisons inter-curriculaires sont difficiles.

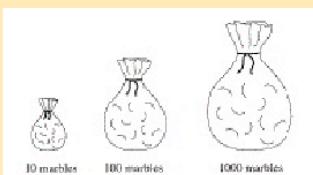
Même des notions qui ont un statut clair dans le champ mathématique peuvent avoir une signification floue lorsqu'il s'agit de mathématiques scolaires. C'est par exemple le cas des probabilités.

⁶ NLSAM : New Longitudinal Study of Mathematical Abilities

⁷ cf Robitaille 1993, et traduction française d'Antoine Bodin à l'IREM de BESANCON (Dossier TIMSS).

**Une question de probabilité ?
TIMSS 1995 Pop 2 (Question M03)**

Il n'y a qu'une bille rouge dans chacune de ces sacs.



Sans regarder à l'intérieur des sacs, tu dois retirer une bille de l'un d'entre-eux. Quel sac te donnera le plus de chance de tirer une bille rouge ?

- A. Le sac qui contient 10 billes
- B. Le sac qui contient 100 billes
- C. Le sac qui contient 1000 billes
- D. Tous les sacs donnent la même chance

Résultats niveau Quatrième (grade 8)

International	France	Hongrie	Canada
76 %	82 %	82 %	90%

Dans certains pays, les probabilités apparaissent dès l'élémentaire, tandis que dans d'autres, elles ne font leur entrée qu'après l'école obligatoire (classe de Première en France). Il est clair qu'il ne peut s'agir de la même chose.

Signalons que dans l'étude TIMSS, les élèves français de 13 ans (population 2) ont aussi bien réussi les questions dites de « probabilité » que les élèves de pays où la notion était étudiée depuis l'élémentaire. Avec un score moyen de 71%, les questions de statistiques et de probabilités obtenaient même, en France, un score plus élevé que le score obtenu dans chacun des autres domaines de l'étude et un score meilleur que celui des élèves hongrois (66%) qui étudiaient les probabilités depuis plusieurs années (notons au passage que ce qui était vrai pour TIMSS 1995 ne l'est plus pour PISA 2003).

Par exemple, la question ci-contre, où pourtant le terme « probabilité » est utilisé dans l'énoncé de la question, était réussie par 54 % des élèves de quatrième ...

Bien sûr, pour certaines de ces questions, par exemple celle des billes, présentée plus haut, il ne s'agit que d'appel au bon sens ou, si l'on veut, aux préconceptions probabilistes ; ces conceptions mêmes qui ont fait pendant des siècles obstacle au développement des concepts et des méthodes probabilistes. Dans ces conditions, on peut comprendre que moins les élèves auront été exposés à la théorie (ou à la modélisation), plus ils seront à l'aise devant ce type de question (remarque largement confirmée par les études faites en Première dans le cadre d'EVAPM).

De son côté, PISA s'éloigne des découpages classiques des contenus, pour faire place à une organisation de type qu'il est convenu, en France, d'appeler « par problématiques »⁸.

Cependant, l'idée du découpage de PISA est plutôt dérivée des recommandations de la Société Mathématiques Américaine - AMS (cf. « *On the shoulders of the Giants* » qui décrit les grandes idées « *overarching ideas* » reprises par PISA)⁹ :

- Quantité
- Variations et relations
- Espace et formes
- Incertitude

Une question de probabilité : TIMSS 1995 Pop 2 (Question O05)

Chacune des six faces d'un cube est peinte soit en rouge soit en bleu. Lorsqu'on lance le cube, la probabilité qu'il retombe une face rouge vers le haut est $\frac{2}{3}$. Combien y-a-t-il de faces rouges ?

- A. Une
- B. Deux
- C. Trois
- D. Quatre
- E. Cinq

Résultats niveau Quatrième (grade 8)

International	France	Angleterre	Singapour
47 %	54 %	39 %	88 %

⁸ Voir les problématiques de l'APMEP élaborées sous l'égide de Régis Gras.

⁹ Voir sur le site de l'APMEP (Informations sur PISA) une description de ces problématiques.

Cette centration sur les types de problèmes, plutôt que sur les contenus susceptibles d'intervenir, est favorable à la production de questions plus signifiantes (mais qui s'avèrent, on le sait, plus difficiles pour les élèves), où les mathématiques interviennent comme outil et ne sont pas simplement étudiées comme objet. Mais aller dans ce sens suppose aussi des questions bien calibrées pour que l'on sache bien ce que, en fin de compte, la question pourra contrôler (ou révéler).

Comme cela a déjà été souligné, la compréhension d'une question, liée à un texte non-mathématique, à structure complexe, est un premier obstacle que beaucoup d'élèves ne franchissent pas.

3.2. Les cadres théoriques

Les typologies précédentes sont surtout établies pour permettre une évaluation curriculaire (au regard de ce qui figure dans les programmes d'enseignement) ; beaucoup moins pour prendre en compte le fait qu'à exposition égale aux contenus, les acquisitions dans les divers sous-domaines varient en fonction des élèves, ni d'ailleurs pour reconnaître que les corrélations entre les résultats des divers domaines peuvent être assez faibles.

Un autre genre de classification, qu'il est convenu d'appeler taxonomique (il s'agit de taxonomies cognitives), est passé par différentes étapes et a produit divers outils dont aucun n'est vraiment satisfaisant mais qui permettent de prendre un peu de distance par rapport aux contenus.

- La Taxonomie de Bloom a ouvert la voie. Cette taxonomie s'est révélée peu valide pour les mathématiques (on voulait une taxonomie hiérarchisée, valable pour l'ensemble des disciplines et susceptible de produire des échelles de Guttman¹⁰ - le rêve de l'éducateur !).

Si tel était le cas, les niveaux A, B, C, D, E, F de la taxonomie seraient tels que la réussite au niveau F impliquerait la réussite au niveau E, etc.

$$F \Rightarrow E \Rightarrow D \Rightarrow C \Rightarrow B \Rightarrow A$$

Bien sûr, on savait bien qu'il ne pouvait pas s'agir d'une implication logique, mais tout au plus d'une implication statistique dans un sens qui restait à définir¹¹.

- La Taxonomie NLSMA (*National Longitudinal Study of Mathematical Abilities*) construite pour les seules mathématiques et qui cherchait à corriger les insuffisances de la taxonomie de Bloom.
- Le modèle à trois dimensions de TIMSS :

Ce modèle cherche à prendre en compte l'idée que le croisement entre les contenus et les niveaux taxonomiques (tables de spécification classiques) est insuffisant pour caractériser une question d'évaluation ou un élément de curriculum. Dans ce modèle de tels éléments sont repérés par un système tri-vecteur, dont les vecteurs constitutifs appartiennent à des espaces respectivement générés par :

- Les contenus (syllabus),
- Les types de compétences (démarches sollicitées et produits attendus),

¹⁰ Une échelle de Guttman parfaite serait constituée d'un ensemble d'items rangés par ordre de difficulté croissante et tels que la réussite à l'item de rang n implique la réussite à chacun des items de rang inférieur à n .

¹¹ L'analyse implicative développée par Régis Gras et ses élèves (cf. références) est, de ce fait, susceptible de recevoir des applications intéressantes dans le domaine de l'évaluation.

- Les perspectives (Attitudes face aux mathématiques – rapport au domaine).

Voir le cadre de référence (frameworks) de TIMSS.

- Le cadre de référence de PISA, le plus récent est sans doute aussi le mieux élaboré, mais essentiellement dans la perspective « littéracie mathématique » qu'il définit en même temps.
- EVAPM s'appuie sur un cadre de référence multidimensionnel qui utilise en particulier un classement en niveau de compétences et une taxonomie de la complexité cognitive issue des travaux de R. Gras¹².
- Les évaluations nationales de la DEPP s'appuient sur des grilles de compétences assez pauvres, mais il est clair que leur cadre de référence est plutôt constitué par les programmes officiels.

La tendance générale est de relativiser l'importance des contenus isolés pour chercher à évaluer des compétences qui, par définition, articulent des contenus entre eux avec des situations dans lesquelles ces contenus deviennent des outils. Cela, du moins en théorie ! Dans la pratique, certains étiquetages en « compétences » ne sont qu'une concession à la mode et ne sont guère plus qu'un ré-étiquetage de ce qui était autrefois appelé capacités.

Ce reproche ne vaut que très partiellement pour l'étude PISA qui, aidée en cela par son relatif désintérêt pour les curriculums, est la première à avoir réussi à se centrer sur l'évaluation de compétences (même si la définition même de ces compétences continue à poser des problèmes).

3.3. Les types de questions

Les QCM

Dans les études internationales, en opposition avec les conceptions et les pratiques les plus courantes en France, la prédominance, voire le caractère exclusif des questions à choix multiple a longtemps été la règle. De plus, ces QCM sont en général des questions pauvres, à réponse unique.

La question présentée ci-contre (TIMSS 95) est typique de cette catégorie de questions.

D'autres types de QCM à réponses multiples, plus intéressantes sous bien des points de vue, commencent cependant à être utilisées (voir des exemples dans les études EVAPM).

Les QROC (Questions à Réponses Ouvertes et Courtes)¹³

En France, les évaluations n'utilisent le format QCM que de façon marginale. La plupart des

Un exemple de QCM : TIMSS 1995 Pop 2 (Question O2)

Si le prix d'une boîte de chocolats passe de 60 francs à 75 francs, quel est le pourcentage d'augmentation du prix

- A. 15%
- B. 20%
- C. 25%
- D. 30%

Résultats niveau Quatrième (grade 8)

International	France	Norvège	Singapour
28 %	29 %	29 %	78 %

Un exemple de QROC : TIMSS 1995 Pop 2 (Question R14)

Pierre a acheté 70 objets et Hélène a acheté 90 objets. Chaque objet coûte le même prix, et ensemble ils ont coûté 800 F.

Combien Hélène a-t-elle payé ?

Réponse : Hélène a payé :

Résultats niveau Quatrième (grade 8)

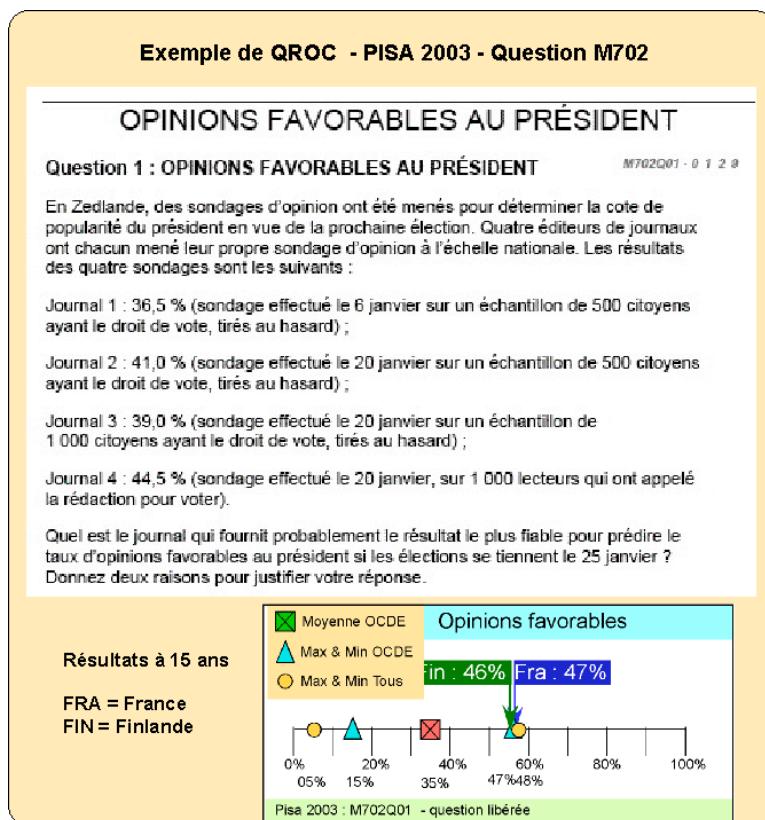
International	France	Angleterre	Singapour
47 %	54 %	39 %	88 %

¹² On trouvera tous ces documents sur le site de l'APMEP à l'entrée EVAPM

¹³ Short answer questions.

questions utilisées sont de type QROC, et, plus souvent encore, à réponses complètement rédigées (QRR).

À partir de TIMSS, les études internationales se sont partiellement ouvertes aux QROC (lesquelles représentent 40% des questions dans le cas de PISA). L'ouverture ne porte cependant que sur la réponse et non sur la question. De plus, « court » peut simplement signifier que l'élève est appelé à écrire lui-même une réponse, par exemple numérique, sans justification (comme, par exemple, dans la question présentée ci-contre). Il reste donc peu de place pour la rédaction, l'expression des idées, l'explicitation des démarches, la justification des résultats. L'absence relative de prise en compte de ces éléments dans les études internationales, éléments auxquels les enseignants français, comme le curriculum, sont très attachés, a contribué au désintérêt que la France a pu montrer pour ces études.



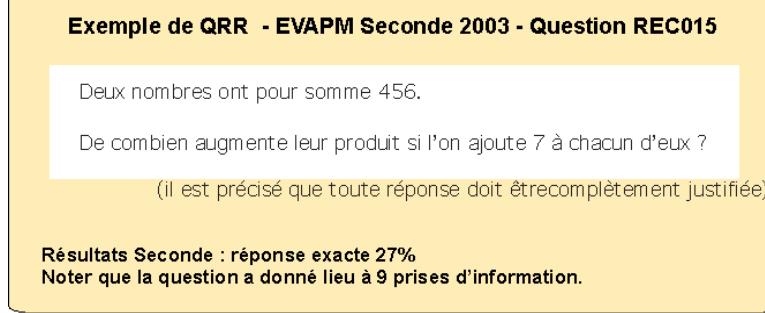
Les QRR (Questions à Réponses Rédigées)¹⁴

La limite entre QROC et QRR n'est pas bien établie. Une question qui serait considérée comme une QROC en France sera facilement considérée comme une QRR dans d'autres pays.

L'utilisation de questions qui pourraient justifier l'appellation QRR est très rare dans les études internationales et dans la plupart des études nationales. De ce point de vue, EVAPM, qui propose souvent des QRR, constitue une exception.

C'est évidemment le souci de fidélité, et l'obsession liée à la construction d'échelles aux qualités qui voudraient être indiscutables, qui a conduit à privilégier les QCM dans les études internationales. À cet aspect fidélité, il faut en plus associer celui de la facilité du codage, lequel peut être entièrement confié à la machine.

Les insuffisances des QCM sont cependant largement reconnues. Outre leur tendance à privilégier certains types de connaissances et



¹⁴ Long answer ou extended answer questions.

de compétences, leur validité est souvent discutable et l'est d'autant plus que l'on prétend faire une évaluation authentique (aussi proche que possible des conditions d'utilisation «normales» des connaissances).

D'un point de vue édumétrique (qui est en fait celui de la psychométrie), les qualités recherchées de fidélité et de validité jouent de façon antagoniste. Tant que l'on privilégie les échelles et que l'on a une conception unidimensionnelle du domaine à évaluer, c'est la fidélité qui l'emporte, et donc les QCM. Lorsque les réflexions épistémologique et didactique parviennent à se faire entendre, le souci de validité prend sa place et les études s'ouvrent à des questions à réponses ouvertes. C'est ce qui s'est largement passé pour les études internationales depuis TIMSS, malgré la difficulté évidente que représente le codage des réponses de ce type dans un contexte multiculturel et multilinguistique.

Parallèlement, les pratiques françaises, jusqu'ici rétives à l'utilisation des QCM, commencent à leur faire une certaine place.

4. L'étude PISA

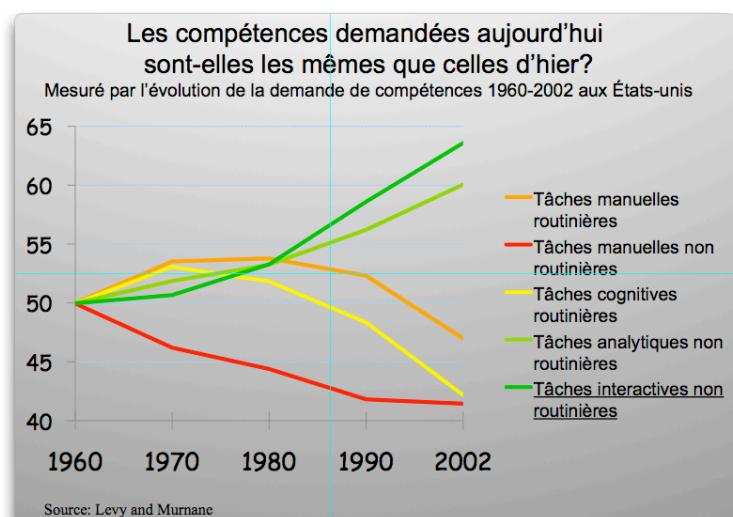
4.1. Origine et buts de l'étude PISA (Qui ? Pourquoi ?)

L'acronyme PISA désigne le programme international pour l'évaluation des élèves mis en place par l'OCDE depuis l'année 2000. Les 30 pays de l'OCDE plus un certain nombre de pays dits partenaires (58 pays en tout en 2006, 63 en 2009,...) participent tous les trois ans à une évaluation commune des compétences de base de tous les jeunes de 15 ans (encore scolarisés), à quelque place qu'ils se trouvent dans les systèmes éducatifs concernés. Il s'agit d'évaluer, de façon indépendante des programmes d'enseignement (les curriculums), la façon dont les jeunes sont prêts, vers la fin des scolarités obligatoires, à affronter les défis du monde dans lequel ils sont appelés à vivre.

L'OCDE part du principe que le monde moderne exige de tous les citoyens des connaissances et des compétences de plus en plus élevées. Le graphique ci-contre, illustre bien ce point de vue. Bien que limité à des données nord-américaines, il est utilisé par l'OCDE comme argument universel.

Cette idée de connaissances et de compétences nécessaires à tous est très liée à l'idée d'émergence d'une société de la connaissance et de formation tout au long de la vie.

La présente communication étant faite au Sénégal, la question de la pertinence de ces réflexions pour le continent africain ne peut manquer d'être posée. Nous laisserons à nos collègues africains le soin d'y répondre en leur suggérant d'étudier de près la pertinence des cadres de référence et du questionnement utilisé avant même de s'intéresser, s'ils le souhaitent encore, aux résultats enregistrés par les quelques pays du continent ayant participé aux études.



Avec PISA, l'OCDE s'adresse en premier lieu aux décideurs et aux gestionnaires auxquels elle fournit des indicateurs pour le pilotage des systèmes éducatifs. Cela justifie, dans une certaine mesure, le nombre important des indicateurs produits et l'attention apportée à leur qualité technique. Les méthodologies mises en œuvre doivent, en particulier, recevoir l'assentiment de l'ensemble des gouvernements concernés et les indicateurs obtenus doivent supporter la comparabilité dans le temps et dans l'espace géographique. Dans cet article, nous n'évoquerons que quelques-uns de ces indicateurs et nous renverrons à d'autres études pour ce qui concerne les indicateurs de nature économiques, socio-économiques, sociaux, ainsi que les relations de ces indicateurs avec les indicateurs plus directement liés à l'éducation et aux résultats de l'éducation.

La carte ci-contre résume la place que PISA a prise dans un grand nombre de pays. L'absence quasi totale du continent africain est remarquable. Elle s'explique largement par des raisons économiques.

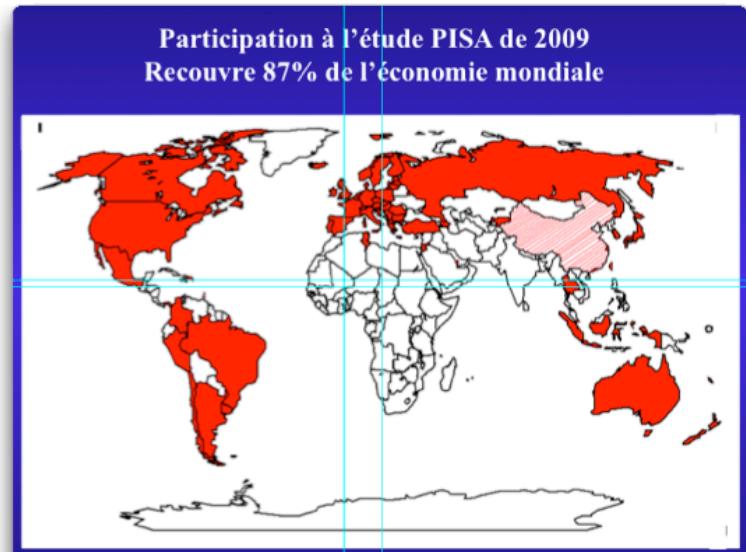
Précisons davantage les objectifs de PISA en utilisant les termes mêmes utilisés par l'OCDE.

« L'enquête PISA vise à évaluer dans quelle mesure les jeunes adultes de 15 ans, c'est-à-dire des élèves en fin d'obligation scolaire, sont préparés à relever les défis de la société de la connaissance.

L'évaluation est prospective, dans le sens où elle porte sur l'aptitude des jeunes à exploiter leurs savoirs et savoir-faire pour faire face aux défis de la vie réelle et qu'elle ne cherche pas à déterminer dans quelle mesure

les élèves ont assimilé une matière spécifique du programme d'enseignement.

Cette orientation reflète l'évolution des finalités et des objectifs des programmes scolaires : l'important est d'amener les élèves à utiliser ce qu'ils ont appris à l'école, et pas seulement à le reproduire. » (OCDE 2004 - Apprendre aujourd'hui, réussir demain Premiers résultats de PISA 2003



Ce qui est évalué est la littéracie, que l'OCDE définit ainsi :

« ...la notion de « littéracie¹⁵ », ...renvoie à la capacité des élèves d'exploiter des savoirs et savoir-faire dans des matières clés et d'analyser, de raisonner et de communiquer lorsqu'ils énoncent, résolvent et interprètent des problèmes qui s'inscrivent dans divers contextes. » (ibidem)

Enfin, ce qui concerne les mathématiques :

« La littéracie mathématique est l'aptitude d'un individu à identifier et à comprendre le rôle que les mathématiques jouent dans le monde, à produire des jugements fondés sur les mathématiques, et à s'engager dans des activités mathématiques, en fonction des exigences de sa vie en tant que citoyen constructif, impliqué et réfléchi.¹⁶ » (ibidem)

¹⁵ Littéracie et non "culture" comme cela est souvent écrit en français. Le terme anglais *literacy* ne correspondant en rien à l'idée que l'on se fait de la culture dans la plupart des...cultures !

¹⁶ Traduit de l'anglais par l'auteur. La définition habituellement donnée en français est incorrecte (et stupide !).

"... la littéracie mathématique ne peut se réduire à la connaissance de la terminologie mathématique, de propriétés et de procédures, ni aux savoir-faire permettant d'effectuer certaines opérations ou d'appliquer certaines méthodes, tout en présupposant, bien sûr, l'existence de ces compétences. Ce qui caractérise la culture mathématique est la mise en œuvre créative de ces compétences pour répondre aux exigences suscitées par les situations externes où se trouve l'individu."

Il est donc clair que les objectifs pris en compte par PISA ne recouvrent pas les objectifs de notre système éducatif. Dans un article précédent (Bodin, 2005), nous avions pu estimer que, pour les mathématiques, le questionnement de PISA recouvrerait 15 à 20% des programmes du collège. Certes, il s'agit, a priori, des 15 à 20% considérés comme les plus utiles à tous ; il ne serait donc pas anormal de leur accorder une attention particulière, mais cela montre bien que l'on ne peut pas considérer que PISA évalue la qualité globale de notre système éducatif. En fait, selon les pays, l'adéquation de PISA aux curriculums en vigueur est plus ou moins grande, ce qui, a soi seul, explique, sans les justifier, une grande partie des différences observées.

4.2. L'organisation des études PISA

L'organisation est lourde et complexe ; nous renvoyons aux documents cités en référence pour plus de détails. Résumons seulement les points principaux :

- L'étude est périodique de période 3 ans (2000 ; 2003 ; 2006 ; 2009 ; ...). À chaque occurrence de l'étude, une partie des questions est gardée secrète pour permettre les comparaisons ultérieures.
- Trois domaines sont concernés (lecture, mathématique et science) avec en plus, en 2003, un domaine considéré comme transdisciplinaire : la résolution de problèmes.
- Pour chacune des opérations, l'accent est porté sur l'un des domaines. En 2003, c'était les mathématiques avec environ le 2/3 du temps de passation des épreuves consacré aux questions de ce domaine (en 2006, l'accent portait sur les sciences ; en 2000 il portait sur la lecture ; l'accent sur à nouveau les mathématiques en 2012).
- Les échantillons d'élèves passant les épreuves sont censés représenter statistiquement l'ensemble des jeunes de 15 ans des pays ou des systèmes concernés. Des procédures strictes de contrôle de la qualité des échantillons sont mises en œuvre. Des pays peuvent être sortis de l'étude pour manquement aux règles imposées pour l'échantillonnage (ce fut le cas du Royaume-Uni en 2003).
- Tous les élèves ne passent pas toutes les questions de l'évaluation et, par le jeu des livrets d'évaluation, tous les élèves ne passent pas les questions dans le même ordre. Signalons au passage que les livrets d'évaluation étaient, en 2003, formés de 3 modules de mathématiques et d'un module d'un autre domaine.
- Ce ne sont pas les enseignants des élèves qui administrent les épreuves et les codages des réponses sont faits par des personnes spécialement entraînées et indépendantes des établissements des élèves testés (théoriquement !).

Précisons que le temps de passation par livret est fixé à 2 heures et cela pour un nombre d'items de l'ordre d'une cinquantaine. Le temps moyen alloué pour répondre à un item est d'environ 2 minutes (mais une question peut comporter plusieurs items). Il suffit de jeter un œil sur les questions de lecture, ou de sciences, mais aussi de mathématiques, pour constater que les élèves n'ont que peu de temps pour chercher et qu'ils doivent décider rapidement de leurs réponses.

4.3. Comprendre les scores de PISA

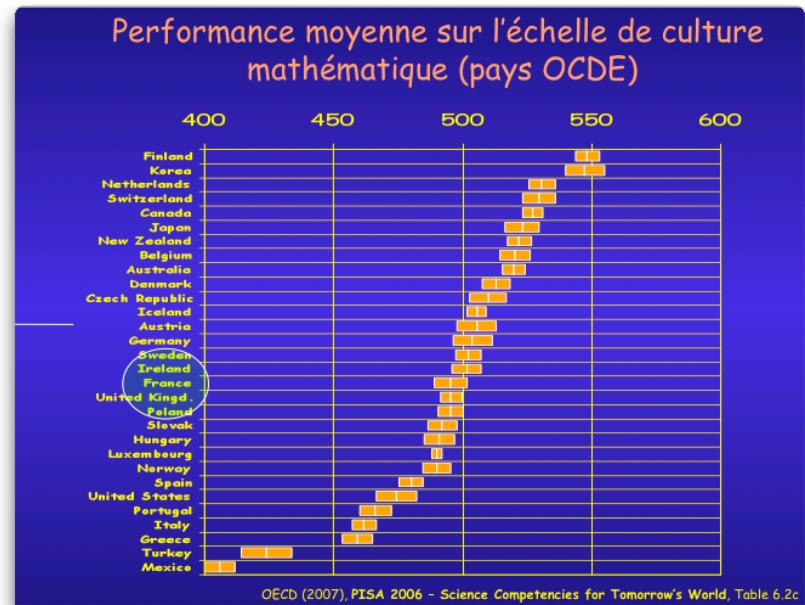
Le dessin ci-contre présente le classement classique des résultats selon les pays. Les résultats de y sont rapportés à une échelle qui reste mystérieuse pour la plupart de ceux mêmes qui l'utilisent. En effet, que signifie la phrase suivante : « *En mathématiques, en 2003 ; le score de la France était 511, tandis que le score de la Finlande était de 544, soit un écart de 32 points....* » ? Sur une échelle d'amplitude apparente de 1000 unités, 32 points, est-ce beaucoup ou est-ce négligeable ?

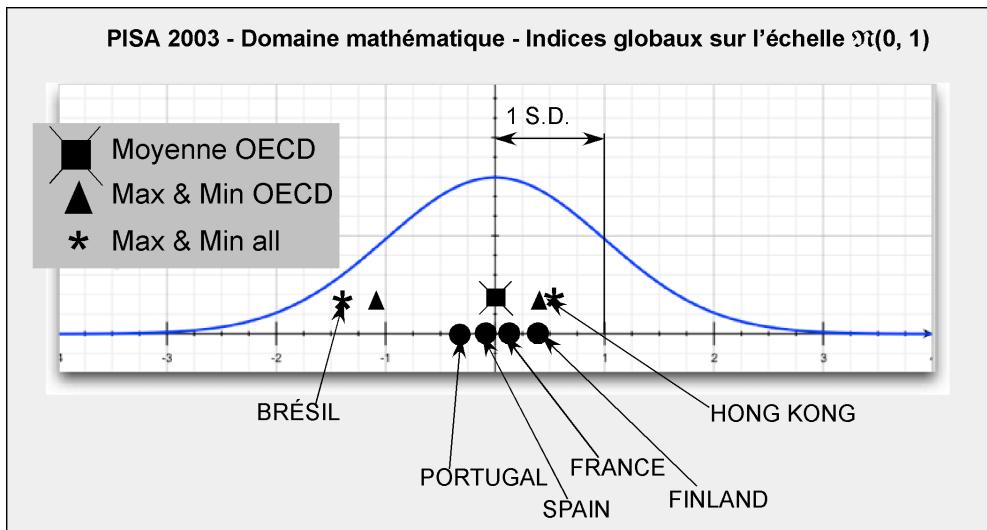
Pour pouvoir répondre à cette question, il faut avoir une idée de la façon dont les données sont traitées et dont ces "scores" sont calculés. En fait, une technologie complexe est utilisée dont il n'est possible de donner ici qu'un aperçu.

Dans un premier temps, on connaît, par pays, les réussites-échecs de chaque élève à chacune des questions auxquelles il a été soumis. Cela permet de déterminer, toujours par pays, les taux de réussite à chacune des questions de l'étude (telle question a été réussie dans tel pays par x% et dans tel autre pays par y% des élèves l'ayant passée).

Partant des résultats bruts (en fait des codages en [0 ; 1]), une procédure d'affectation probabiliste permet de traiter les élèves n'ayant pas passé certaines questions comme s'ils les avaient passées. D'autres corrections et ajustements sont faits pour tenter de réduire certains biais qui auront été détectés et en particulier pour tenir compte des biais d'échantillonnage. La procédure d'attribution qui conduit finalement à attribuer un « score » à chaque élève est assez complexe et comporte de nombreuses itérations. Cette procédure est destinée à rendre comparable les résultats d'élèves n'ayant pas été soumis aux mêmes questions. Elle vise aussi à inférer sur l'ensemble des jeunes de 15 ans les résultats observés sur un échantillon représentatif (en France, par exemple, seuls 4300 élèves ont réellement passé des épreuves, c'est à dire environ 1000 élèves pour chacune des questions). Ce que l'on obtient à ce moment, c'est un indice qui reste assez proche des scores réels observés, mais ce n'est déjà plus un score au sens strict. La distribution des scores obtenus à ce premier niveau est alors transformée pour être ajustée à la distribution normale réduite $\mathcal{N}(0 ; 1)$. On obtient ainsi un indice de réussite (il ne faudrait plus parler ici de score).

L'indice de réussite de l'ensemble des jeunes de 15 ans de l'OCDE est donc ajusté à la loi normale $\mathcal{N}(0 ; 1)$ et l'on peut alors replacer les résultats d'un pays sur cette échelle (en se restreignant aux résultats de ce pays). On peut de même placer chaque individu sur cette échelle et parler, par exemple, d'un individu de niveau 2 par rapport à cet indice, cela pour dire que ses résultats se trouvent à deux écart-types de la moyenne par rapport à l'ensemble des jeunes de l'OCDE. Voici un exemple de présentation possible des indices de quelques pays.

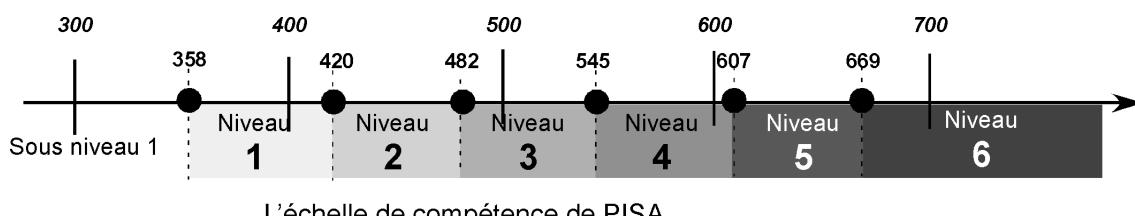




Insistons sur le fait qu'à ce niveau, on a totalement perdu de vue les scores. La seule chose que l'on puisse dire de l'écart, par exemple, entre la France et la Finlande est qu'elle est de 33 centièmes d'écart-type sur l'échelle ainsi construite. Pour des raisons de lisibilité, on effectue une nouvelle transformation pour ajuster notre distribution à la distribution normale de moyenne 500 et d'écart-type 100. L'échelle de compétence mathématique de PISA est donc, finalement, l'échelle $\mathcal{N}(500 ; 100)$.

Conformément aux techniques issues de la psychométrie (théorie des réponses aux items), PISA définit alors l'indice de difficulté d'un item comme étant la valeur de l'indice à partir duquel un individu a une probabilité au moins égal à 0,5 de réussir l'item.

Cette organisation a permis à PISA de définir des niveaux de compétence. Là encore, la définition de ces niveaux est assez complexe et fait interagir une démarche qualitative (jugement d'experts) et une démarche quantitative. Après quelques itérations du processus et stabilisation du résultat, on a obtenu un découpage de l'échelle de compétences en 6 niveaux (plus un).



Un élève est donc au niveau 6 s'il a un indice de compétence égal ou supérieur à 669, tandis que dire qu'un item est au niveau 6, c'est dire que la probabilité d'un élève de niveau 6 de réussir cet item est supérieure ou égale à 0,5. Notons que la définition de ces niveaux permet d'assurer que, si l'on considère un ensemble d'items dont les indices de difficulté appartiennent tous, par exemple, à l'intervalle [607 ; 669], l'espérance mathématique du score d'un individu de niveau 5 sur cet ensemble d'items est supérieur ou égal à 50% (espérance mathématique de la loi binomiale de paramètre 0,5).

La façon dont la construction de cette échelle prend en compte l'analyse des tâches permet de donner un sens à ces niveaux et permet de les décrire (cf. Bodin 2005) On peut alors, par exemple, comparer les proportions d'élèves qui, dans chaque pays, se trouvent à tel ou tel niveau de compétence.

4.4. Les mathématiques dans PISA

S'il convient de savoir interpréter les scores de PISA, il est nécessaire de connaître le cadre de référence, lequel donne accès aux conceptions qui orientent l'étude, et le questionnement lui-même. Ces éléments sont facilement accessibles (cf. références).

PISA a une approche utilitaire des mathématique et se demande dans quelles classes de problèmes les compétences mathématiques pourront s'utiliser. L'idée n'est pas nouvelle et rejoint, en particulier, celle des problématiques de l'APMEP. Cependant, comme nous l'avons signalé plus haut, sur ce point, PISA s'est directement inspiré d'un ouvrage publié par le conseil national de la recherche des USA (Steen, 1990).

Le domaine mathématique est découpé en quatre sous-domaines : Quantité, Espace et forme, Relations et variations, Incertitude. Les expressions utilisées reflètent bien l'idée que ce ne sont pas les connaissances dans les domaines mathématiques classiques (géométrie, algèbre, ...) qu'il s'agit d'évaluer, mais bien la façon dont ces connaissances peuvent être mobilisées dans des situations relevant d'une analyse non scolaire des besoins.

Le terme d'incertitude, par exemple, recouvre ce que l'on peut appeler l'aléatoire ; du moins si l'on admet que les statistiques, lorsqu'elles ne se limitent pas au dénombrement, participent de l'aléatoire (choix des échantillons, etc...)

«Le terme ‘uncertainty’ est utilisé pour suggérer deux sujets liés : données et hasard. Aucun des deux n'est un sujet mathématique. D'une façon un peu rapide, on peut dire que les statistiques et les probabilités sont les domaine des mathématiques qui prennent en charge, respectivement, les données et le hasard» (David Moore dans Steen, op cité)

Cette distance prise avec les mathématiques savantes illustre le type de double transposition valorisé par PISA. Le « réel » est supposé constituer la source comme le cadre des situations d'évaluation proposées, tandis que les mathématiques plus ou moins savantes, apprises au cours de la scolarité, sont supposées se mobiliser naturellement dans les dites situations. D'une certaine façon, PISA porte sur ce que l'on a pu appeler les mathématiques mixtes, mais traduire, comme cela est souvent fait, « data » par statistiques et « chance » par probabilités abolit cette distance ; distance dont il n'est alors même plus possible de discuter la pertinence épistémologique.

PISA considère 3 grandes classes de compétences : Reproduction, Connexions et Réflexion. Ces classes servent, simultanément, à l'analyse des compétences (ce qu'il faut être en mesure de mettre en œuvre dans tel type de situation), à la création et à l'analyse des tâches proposées (cf. annexe 3).

4.5. Réflexions d'ordre épistémologique et didactique

Le domaine mathématique peut être étendu ou restreint suivant les conceptions que l'on peut avoir. Certaines questions du domaine mathématique de PISA sont de nature à surprendre les professeurs français. Ils n'y reconnaissent pas nécessairement les mathématiques qu'ils s'efforcent d'enseigner. En même temps, ils reconnaissent en général l'utilité sociale du savoir impliqué par ces questions. Même chose pour les mathématiciens (professionnels) lorsqu'ils s'intéressent à la question : en effet, l'appartenance des questions de PISA dans la construction des théories mathématiques ne paraît pas toujours évidente.

Quantités, variations et relations, espace et formes, incertitude,...sont modélisés dans les théories mathématiques, mais sont aussi présentes dans les situations communes, utilisant le sens commun et le langage commun.

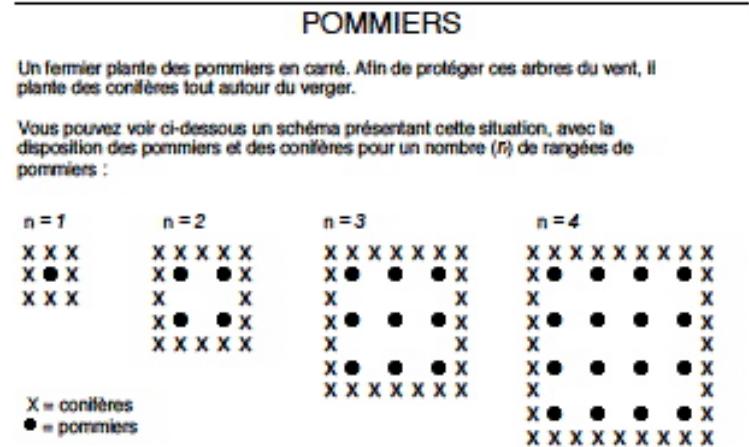
PISA, dans son souci de coller au réel, ne peut pas éviter d'utiliser abondamment le langage courant pour présenter ses questions. Dans certains cas, la compréhension de textes qui ne peuvent, en aucune façon, être qualifiés de textes mathématiques, devient la principale difficulté que les élèves ont à affronter. Bien sûr cela fait partie de l'activité mathématique normale, mais le vrai travail mathématique ne commence qu'une fois cette étape franchie. Ici, il n'est jamais sûr que l'habillage des questions où le type de langage utilisé ne soit pas ce qui empêche nombre d'élèves de résoudre des questions qui, sur le plan mathématique, peuvent apparaître comme tout à fait triviales.

En fait, nombres, quantités, etc... apparaissent aussi dans les questions de PISA relatives aux domaines "science" et "problem solving"¹⁷. Il n'est pas toujours évident de comprendre pourquoi PISA place telle ou telle question dans l'un des domaines "science" ou "problem solving" (ou même "lecture") plutôt que dans le domaine mathématique. En particulier, certaines questions du domaine "problem solving" qui induisent des activités de type logique auraient pu être ajoutées aux questions du domaine mathématique pour l'analyse que nous faisons ici.

L'exemple de la question "pommiers"

Cette question est typique des mathématiques de la vie réelle (et de l'évaluation authentique) que l'OCDE cherche à promouvoir. Dans ce contexte, selon le cadre de référence de l'étude, une "bonne question" doit amener l'élève à :

1. Partir de la réalité.
2. Organiser le problème en fonction de concepts mathématiques.
3. Effacer progressivement la réalité au travers de divers processus, tels que la formulation d'hypothèses concernant l'identification des principales caractéristiques du problème, la généralisation et la formalisation (...et de transformer le problème réel en un problème mathématique qui soit le reflet fidèle de la situation).
4. Résoudre le problème mathématique.
5. Comprendre la solution mathématique et l'appliquer à la situation réelle (ce qui implique aussi d'identifier les limites de la solution).

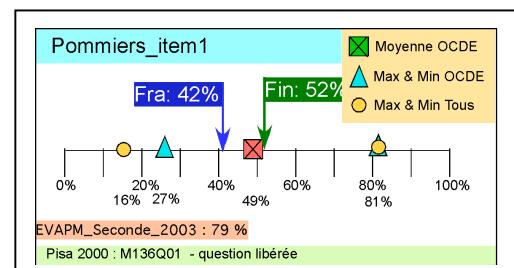


Question 1 : POMMIERS

M136Q01 - DF 02 11 12 21 89

Complétez le tableau :

n	Nombre de pommiers	Nombre de conifères
1	1	8
2	4	
3		
4		
5		



¹⁷ Qu'il serait trompeur de traduire par "résolution de problèmes"

Question 2 : POMMIERS

M136Q02- 00 11 12 13 14 15 99

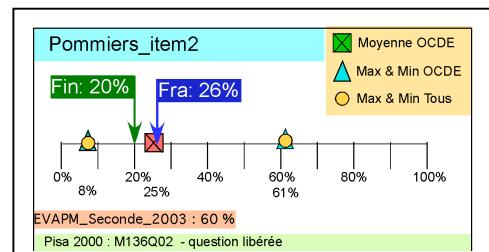
Il existe deux expressions que vous pouvez utiliser pour calculer le nombre de pommiers et le nombre de conifères dans cette situation :

$$\text{Nombre de pommiers} = n^2$$

$$\text{Nombre de conifères} = 8n$$

où n est le nombre de rangées de pommiers.

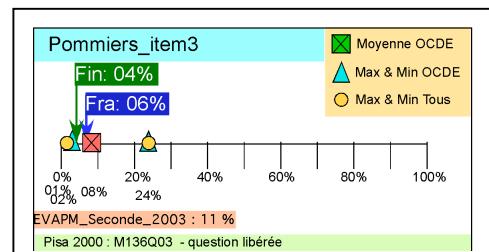
Il existe une valeur de n pour laquelle le nombre de pommiers est égal au nombre de conifères. Trouvez cette valeur de n et expliquez votre méthode pour la calculer.



Question 3 : POMMIERS

M136Q03- 01 02 11 21 99

Supposez que le fermier veuille faire un verger beaucoup plus grand, avec de nombreuses rangées d'arbres. Lorsque le fermier agrandit le verger, qu'est-ce qui va augmenter le plus vite : le nombre de pommiers ou le nombre de conifères ? Expliquez comment vous avez trouvé votre réponse.



Pour l'item 1, il s'agit d'abord de comprendre la situation (l'histoire !) et, ensuite, d'être capable de compléter et d'extrapoler un tableau. Il suffit de compter pour pouvoir compléter les quatre premières lignes. Pour la cinquième ligne, l'élève peut soit prolonger la suite de figures et compter, soit utiliser les régularités du tableau¹⁸.

Les 10% de différence entre les résultats finlandais et français (voir encadré Pommiers_item 1)¹⁹ illustrent la relative timidité des élèves français devant ce genre de questions (traduisant un manque certain d'initiative). Une partie d'entre eux, ne disposant pas d'une procédure apprise, toute prête, pour répondre, sont arrêtés à ce point.

À l'inverse, ceux qui surmontent cette première difficulté réussissent nettement mieux l'item 2 que leurs homologues finlandais (62% au lieu de 38%, si l'on ne prend en compte que les élèves qui ont franchi la première étape). Ce cas est généralisable.

Pour cet item, la mathématisation est quasi évidente et conduit à une équation à résoudre dans IN : $n^2 = 8n$.

Les élèves français sont familiarisés avec ce type d'équation qu'ils rencontrent le plus souvent directement sous forme ... formelle, non réaliste. Nous pouvons même supposer qu'une partie d'entre eux auront fait une résolution mathématiquement correcte, factorisant $[n(n - 8) = 0]$, trouvant les deux valeurs 0 et 8, puis éliminant la valeur 0 et reconnaissant que la valeur 8 est la seule qui convienne pour le problème donné (point 5 ci-dessus).

¹⁸ La littérature anglo-américaine parle dans ce cas de "pattern".

¹⁹ Les pourcentages donnés sont les pourcentages de réussite observés sur l'ensemble des élèves ayant passé la question. Ici, 52% des finlandais ont réussi l'item 1 contre 42% des français. Le cadre indique aussi les scores maximums enregistrés pour les pays de l'OCDE (81%, pour la Corée) et pour l'ensemble des pays ayant participé à l'étude (dans ce cas, aussi 81%) et les scores minimums pour les pays de l'OCDE (27%, pour le Mexique) et pour l'ensemble des pays ayant participé à l'étude (16%, pour le Brésil).

Les différences de pourcentages évoqués sont obtenus à partir des pourcentages bruts enregistrés dans les pays.

Cependant une partie des élèves (en France comme en Finlande) auront donné la réponse exacte en utilisant la démarche incorrecte suivante :

$$n^2 = 8n \Leftrightarrow n = 8 \text{ or } n \times n = 8 \times n \Leftrightarrow n \times n = 8 \times n.$$

Une autre procédure consistait à étendre le tableau jusqu'à la ligne $n = 8$.

Ces procédures parmi lesquelles au moins une est mathématiquement incorrecte ont été considérées comme correctes (il est vrai avec attribution d'une part des points ou d'une partie selon le cas). Cela soulève une question de nature épistémologique : quelles mathématiques sont en jeu ? qu'est-ce qui est valorisé ?

Il ne s'agit pas pour nous de mettre en cause l'intérêt de cette question, ni sa pertinence dans le test, ni même sa légitimité à être intégrée de la façon choisie dans la construction d'une échelle (où tout ce qui permet d'obtenir la réponse attendue tend à être considéré comme valable !). Ce que nous voulons soulever est le besoin d'études qualitatives complémentaires destinées à analyser d'un point de vue mathématique et didactique les procédures suivies par les élèves.

L'item 3 demande de comparer des "vitesses" de variation. Cela pourrait conduire à comparer la croissance des dérivées des fonctions f telle que $f(n) = n^2$ et g telle que $g(n) = 8n$, et donc, en fin de compte à comparer des dérivées seconde.

Cependant, les élèves concernés ne sont pas supposés avoir étudié les dérivées... il leur est seulement demandé d'avoir une approche personnelle de la question. Plusieurs procédures sont possibles, qui ont des intérêts mathématiques différentes, mais qui sont traitées comme équivalentes.

On notera que la question n'est aucunement triviale et il n'est pas surprenant que si peu d'élèves, non seulement en Finlande et en France, mais aussi dans le reste du monde parviennent à la résoudre d'une façon ou d'une autre.

Il est intéressant de constater, à travers ce type d'étude (ce fut aussi le cas pour TIMSS), que les vraies difficultés mathématiques, c'est-à-dire les difficultés liées aux concepts et non à l'habillage des questions sont ressenties de façon assez semblable à travers le monde.

Les 6% de succès indiqués pour l'item 3 en France, et les 4% indiqués pour la Finlande ne concernent que les procédures totalement correctes. Ils sont à comparer avec les 11% enregistrés dans les mêmes conditions au Japon et aux 11% obtenus en France, par EVAPM, en fin de seconde. Le seul pays qui fasse nettement mieux étant la Corée avec 24%.

La question des pommiers a en effet été utilisée en France dans le cadre de l'étude EVAPM de fin de seconde en 2003. Les résultats obtenus à chacun des items apparaissent dans les rectangles de résultats.

L'exemple des étagères

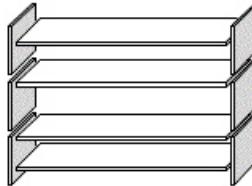
La question suivante est typiquement un cas de question mal adaptée au curriculum français : plus exactement, cette question paraît plutôt adaptée pour la fin de l'école primaire.

ETAGERES

Question 1 : ÉTAGÈRES

Pour construire une étagère complète, un menuisier a besoin du matériel suivant :

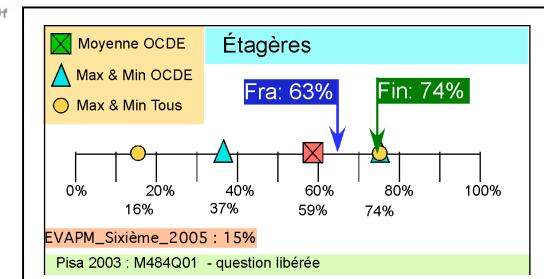
- 4 planches longues ;
- 6 planches courtes ;
- 12 petites équerres ;
- 2 grandes équerres ;
- 14 vis.



Le menuisier dispose d'un stock de 26 planches longues, 33 planches courtes, 200 petites équerres, 20 grandes équerres et 510 vis.

Combien d'étagères complètes le menuisier peut-il construire ?

Réponse :



En même temps, en France, chacun s'attend (et en particulier les enseignants de mathématique) à ce que les jeunes de 15 ans soient en mesure de résoudre cette question.

Le taux de succès est de 10% plus élevé en Finlande qu'en France ; cela illustre ce qui a déjà été dit sur les questions de la vie réelle.

Mais on peut se demander s'il s'agit bien d'une question de mathématiques. Ou, si l'on veut, doit-on considérer toute question impliquant des nombres comme une question de mathématiques (à tout niveau) ? Dans certains pays, cette question aurait mieux sa place dans une évaluation du domaine technologique.

Une solution mathématique pourrait être :

$$N = \min\left(\left[\frac{26}{4}\right]; \left[\frac{33}{6}\right]; \left[\frac{200}{12}\right]; \left[\frac{20}{2}\right]; \left[\frac{510}{14}\right]\right)$$

où N est le nombre maximum d'étagères que le menuisier peut faire et où $[x]$ désigne la partie entière de x .

Ici encore, il n'est pas attendu que les élèves écrivent une telle formule. En général, ils procèdent par une méthode du genre *essai-erreur*. Toutefois, s'ils devaient expliciter une preuve et s'ils devaient la formuler en langage courant, cela pourrait être encore plus difficile que d'écrire la formule ci-dessus. Ils devraient alors écrire le contenu et traduire le sens de cette formule.

Heureusement pour PISA, aucun élève ne pense à écrire une telle formule (nous ne le ferions pas davantage ailleurs que pour cet article !). De ce fait, les taux de réussite sont assez élevés : ils appartiennent pour la plupart à l'intervalle [50% ; 70%].

Mais s'agit-il encore de mathématiques ? Une bonne préparation aux questions concrètes est-elle aussi une bonne préparation à l'accès aux mathématiques plus abstraites ? Alors que de nombreux systèmes éducatifs incitent les enseignants à mettre l'accent sur le concret et la "vie réelle" (certains dans le but affiché d'obtenir de meilleurs résultats aux études internationales !), la question mérite d'être posée.

Cette question, avec beaucoup d'autres, montre aussi la faible importance que l'étude PISA accorde à la question de la preuve. Sans aller jusqu'à l'idée de démonstration (totalement absente), les processus d'explication et de justification sont très peu pris en compte. Cela fait évidemment, une grande différence avec les conceptions françaises habituelles concernant les acquis mathématiques.

Vers un minimum d'analyse didactique

La question précédente soulève directement des questions d'ordre didactique.

Quelles situations d'enseignement peuvent aider les élèves à acquérir, simultanément, des compétences dans les mathématiques du quotidien (qui sont pour une part les mathématiques du sens commun) et dans des niveaux mathématiques davantage abstraits et symboliques.

Certains objecteront que la question n'est pas pertinente et qu'il y a continuité entre les savoirs communs et le savoir théorique.

À l'opposé, comme le travail de l'école française de didactique nous a aidé à le comprendre, nous pensons que nombre de ruptures sont nécessaires et sont constitutives de la construction du savoir. Nous pouvons donc craindre que l'accent trop fort mis sur la vie réelle et sur les situations concrètes puisse, en retour, avoir des effets négatifs sur les apprentissages.

Voici encore un exemple.

La question des bonbons de couleur

Cette question est placée par PISA dans le domaine incertitude. Une valeur de probabilité est demandée.

L'étude des probabilités ne figure pas dans les programmes d'étude suivis par 98% des élèves sur lesquels porte l'étude. Toutefois, les résultats en France sont du même niveau que ceux qui sont enregistrés dans l'ensemble de l'OCDE et peu inférieurs à ceux qui sont obtenus dans certains pays où les probabilités sont enseignées depuis l'école primaire.

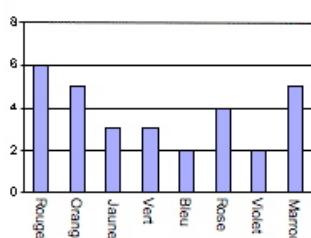
Nous avons observé le même phénomène avec l'étude TIMSS menée sur les élèves de 13 ans. Alors que l'étude des probabilités ne figurait pas au programme du niveau concerné (Quatrième en France), les élèves français obtenaient de meilleurs résultats que les élèves de pays où les probabilités figuraient dans les programmes. D'autres observations (EVAPM) montrent qu'au début de l'introduction des concepts mathématiques correspondants, les élèves ont davantage de difficulté à répondre à ce genre de question qu'avant d'avoir reçu un enseignement sur le sujet.

BONBONS DE COULEUR

Question 1 : BONBONS DE COULEUR

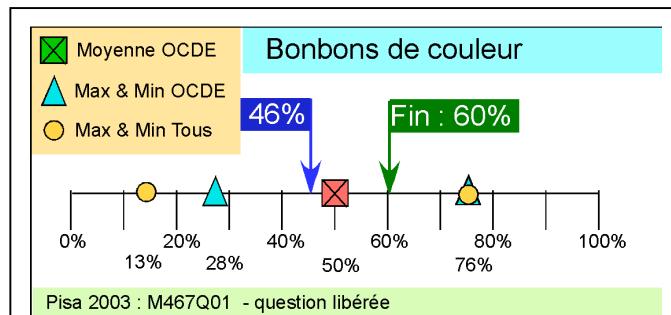
M467Q01

La mère de Robert lui permet de prendre un bonbon dans un sachet. Robert ne peut pas voir les bonbons. Le nombre de bonbons de chaque couleur qu'il y a dans le sachet est illustré dans le graphique suivant :



Quelle est la probabilité que Robert prenne un bonbon rouge ?

- A 10 %
- B 20 %
- C 25 %
- D 50 %



À nouveau, nous pouvons parler de savoir commun : comprendre le diagramme, compter le nombre total de bonbons (30), remarquer que 6 d'entre eux sont rouges et finalement interpréter les 6 chances sur 30 de tirer un bonbon rouge comme une probabilité.

Il s'agit plutôt de langage commun et de préconceptions relatives au concept de probabilité. Donner de l'importance à ce type de tâche, qui plus est en utilisant des questions à choix multiple, et laisser les élèves (et beaucoup d'autres) croire qu'ils ont acquis des connaissances dans le domaine des probabilités ne peut que conduire à des malentendus.

Bien d'autres questions de PISA nous semblent mériter ce type d'analyse.

5. Principaux enseignements des évaluations internationales et nationales

5.1. Remarques générales.

Que l'on s'intéresse aux curriculums, aux acquis des élèves, ou aux relations entre les variables de contexte et celles, personnelles, relatives aux élèves, les études permettent de mettre en évidence des différences importantes, mais aussi des similarités entre les pays ou les contextes éducatifs.

Il convient de noter que le pays n'est pas toujours l'unité la plus adéquate pour les comparaisons : dans de nombreux pays coexistent des systèmes éducatifs différents, et même lorsque ce n'est pas officiellement le cas, l'unité de façade est souvent contredite par les faits.

Les variations à l'intérieur d'un pays (entre régions par exemple) peuvent être aussi grandes que les variations entre les pays.

Par exemple, pour l'étude PISA 2003, l'Italie obtient un « score » considéré comme faible (indice 466), mais la région du Nord-Est de ce pays obtient exactement le même niveau d'indice (511) que la France.

Nous avons vu plus haut que ces indices qui font les délices des médias, et qui sont appelés scores de façon trompeuse, ne sont en fait que des valeurs prises par des indicateurs dont la construction est complexe et dont la signification est rarement explicitée.

Par exemple, pour PISA 2003, le *palmarès*, 511 pour la France, 548 pour la Finlande, signifie un différence de 0,33 écart-type sur l'échelle normale réduite (moyenne 0, écart-type 1). Cela, il est vrai, au voisinage d'un point de distribution où la densité de probabilité est maximale (près de la moyenne). Cette différence correspond à une différence moyenne de quelques pourcents dans les scores des réponses exactes question par question (de l'ordre de 3,5%²⁰).

Le fait que ces indicateurs soient le résultat de traitements de nature à réduire divers biais, ainsi que le fait qu'ils ont tous la même moyenne et la même dispersion sont de nature à faciliter certaines comparaisons. Il conviendrait cependant de ne les utiliser qu'avec prudence, et, aussi souvent que possible, leur associer des mesures plus directes (telles que des pourcentages).

Observons aussi que, si par malheur, les capacités intellectuelles de l'humanité (de chacun des humains vivants) se trouvaient d'un coup divisées par deux (capacités intellectuelles dans un sens qui peuvent, dans notre cas, rester non précisées), les valeurs de chacun des indicateurs en question resteraient inchangées.

²⁰ Valeur approximative estimée – un calcul plus précis reste à faire.

Dans le cas de la comparaison France-Finlande, par exemple, les amplitudes des différences, selon les items étudiés, s'étaisent entre +30% en faveur des élèves finlandais et +25% en faveur des élèves français, pour, comme nous l'avons dit plus haut, une moyenne des différences entre les scores obtenus aux items, d'environ 3,5% pour l'ensemble des items. Voilà de quoi relativiser le « miracle finlandais ». Ce miracle, s'il se confirmait, devrait plutôt conduire à regarder du côté de la qualité des rapports des élèves et des enseignants entre eux, et avec le système scolaire, ainsi que du côté du rapport au savoir et de la motivation des élèves, ce qui ferait déjà beaucoup et justifierait que l'on s'y arrêtât, que du côté d'une supériorité absolue dans la maîtrise des savoirs²¹.

Pour donner un autre point de repère, il suffirait, pour la France, d'exclure de l'étude les 10% des élèves qui réussissent le moins bien pour que les 90% restants affichent des résultats supérieurs à ceux de la Finlande.

Cela n'est pas dit pour minimiser les différences mais pour les situer plus correctement. Dit ainsi, on souligne aussi un échec, au moins relatif, du système français en ce qui concerne les élèves les plus en difficulté. Ce point avait déjà été souligné lors de l'étude TIMSS 1995 dans sa partie « *literacy* » au niveau des classes terminales.

On peut aussi penser que les résultats obtenus dépendent de l'importance donnée aux études dans les divers pays concernés. Dans certains pays, les enjeux des évaluations internationales sont ressentis comme importants, tant par les responsables aux plus hauts niveaux, que par les enseignants et par les élèves.

En France, notons que, pour les études EVAPM, les élèves obtiennent des résultats significativement meilleurs lorsque les enjeux de leurs performances les concernent directement (cas où leurs scores sont transformés en notes par leurs professeurs et utilisés comme telles).

Dans l'ensemble, les évaluations s'accompagnent de réductions importantes des enjeux curriculaires assumés par les systèmes éducatifs sur ce qu'il apparaît possible d'évaluer²². Sauf à considérer que le cadre de référence de telle étude s'imposerait comme standard universel et absolu pour tous les systèmes éducatifs, il n'est pas acceptable de considérer les résultats obtenus comme des indicateurs généraux de l'efficacité globale des systèmes éducatifs concernés.

Précisons ici certaines des réductions opérées :

- Pour TIMSS, il s'agissait de se limiter à un curriculum central (*core curriculum*) supposé être commun aux quelque 40 pays participants. Des pans entiers des compétences et des connaissances développées dans chaque pays sont ainsi laissés de côté.
- Pour PISA, le point de départ ayant été les besoins supposés du citoyen, il est clair qu'une partie importante des objectifs habituellement poursuivis dans les différents systèmes éducatifs n'a pas été prise en compte. Dans le cas Français, nous avons estimé à seulement 15% la part du curriculum couvert par PISA (A. Bodin, 2005). Il est toutefois exact que ces 15%, s'ils manquaient à la formation d'un individu seraient de nature à l'handicaper plus sérieusement que l'absence des 85% restants.

²¹ Ce point de vue fait écho aux échanges que nous avons eus avec des collègues finlandais lors du colloque franco-finlandais d'octobre 2005.

²² C'est l'« écrasement des objectifs sur l'évaluation » dénoncée en France par Guy Brousseau depuis les années 80.

5.2. Quels résultats ?

En ce qui concerne les seules mathématiques, on peut envisager de s'intéresser à plusieurs types de résultats produits par ces études. Il s'agit des connaissances que ces études apportent :

- sur les acquis des élèves,
- sur l'état des curriculums à travers le monde et sur les différences curriculaires,
- sur les relations entre les acquis des élèves et les curriculums correspondants,
- sur les relations entre les acquis des élèves et les différents facteurs, sociaux, contextuels et personnels, qui les accompagnent.

Chacun de ces points justifierait un long développement. L'ensemble des documents officiels des seules études internationales dont il est question dans cet article représente plus de 10 000 pages. Une bonne partie des synthèses reste à faire et cet article ne peut prétendre les faire. Comme nous l'avons dit précédemment, nous nous contenterons de donner quelques pistes concernant le premier point : les acquis des élèves.

Pour cela, on peut déjà répartir les pays en trois groupes :

- Groupe A : quatre ou cinq pays du continent asiatique obtiennent régulièrement des résultats nettement plus élevés que tous les autres pays : la Corée, Singapour, Hongkong, Taiwan et le Japon.

La supériorité de ces pays a parfois été mise en doute, mais les dernières études (TIMSS 2003 et PISA 2003) sont sans appel. Comme dans le cas des études précédentes, les résultats des élèves de ces pays sont effectivement meilleurs, et en général de beaucoup²³, que ceux des autres pays, et cela à tous les niveaux de la taxonomie (et non comme on l'a cru parfois, essentiellement au niveau « reproduction »).

- Groupe B : Les pays en voie de développement ou sous-développés, lorsqu'ils participent, obtiennent des résultats bas ou très bas. Cela montre surtout qu'il ne suffit pas d'imposer un curriculum exigeant pour que l'enseignement suive et pour que les apprentissages correspondants se fassent. Nous laisserons de côté ce cas qui, à l'évidence, a davantage à voir avec la question du développement qu'avec notre propos.

Notons cependant que si l'on s'intéressait à l'efficience, c'est-à-dire au rapport entre les investissements consentis et les résultats en termes d'acquis des élèves, certains des pays de ce groupe se situeraient très bien et auraient sans doute des enseignements à nous apporter.

- Groupe C : À côté de ces cas extrêmes, il reste la plupart des pays d'Europe, l'Amérique du Nord et la Russie. Ces pays présentent suffisamment de points communs, aussi bien en ce qui concerne les diverses variables contextuelles qu'en ce qui concerne les résultats des élèves, pour que les comparaisons directes présentent quelque intérêt.

²³ Nous évitons l'expression « significativement meilleur » souvent utilisée, mais qui est souvent trompeuse. On sait en effet que plus les échantillons observés ont des effectifs importants et plus la moindre différence devient significative... d'un point de vue statistique. Autrement dit, et contrairement à ce que laisse entendre le langage courant, dire qu'une différence est significative ne renseigne en rien sur l'importance de cette différence.

- En ce qui concerne les acquis des élèves, dans les diverses études internationales, les scores moyens de réussites de ces pays se situent dans l'intervalle [50% - 70%], et sont assez regroupés autour du milieu de cet intervalle (60%).
- Au niveau des questions d'évaluation, les différences observées dans les scores peuvent être importants, sans être systématiquement ordonnés de la même façon que les scores moyens des pays.
C'est à ce niveau, sans doute, que les comparaisons internationales sont susceptibles de présenter le plus d'intérêt pour la compréhension des phénomènes d'enseignement et d'apprentissage des mathématiques.
- Seules ne semblent bien réussies (au-dessus de 70% de réussite) que les questions qui font l'objet d'un entraînement intensif, et qui sont posées dans des formes familières aux élèves. Bien sûr, de telles questions obtiennent des scores de réussite très différents selon les curriculums concernés (une nouvelle fois, c'est davantage le curriculum que le pays qui constitue la variable indépendante).
- D'une façon générale, dans les trois groupes de pays, les différences varient selon les domaines et sous-domaines concernés (numérique, géométrique,...), ce qui peut informer sur les effets des curriculums sur les résultats. Dans l'ensemble, on observe surtout que le niveau des résultats dans un domaine est positivement lié à l'imposition aux contenus correspondants. Le cas des probabilités souligné plus haut constituant toutefois une exception.
- Très souvent, des questions qui paraissent faciles, mais qui demandent un petit pas de côté par rapport à ce qui est bien entraîné, sont assez mal réussies.
 - Par exemple, dans l'étude TIMSS 95, un rectangle étant donné dessiné, on demandait aux élèves de 13 ans de dessiner un second rectangle dont la longueur devait être une fois et demi la longueur du rectangle donné et dont la largeur devait être la moitié de la largeur du rectangle donné. Sur 45 pays, il ne s'en est trouvé qu'un (la Corée) pour lequel le taux de réussite à la question a été supérieur à 50%. (France : 43%). Il serait facile de multiplier les exemples de ce type.
 - Partout, on constate que les compétences, les vraies, celles qui correspondent à des tâches dont le niveau taxonomique dépasse celui de la reproduction, mettent en difficulté un pourcentage important d'élèves.
- Les concepts importants résistent partout. Par exemple, dans l'étude TIMSS 95, un des items les plus faciles du champ « proportionnalité » demandait ce qui se passerait si, dans une classe dont les trois-cinquième des élèves étaient des filles, on ajoutait 5 filles et 5 garçons. Pour l'ensemble des pays, moins des deux-tiers des élèves ont identifié la bonne réponse : à savoir qu'il y aurait encore plus de filles que de garçons dans cette classe.
- Dans la plupart des cas, les compétences manifestées par les élèves dans l'exécution des tâches proposées sont bien inférieures à celles qui pourraient être inférées de l'analyse des programmes, des commentaires officiels, des manuels, ou... des sujets d'examen. Autrement dit, il y a loin du curriculum officiel au curriculum atteint.

5.3. Les acquis des élèves en France

Dans un rapport récent de l'IGEN²⁴ sur l'évaluation des acquis (cf. références), il est clairement exprimé que l'on (l'institution ?) ne sait pas grand-chose sur la réalité des acquis des élèves, et que le peu que l'on sait est peu utilisé. On peut en particulier y lire :

(On a l') « impression que l'école, en France, depuis longtemps n'est pas au clair, comme elle a pu l'être dans le passé ou comme le sont les écoles de certains pays, sur les rapports entre la prescription nationale des programmes et la réalité des apprentissages et des acquis des élèves. « On ne sait pas bien » ce que les élèves apprennent et cette ignorance est néfaste pour tous : les pilotes approchent avec difficulté les « résultats », les professeurs y trouvent confirmation de leur fréquent malaise et les élèves de leur passivité tout aussi fréquente. Le temps semble urgent, dans un contexte général qui fait parfois douter l'école d'elle-même, de reprendre la question et de lui rendre tout motif de croire en elle en l'assurant mieux sur la construction, le suivi, la mesure et la certification de ce qu'elle enseigne. »

Le jugement est sévère et porte sur l'ensemble des disciplines. En ce qui concerne les mathématiques, cependant, ce jugement doit être largement nuancé.

Il est exact que la participation de la France aux études internationales est trop erratique et que les analyses nationales qui sont faites sont trop lacunaires pour que l'on puisse en tirer des enseignements sur la durée, comme peuvent le faire un certain nombre de pays, mais on peut déjà en tirer pas mal d'informations utiles.

En prenant en compte l'ensemble des études internationales et nationales (DEPP, mais aussi EVAPM), on sait en effet déjà beaucoup de choses sur les acquis des élèves en mathématiques. Il manque toutefois :

- Des études portant sur le primaire et la participation de la France aux études internationales à ce niveau est un manque que l'on ne peut que regretter. Doit-on y voir une stratégie du type de celle de l'autruche, un mépris pour ce qui se passe ailleurs, encore ou confiance définitive dans notre supériorité ? L'argument souvent avancé du coût des études paraît en tout cas assez faible.
- Des études s'inscrivant dans la durée et se prolongeant au delà de l'enseignement secondaire.
- Des études portant sur les niveaux supérieurs des taxonomies.
- Des études ciblées, en particulier sur :
 - Les 10% des élèves qui rencontrent le plus de difficulté.
 - Les 10% des élèves qui réussissent le mieux, en particulier en terminale scientifique.
- Des synthèses prenant en compte les aspects temporels et les niveaux scolaires. De telles synthèses devraient être faites par des équipes de recherche pluri-disciplinaires et pluri-institutionnelles.

Les études de la DEPP, pour intéressantes qu'elles soient, ne portent que sur quelques niveaux et, comme nous l'avons déjà signalé, sont orientées vers l'évaluation diagnostique. Elles sont de nature

²⁴ Inspection Générale de l'Éducation Nationale

à faciliter la mise en place des situations d'apprentissage, mais ne permettent pas de faire le bilan des acquis.

Les études EVAPM sont, par construction aussi exhaustives qu'il est possible, même si elles laissent, elles aussi, passer une partie de ce qu'il serait souhaitable d'évaluer. Ce sont des études à grande échelle, qui portent régulièrement sur des milliers de classes. Leur représentativité a pu être mise en doute, compte tenu du caractère volontaire de la participation des enseignants. Cependant une des études a pu être coordonnée avec une étude sur échantillon représentatif menée par la DEPP (Troisième 1990). La comparaison a permis de constater que le biais, s'il existe, est faible et correspondait à une différence maximum de 4% sur les taux de réussite aux items, ce qui est sans importance pour la plupart des conclusions que l'on peut tirer de ce genre d'étude.

En ce qui concerne les observations qu'il est possible de faire à partir des sources citées, et non plus, seulement, des études internationales, nous proposerons quelques pistes générales :

- Plus l'évaluation se situe près des apprentissages et plus les résultats semblent bons, du moins tant que l'on se limite, ce qui est fréquent, aux premiers niveaux des taxonomies. L'identification du fonctionnement de ces savoirs dans des situations éloignées de l'apprentissage (tant d'un point de vue temporel que d'un point de vue contextuel), réserve souvent de mauvaises surprises (utilisation de la proportionnalité en chimie à l'université par exemple).
- Une partie des études nous renseigne, de façon pointilliste, sur la présence à un moment donné d'une connaissance ou d'un savoir-faire isolé. Avec EVAPM et avec PISA, en particulier, on sait en plus pas mal de choses sur les compétences de niveaux supérieurs, et sur le caractère (peu) mobilisable et sur le caractère (peu) disponible des savoirs.
- L'appel aux connaissances des années antérieures, non entraînées en cours d'année, donne en général des résultats faibles, voire « catastrophiques ».
- L'éloignement des conditions habituelles de présentation et d'utilisation des notions, les habillages originaux, les appels à modélisation, amènent toujours des résultats faibles et sont régulièrement source de déception pour tous ceux qui cherchent à faire évoluer l'enseignement des mathématiques, d'autant plus, que dans le cadre du contrat didactique local, les élèves peuvent avoir manifesté des compétences qu'ils ne manifestent plus dans des conditions éloignées de ce contrat.
- Les résultats rapportés aux différents niveaux taxonomiques sont assez faiblement corrélés. En clair, des élèves à l'aise dans des situations faisant appel à l'analyse, la synthèse, la créativité, mais demandant des connaissances explicites réduites, peuvent être moins performants dans des situations ne demandant que l'application de connaissances isolées, et réciproquement.
- Les enseignants sont très souvent déçus, pour ne pas dire davantage, par les résultats obtenus par leurs élèves et cela malgré les jugements souvent sévères qu'ils portent sur eux « je les savais faibles, mais à ce point !... ». Ils comprennent difficilement qu'ils « ne sachent pas » ou qu'ils « ne sachent plus » !. Nous avons déjà dit qu'il s'agissait là, pour une part, d'un effet de rupture de contrat et il n'y a pas lieu de penser que le jugement porté directement par les enseignants sur les acquis de leurs élèves soit systématiquement moins fondé que celui que l'on peut tirer d'une observation ponctuelle externe.

Dire comme c'est souvent le cas que, dans le cas des études internationales, que la France se situe dans la moyenne des pays concernés ne devrait pas suffire à nous rassurer. On peut aussi bien dire que la France apparaît comme médiocre.

Par exemple, dans le cas de PISA, la moyenne des scores des pays de l'OCDE est de 58,2% tandis que les moyenne des scores en France est de 58,8%. Sachant que les pays de l'OCDE incluent des pays qui rencontrent des difficultés importantes qui les classeraient plutôt dans le groupe B ci-dessus, chanter cocorico est pour le moins déplacé.

Les études PISA (2000 et 2003) montrent en particulier, et cela est confirmé par les autres études, que, en France :

- Les élèves ont des connaissances, mais que ces connaissances sont peu disponibles. Pour la plupart d'entre eux, si on ne leur dit pas explicitement quelles connaissances mathématiques particulières il convient d'utiliser dans une situation donnée, ils ne le trouveront pas d'eux-mêmes, même s'ils possèdent le ou les éléments de connaissance correspondants.
- Les élèves font preuve de peu d'autonomie. Ils ne s'attaquent qu'aux questions qu'ils pensent pouvoir résoudre, ils ne disposent pas de stratégies pour aborder un problème qui ne leur est pas familier : essayer, expérimenter, bricoler... ne font pas partie de leurs modes d'approche possible.
- Les élèves sont légèrement défavorisés par les QCM. Cela, surtout parce qu'ils ne répondent pas n'importe comment lorsqu'ils ne savent pas. Ce point devrait donc plutôt être considéré comme positif. L'attribution de points négatifs pour réponse erronée serait plus correcte et supprimerait cet inconvénient.

Si l'on s'intéresse aux mathématiques pour tous, PISA ne fait que confirmer ce que l'étude TIMSS 1995 montrait à l'évidence :

- Nos élèves qui ne feront pas de mathématiques spécialisées partent mal armés pour la vie, pour comprendre le monde qui les entoure, pour participer aux divers débats de société qui les attendent. Des efforts de tous types devraient être consentis si l'on veut faire face à cette situation inquiétante.

À l'autre extrême, si l'on s'intéresse aux élèves qui poursuivront des études demandant des mathématiques à un niveau plus ou moins spécialisé, on peut observer que :

- D'un point de vue absolu, on ne peut que s'inquiéter de la distance entre les attentes et les observations que l'on peut faire sur les acquis des élèves (cf. EVAPM terminale 1999).
- Cela ne signifie pas cependant que les élèves de cette catégorie n'ont pas acquis des compétences et des connaissances, mais que, bien souvent, la distance entre ce qu'ils peuvent faire et ce qu'on leur demande de faire est trop grande. Cela est valable, sans doute, pour l'évaluation, mais cela l'est aussi pour l'enseignement : on n'apprend pas lorsque la distance entre le connu et le nouveau est trop importante (théorie de la zone proximale de développement – L. Vygotsky).
- D'un point de vue comparatif, seule l'étude TIMSS 1995 relative aux « spécialistes » est susceptible de nous renseigner. Dans cette étude, la France se distinguait nettement dans le peloton de tête, mais... aucun pays asiatique n'y avait participé. Il s'agissait de plus, pour l'essentiel, d'évaluer des savoirs formels, ne correspondant qu'à une partie des compétences susceptibles de prédire la réussite dans les études ultérieures. D'autre part, il semblerait que la

situation se soit dégradée entre 1995 et 2005 (voir plus loin).

Lorsque l'on évoque les résultats français dans les études internationales, il est souvent évoqué l'idée que ces études seraient culturellement marquées et, donc, culturellement biaisées. Il serait plus exact de dire qu'elles sont curriculairement biaisées, ce qui vaudrait aussi, plus ou moins, pour tous les pays participants. Les biais culturels, lorsqu'ils existent, sont de nature à relativiser la faiblesse des résultats, non à les nier²⁵. D'autre part, comme cela a déjà été dit, il ne tiendrait qu'à la France d'être plus influente dans ces études.

Outre ce qui a déjà été dit sur le cas français, les études EVAPM nous permettent de faire quelques remarques supplémentaires.

Malgré l'importance accordée dans notre curriculum à l'expression écrite, à l'argumentation et à la démonstration, les résultats que l'on peut observer dans ces domaines sont très loin des attentes.

- Au niveau du premier cycle, il est rare qu'une demande d'argumentation, voire même de justification obtienne plus de 20% de réussite.
- Au niveau du second cycle, si l'on prend en compte l'ensemble des élèves, il est rare qu'une demande de démonstration soit réussie par plus de 10% des élèves.

La question présentée ci-contre, posée en seconde et réussie par 10% des élèves, n'est qu'un exemple qui peut être généralisé.

On observera cependant que le taux de réussite des élèves qui s'orientent vers une première scientifique est supérieur à 25%.

D'autre part, cette question a donné lieu à six prise d'information qui permettent de préciser que, en ce qui concerne les futurs scientifiques, plus de la moitié d'entre eux ont produit des éléments de démonstration corrects.

- On peut tout aussi bien déplorer la faiblesse de ces résultats que penser que ces résultats sont le signe d'apprentissages en cours. Toutefois, les études diachroniques qu'EVAPM permettent de faire mettent en évidence une diminution dans le temps des compétences des élèves en ce qui concerne l'argumentation et la démonstration.
- Cette baisse concerne en fait l'ensemble des domaines sur lesquels les comparaisons sont possibles. On aurait pu penser que ces baisses, qui, concernent surtout des aspects techniques, seraient compensées par l'acquisition de nouvelles compétences (aptitude à organiser une situation, à chercher, à mobiliser ses connaissances,...). Malheureusement, les études PISA sont loin de confirmer cet espoir !

**Exemple de question demandant une démonstration
EVAPM - Seconde 2003 - Question GES03**

On donne la figure suivante sur laquelle les hypothèses ont été directement codées.

Démonstration correcte

Démontrer que le triangle ABC est isocèle.

Ensemble des élèves de seconde : 10%
Elèves orientés en première scientifique : 26%

²⁵ Le lecteur trouvera sur les sites de l'APMEP et de la SMF un diaporama présentant l'ensemble des questions de PISA qui ont été libérées, avec les résultats obtenus en France et dans quelques autres pays. Il pourra ainsi se faire une idée personnelle du caractère éventuellement culturellement biaisé de ces questions.

Dans le domaine du calcul, les résultats ne correspondent pas davantage aux attentes, et cela à tous les niveaux.

Nous présentons ici deux questions qui illustrent ce constat. Là encore, il ne s'agit que d'exemples de ce qui constitue le cas général et concerne tous les niveaux.

Pour davantage de précisions sur les différences observées au cours des 10 dernières années en ce qui concerne les acquis des élèves en mathématiques nous renvoyons à un texte publié dans un bulletin de l'APMEP (alerte aux maths ? – cf. référence) ainsi qu'aux notes de synthèse des études EVAPM 2005 de fin de sixième et de fin de première.

Exemple de question du domaine numérique
Suivi EVAPM - seconde 1991 à seconde 2003 - Question NAL068

Développer et réduire chacune des expressions suivantes :

($3x + 2y$)² = $3(x - 2yx) - 2x(x^2 - 3y)$ =

Réponse exacte :
Seconde 1991 : 72%
Seconde 2003 : 50%

Réponse exacte :
Seconde 1991 : 64%
Seconde 2003 : 46%

Exemple de question du domaine numérique
Suivi EVAPM - seconde 1991 à seconde 2003 - Question NAL068

Une personne a emprunté sans intérêt 1000 F.
Elle a déjà remboursé une somme S.
Il lui reste à rembourser une somme égale aux $\frac{2}{3}$ de la somme S déjà rendue.
Calculer S en laissant le détail des calculs.

Explications

Réponse exacte (en fin d'année) :
Quatrième 1989 (EVAPM) : 12%
Troisième 1990 (EVAPM) : 31% (48% pour les élèves orientés en seconde)
Troisième 1984 (étude Spresse - MEN) : 23%
Seconde 1991 (EVAPM) : 58%
Seconde 2003 : 30%

Sur ce type de question, les élèves de seconde 2003 ne sont pas au niveau des élèves de troisième 1990 orientés en seconde.
Par rapport aux élèves de seconde 1991, la chute est de près de 50%.

Précisons qu'il n'est pas dans nos intentions de dire que rien ne va plus. Le point d'interrogation du titre « alerte aux maths ? » n'est pas fortuit et il reste de la place pour l'interprétation et pour le jugement. On peut toutefois noter que l'encadrement mathématique des élèves ayant subi des baisses considérables ces dernières années il est assez prévisible que ces baisses puissent avoir un effet sur le niveau des acquis des élèves. Cependant, on ne peut pas exclure que les changements de programmes et de pratiques aient des effets positifs que les études évaluatives faites à ce jour ne permettent pas de mettre en évidence.

L'absence de la France des études TIMSS après 1995 ne permet malheureusement pas l'étude des variations dans le temps des acquis des élèves, comme cela a pu être fait dans d'autres pays, avec dans certains cas l'enregistrement de progrès significatifs correspondant à des prises de consciences et à des investissements publics importants (cas, par exemple, des USA). Dans beaucoup de pays toutefois, les comparaisons entre les résultats des études TIMSS 1995, 1999 et 2003 montrent des baisses qui semblent importantes.

Pour revenir au cas français, les études internationales ne permettent aucune comparaison sur ce qui constitue nos spécificités nationales, telles que la place des aspects formels des mathématiques, de la preuve et de l'argumentation. Il faudrait que des études complémentaires soient menées sur ce point (au niveau européen par exemple). Pour l'instant rien ne prouve que nos faiblesses absolues soient aussi des faiblesses relatives. Tous les systèmes éducatifs éprouvent des difficultés à faire coïncider leurs ambitions avec les résultats observés. Il serait important que des mises à plat soient faites qui permettent de baser des actions sur l'état réel de la situation et non sur un imaginaire dépassé.

6. Retombées des études

6.1. En général

Les études et les évaluations dont nous parlons ont toutes pour objectif déclaré l'amélioration de la connaissance des résultats de la formation des élèves pour, finalement, contribuer à l'amélioration de cette formation.

Cette amélioration passe en général par des actions sur le contexte, ou, du moins, sur la prise en compte différenciée de ce contexte (incitation politique à l'amélioration des conditions de vie, discrimination positive,...) qui, malgré leur importance, seront laissées hors du champ de cet article.

Les retombées directes de ces études portent essentiellement sur des modifications des programmes d'enseignement, la formation des enseignants et les instructions qui peuvent leur être données.

Ces retombées dépendent largement de l'intérêt porté à ces études par les responsables des systèmes éducatifs et par la société elle-même. À ce propos, citons deux faits :

Aux USA « en 1988, après la publication finale des résultats de TIMSS, le président Clinton réunit les responsables du gouvernement, du monde des affaires, de l'éducation et des chercheurs en éducation, pour discuter de la façon dont le pays pouvait répondre aux performances (mauvaises) révélées par l'étude. Depuis cette réunion, beaucoup d'autres ont eu lieu et TIMSS continue à avoir un impact sur l'éducation aux USA... » (selon Robiltaille, 2000)

Au Japon "Les résultats de TIMSS ont fait l'objet d'une communication importante à la télévision et par l'intermédiaire de sociétés professionnelles telles que la société japonaise des formateurs d'enseignants de mathématiques... » (selon Robiltaille, 2000)

Retombées sur les programmes d'enseignement

Les programmes d'enseignement constituent la variable sur laquelle il est le plus facile d'agir ; on ne s'étonnera donc pas que, assez souvent, les retombées les plus visibles aient lieu à ce niveau et que parfois elles s'y arrêtent.

Le pilotage du curriculum par l'évaluation a largement été dénoncé dans les années 80 (H. Freudenthal, Westbury,...). Il conduit à adapter les programmes aux résultats et donc à diminuer les attentes, ouvrant la voie à un cycle d'ajustements successifs vers le bas.

Il est toutefois des retombées positives, telles que l'introduction généralisée de questions liées à l'incertitude (statistiques et probabilités). Des pays qui avaient un enseignement extrêmement formel et procédural ouvrent leurs curriculums à la résolution de problèmes non stéréotypés, plus ouvert sur l'aspect outil des mathématiques.

La place de la géométrie et en particulier de la géométrie de l'espace est souvent revisitée en relation avec les études évaluatives.

Retombées sur la formation des enseignants

Moins évidente et plus importante est la retombée sur la formation initiale et continue des enseignants.

Au pire il peut s'agir d'un simple conditionnement pour essayer d'obtenir des meilleurs résultats à la prochaine étude internationale. C'est malheureusement ce qui a tendance à se passer, dans quelques pays, avec les résultats de PISA.

Bien entendu, si l'on considère que PISA produit une « mesure » valide de la formation mathématique dans son ensemble, ce que PISA ne fait pas et ne prétend pas faire, alors on peut trouver souhaitable que tout l'effort de formation des enseignants porte sur leur adaptation au cadre de référence de PISA et aux questions de PISA, pour que les élèves soient bien préparés aux futures épreuves (PISA 2006).

Là encore, il s'agit d'une façon détournée de piloter le curriculum par l'évaluation au lieu, comme cela est souhaitable de laisser le curriculum piloter l'évaluation.

Mais chaque fois que l'on préserve un bon équilibre entre la réflexion curriculaire (que faut-il enseigner, pourquoi ? comment ?) et l'utilisation des observations faites dans le cadre des études évaluatives, des actions de formation des enseignants peuvent être organisées avec profit.

Cela est d'autant plus vrai que les enseignants sont associés à la réflexion et ne reçoivent pas des instructions préparées dans un cénacle d'initiés.

Dans de nombreux pays, suite aux études internationales, selon les besoins détectés, des formations d'enseignants ont lieu sur la recherche de problème, sur la mathématisation des situations issues de la vie « réelle », sur les statistiques et la pensée probabiliste, sur la géométrie,...

Retombées sur les pratiques pédagogiques

Les retombées sur les pratiques sont plus difficiles à mettre en évidence.

Les études internationales véhiculent toutefois l'idée d'un élève actif, acteur de ses apprentissages et d'un enseignant qui n'est pas seulement un distributeur de connaissances, mais qui est surtout un organisateur d'apprentissages. Cette conception semble faire son chemin dans tous les pays et les études dont nous parlons ne sont pas étrangères à cette évolution.

Il ne faut toutefois pas trop penser à des effets directs mais plutôt à une sorte de percolation²⁶, laquelle, on le sait, ne se produit que lorsque qu'un certain seuil de déclenchement est atteint. Ce qui peut se passer lorsque les responsables mettent les informations dont ils disposent à la disposition des enseignants et des acteurs de leur formation et lorsque des actions de formation concertées sont mises en place.

Retombées sur acquis des élèves

Signalons que plusieurs études menées aux USA montrent que des gains significatifs, parfois importants, font suite à l'inscription des établissements dans un programme d'évaluation qu'il soit fédéral ou mené au niveau des États. Il y a sans doute là un effet adaptation et il conviendrait d'examiner de plus près la validité des instruments d'évaluation utilisés et des cadres de référence correspondants, mais le fait mérite d'être noté.

6.2. En France

Comparé à ce qui se passe dans d'autres pays, peu d'efforts de communication et assez peu de travail d'approfondissement est fait, en France, autour des études internationales. Des travaux intéressants sont cependant menés par des chercheurs plus ou moins isolés et par la DEPP (voir références). La encore, il manque des synthèses et des productions en direction des enseignants et des formateurs d'enseignants²⁷. Il manque aussi des analyses du contenu même des questions

²⁶ Selon une expression de Guy Brousseau qui nous paraît particulièrement adaptée ici.

²⁷ Voir cependant les articles d'Yves Olivier et de Claire Dupé (cf. références)

utilisées dans ces études, analyses dont certaines devraient être menées par des équipes de didacticiens, de mathématiciens et d'enseignants de mathématiques.

La plupart des travaux se focalisent sur les résultats et sur les relations entre les variables de contexte (inputs) et les résultats (outputs) ; cela dans la tradition du courant de recherche sur l'efficacité des écoles (*orientation school effectiveness*). Aujourd'hui, les recherches s'orientent plutôt sur l'étude des processus en vue de leur amélioration (*orientation school improvement*) (cf. M. Bru, M. N. Bottani, & al.). Ces recherches, qui prennent davantage en compte les pratiques d'enseignement, ne devraient plus faire l'impasse sur la question de la validité épistémologique, ni de la validité didactique, des questions utilisées pour les évaluations.

L'école française de recherche en didactique des mathématiques est particulièrement bien armée pour entreprendre ce genre d'études. Les travaux qui pourraient être menés dans ce cadre seront utiles dans le cas français, mais ils seront aussi de nature à renforcer l'intérêt que notre culture scolaire rencontre dans le monde.

En ce qui concerne l'utilisation des résultats des évaluations, citons à nouveau le rapport déjà cité de l'IGEN :

« Les résultats des évaluations nationales ou internationales diffusent peu en profondeur dans notre système : la DEPP communique largement à travers des conférences de presse, des publications écrites et la mise en ligne de documents ; les inspections générales relaient l'information auprès des inspecteurs territoriaux ; mais la diffusion semble aléatoire et non systématique : ces résultats, s'ils contribuent au pilotage des disciplines, le font inégalement et au hasard des préoccupations du moment. »

Là encore, le jugement est sévère et, en ce qui concerne les études internationales, nous ne pouvons que le partager. Cependant, le jugement paraît moins fondé en ce qui concerne les études nationales, du moins dans le domaine des mathématiques (rappelons que le rapport concerne l'ensemble des disciplines).

Même si ces évaluations n'ont pas bouleversé les conceptions et les pratiques des enseignants en matière d'évaluation, elles ont largement contribué à les modifier.

Dans de nombreux collèges et lycées, les enseignants ont pris l'habitude de se réunir pour analyser les résultats des évaluations et pour envisager les réponses pédagogiques à apporter.

En cela les enseignants ont été soutenus par de nombreuses recherches et par la publication de nombreux articles et documents produits par la DEPP, par les inspections nationales et régionales, et par l'APMEP et les IREM.

Des formations d'accompagnement impliquant en particulier les IREM ont en particulier amené les enseignants à réfléchir aux difficultés conceptuelles du domaine mathématique et à davantage considérer les erreurs comme des symptômes de ces difficultés que comme des fautes.

Certes, la distinction ou l'articulation entre l'évaluation formative et l'évaluation bilan continue à être mal assurée, mais cela tient sans doute autant au jeu des contraintes souvent antagonistes auxquelles les enseignants sont soumis qu'à un manque de clairvoyance de leur part.

Des erreurs ont pu être commises et il reste sans doute beaucoup à faire pour que l'évaluation puisse aider les enseignants à prendre les élèves là où ils en sont pour les amener le plus loin qu'il soit possible, mais il ne semble pas correct de dire que les évaluations qu'elles soient nationales ou internationales soient restées sans effet.

7. Le cas français²⁸

7.1. Résultats globaux

Le tableau suivant rassemble, pour trois pays dont la France, les scores officiels du domaine mathématique pour les volets 2000, 2003 et 2006 de l'étude.

La comparaison avec la Finlande s'impose dans la mesure où ce pays est souvent cité comme exemple. Nous avons ajouté l'Allemagne parce que dans ce pays, les résultats de l'enquête 2000, jugés mauvais ont suscité des réactions importantes et des remises en cause sévères.

À l'observation de ce tableau, qui vient confirmer d'autres alertes, on comprends que ces résultats ne laissent pas (ou plutôt, ne laissent plus) indifférents les responsables de notre système éducatif.

Le tableau suivant présente les

"vrais" scores moyens de l'ensemble des questions du domaine mathématique et de ses différents champs (sous-domaines). On constate que la différence entre les moyennes des scores français et finlandais est d'environ 6 points de pourcentage, soit une différence relative de plus de 10%. Cette différence est hautement significative d'un point de vue statistique.

PISA : les « scores » en mathématiques				
	Maths 2000	Maths 2003	Maths 2006	
FRANCE	517	511	496	
FINLANDE	536	544	548	
ALLEMAGNE	490	503	504	

Scores mathématiques PISA 2003					
	Maths TOUT	Quantité	Relations et variations	Espace et forme	Incertitude
France	53%	62%	53%	50%	48%
Finlande	59%	68%	56%	56%	56%
Japon	58%	65%	55%	59%	52%
OCDE	50%	58%	48%	48%	46%
TOUS	48%	56%	45%	46%	44%

Les différences de même type, pour chacun des champs de l'étude restent du même ordre (par exemple entre 3% et 8% selon les domaines, entre les résultats français et finlandais). Il s'ensuit que les analyses par domaine pour tenter d'expliquer ou de justifier des différences entre pays sont en général assez vaines. Ainsi, pour expliquer les mauvais résultats français relatifs au domaine "incertitude" on a fait valoir que notre curriculum faisait peu de place à l'aléatoire et que les statistiques étaient traitées chez nous d'un point de vue plus quantitatif que qualitatif, contrairement à ce qui se passe dans d'autre pays. Cela est en partie vrai, mais, surtout, les questions du champ "incertitude" sont perçues comme plus difficiles que celles des autres champs. Et cela dans tous les pays. Bien entendu, l'analyse par sous-domaines reste intéressante, mais à condition d'entrer dans le détail du questionnement.

Nous avons vu plus haut que PISA classait aussi les questions de l'étude en trois classes de compétence supposées hiérarchisées : Reproduction, Connexions et Réflexions (cf. Bodin, 2005). On peut alors se demander si la faiblesse relative de la France se traduit de la même façon pour chaque classe de compétences.

Le tableau suivant conduit à répondre oui à la question.

²⁸ Partie reprise de Bodin 2009, Gazette des mathématiciens (cf références)

	Reproduction	Connexions	Réflexion
Finlande	74,1%	55,2%	44,2%
France	68,7%	48,6%	38,6%
Japon	71,7%	53,2%	45,6%
OCDE	65,1%	45,7%	36,3%

7.2. Influence du format des questions ?

Le mode de questionnement de PISA peut aussi être interrogé. Il a souvent été dit que le mode de questionnement en QCM défavorisait nos élèves.

En réalité, seulement le tiers des questions mathématiques de PISA sont des QCM simples ou complexes (28 questions sur 85). Les autres questions sont des QROC (Questions à Réponses Ouvertes et Courtes) : 58 questions sur 85).

On peut encore distinguer les QROC à réponse forcée, notées ici QROCF, questions pour lesquelles on attend un nombre un mot ou une expression, et les questions à réponse étendue, notées ici QROCE, pour lesquelles on attend une justification.

L'examen du tableau ci-contre montre bien que les QCM ne creusent pas les différences. Compte tenu de l'importance donnée à l'argumentation et à la démonstration dans notre système, nous pensions que les résultats français des QROCE se rapprocheraient de ceux des pays qui sont en tête. La encore, le tableau montre qu'il n'en est rien.

	QCM	QROC	QROCF	CROCE
Finlande	59,6%	58,3%	66,9%	48,3%
France	53,3%	51,9%	61,2%	41,1%
Japon	59,1%	56,9%	66,3%	46,0%
OCDE	51,2%	48,8%	58,6%	37,4%

Plus généralement, nous avons essayé de chercher des différences en faveur des résultats français, mais nous devons dire que nous n'avons rien trouvé de ce type. Ce qui frappe plutôt, c'est que, en quelque sorte, nous sommes uniformément moins bons que les têtes de classe (Finlande, Japon). Un peu moins bons ou beaucoup moins bons reste une question d'appréciation.

7.3. Classement suivant les niveaux de compétence

Nous avons vu plus haut que PISA définissait 6 (+1) niveaux de compétence. Le tableau suivant permet de comparer la répartition des jeunes de 15 ans de l'ensemble des pays de l'OCDE et de ceux de la France et de la Finlande par rapport à ces niveaux.

Niveau de compétence mathématiques	Inférieur à 1	1	2	3	4	5	6
OCDE	11,0%	14,6%	21,2%	22,4%	17,6%	9,6%	3,5%
FRANCE	5,6%	11%	20,2%	25,9%	22,1%	11,6%	3,5%
FINLANDE	1,5%	5,3%	16%	27,7%	26,1%	16,7%	6,7%

Ces chiffres confirment d'autres études (TIMSS en particulier), qui, depuis longtemps, mettent en évidence qu'en ce qui concerne les mathématiques pour tous, ou, si l'on veut, les mathématiques du citoyen, notre pays réussissait plutôt mal, en particulier avec les élèves les plus en difficulté. Un autre fait apparaît, qui peut contredire quelques certitudes : nous ne sommes pas très bons, non plus, en ce qui concerne la formation des meilleurs (niveaux 5 et 6), au moins pour les compétences visées par PISA. Moins bons que la Finlande, bien sûr, mais aussi que la Suisse, le Canada, le Japon, la Corée, etc...

Il faut certes éviter d'étendre sans précautions les conclusions ci-dessus à l'ensemble de la formation mathématique. Il est possible que l'insistance mise dans notre curriculum, et dans nos pratiques, sur des mathématiques plus formelles que celles prises en compte par PISA (place de la démonstration, du symbolisme, de l'algèbre, de l'analyse,...), puisse profiter à nos meilleures élèves. Dans les études EVAPM, par exemple, les corrélations entre les réussites observées à des exercices de type PISA (concrets) et des exercices plus formels sont en général assez faibles. De leur côté, nos collègues finlandais dénoncent une focalisation trop grande de leur enseignement secondaire sur les situations de la vie réelle et le fait que cela se paie ensuite par des difficultés d'abstraction, de rigueur, et de formalisation.

Mais l'accès médiocre de l'ensemble des jeunes aux mathématiques reconnues utiles pour la vie et pour la vie citoyenne en particulier, n'est pas sans importance. Les meilleurs d'aujourd'hui seront les élites de demain et si ces élites ne sont, au mieux, à l'aise que dans le formalisme mathématique, c'est l'équilibre même de la société, pour ne pas dire la démocratie qui se trouvera menacée.

7.4. Tentatives d'explications

Pour une part les difficultés rencontrées doivent être rapportées aux difficultés générales de notre système éducatif, lesquelles doivent elles mêmes être rapportées aux difficultés générales de notre société. J'ai montré ailleurs (Bodin, 2005) qu'il suffirait d'"oublier" les 10% des jeunes les plus en difficulté pour nous retrouver au niveau de la Finlande (mais il faut déjà noter que pour des raisons obscures les DOM ont déjà été exclus de l'étude !).

D'autre part, les observateurs convergent pour expliquer que les élèves de notre système ont acquis des connaissances mais qu'ils ne sont pas bien préparés à mobiliser ces connaissances dans des situations qui ne sont pas soigneusement balisées. Ils savent faire (comparativement) si on leur dit ce qu'il faut faire mais sont désarmés s'ils doivent eux-mêmes mathématiser ou modéliser dans une situation qu'ils n'ont pas déjà, explicitement, rencontrée.

Ces mêmes observateurs, qu'ils soient français ou non, tendent à dénoncer une forme d'enseignement trop formel et trop procédural (bien que cela ne semble plus tout à fait vrai, au moins dans les premières années de l'enseignement secondaire).

Il faut encore noter que, entre les attaques visant l'"*impérialisme des mathématiques*", depuis longtemps dépassées en ce qui concerne l'enseignement obligatoire, et les rééquilibrages disciplinaires, la place des mathématiques dans la scolarité n'a pas cessé de se dégrader au fil des ans.

Il n'est donc pas possible de rejeter totalement les faiblesses constatées sur des défauts ou sur des biais dus à l'étude elle-même.

Ces biais existent cependant, mais il y a deux sortes de biais que l'on a trop tendance à confondre dans une même réprobation de PISA : les biais techniques et les biais curriculaires et culturels.

Les biais techniques sont nombreux et bien documentés (Hopmann, 2007). Il en existe à tous les niveaux des opérations : échantillonnage, recueil de l'information, condition de passation des tests, codage des réponses, traitements des données,... À des degrés divers, ces biais existent aussi pour les études nationales et certains sont inhérents à la démarche. Ils rendent ridicule l'excès de précision souvent affichée par les rapporteurs et les commentateurs de PISA (et le présent article n'y échappe pas qui donne des pourcentages à 0,1% près...), mais selon notre expérience personnelle de PISA et selon les comparaisons que nous avons pu faire dans le cadre de contre-enquêtes nationales

(Observatoire EVAPM), ces biais ont peu d'incidence sur les conclusions que l'on peut raisonnablement tirer de cette étude.

Les autres biais invoqués sont les biais curriculaires et les biais culturels : PISA n'évaluerait pas de façon conforme aux programmes, aux pratiques et aux attentes qui sont habituelles dans notre pays ; de plus l'organisation de PISA s'effectuerait dans un cadre largement anglo-saxon.

Ces biais sont réels et puissants. Le biais curriculaire mesure une sorte de distance entre nos conceptions de l'enseignement des mathématiques et celles que l'OCDE cherche à promouvoir. Mais parler alors de biais conduit à penser que notre curriculum serait idéal, ce qui n'a rien d'évident. On peut dire la même chose du biais culturel : le présenter ainsi équivaut à conférer une supériorité absolue à notre culture nationale.

Il n'en reste pas moins que la façon dont PISA recouvre nos objectifs d'enseignement est très faible et qu'aucun locuteur francophone, ni même de langue latine, n'est membre de l'équipe mathématique de PISA, et que d'une façon générale, l'influence française dans l'organisation de cette étude est très faible. Cela est surtout dû aux carences de la gouvernance française (incapable de choisir entre l'exclusion justifiée et la présence active et efficace)²⁹.

Une certaine conception de la "vie réelle" (mot d'ordre directeur de PISA) et des mathématiques pour tous (mathématiques du citoyen) conduit à exclure de l'étude toute démonstration ou recherche de preuve et, pratiquement, toute manipulation symbolique. D'une certaine façon, on peut dire que PISA est plus proche du certificat d'études d'antan que de notre baccalauréat, à ceci près que les qualités d'analyse des situations, d'initiative et de critique y sont beaucoup mieux prises en compte.

On est alors en droit de se demander ce qu'il reste de l'éducation mathématique et craindre qu'une adaptation sauvage aux objectifs de PISA (adaptation en cours dans de nombreux pays) nous éloigne de valeurs essentielles de la formation mathématique. Il y a là matière à débat et sans doute à action pour qu'une grande partie des élèves puisse continuer à faire des mathématiques et à pouvoir accéder à des études scientifiques dans de bonnes conditions. Mais, simultanément, il est souhaitable que quasiment tous les jeunes abordent leur vie d'adultes avec les outils leurs permettant une insertion réussie aux points de vue personnel, social et professionnel. Cela, qui constitue l'objectif principal de PISA est aussi, dans notre pays, et plus généralement en Europe (au moins), l'objectif officiel du socle commun de connaissances et de compétences. Articuler ces deux exigences ne sera pas facile et mérite de mobiliser davantage les efforts de la communauté mathématique.

7.5. Rapport au verbal et aux autres domaines disciplinaires

Les corrélations entre les résultats des différents domaines de l'étude sont très élevées : beaucoup plus que ce que l'on trouve habituellement à l'intérieur même des mathématiques par exemple entre algèbre et géométrie ou géométrie et gestion de données.

PISA 2003 : corrélations OCDE			
	Lecture	Science	Résolution de problèmes
Mathématiques	0,77	0,82	0,89
Lecture		0,83	0,82
Science			0,79

²⁹ L'OCDE a en effet été contrainte par le "gouvernement français" (sans doute quelque obscur gardien du temple) à se débarrasser de l'expert français qu'elle avait choisi pour les mathématiques. Il aurait, il est vrai, pu manquer de docilité.

Si l'on se place, non plus au niveau des individus, mais au niveau des pays, les coefficients de corrélation linéaire sont tous supérieurs à 0,95. Autrement dit, la connaissance du score d'un pays dans l'un des domaines permet d'avoir, directement, une bonne approximation du score de ce pays dans les autres domaines. À l'évidence, il y a un facteur commun à l'ensemble des domaines et ce facteur est très important.

L'importance prise dans le questionnement par la langue de communication est telle qu'il est probable que la langue soit ce facteur commun. En mathématiques comme dans les autres domaines, l'élève se trouve très souvent face à un texte important à lire, à comprendre, duquel il doit tirer les informations utiles et laisser de côté celles qui ne le sont pas. Comme l'exprime assez bien un chercheur anglais : "*dans PISA, il y a souvent peu de mathématiques à mobiliser, mais beaucoup à faire avant de pouvoir les utiliser.*" [8]. On peut alors penser que les caractéristiques linguistiques des langues utilisées influent aussi sur les résultats de PISA, et donc sur les différences entre pays.

Ces observations posent la question de la validité de l'étude relativement à ce qui peut faire la spécificité des mathématiques et donc, de la valeur des conclusions que l'on peut tirer sur la qualité de la formation mathématique des jeunes d'un pays à partir des seuls résultats de PISA.

7.6. Les réactions

Dans le monde, les réactions à PISA ont été très importantes et ont souvent conduit à de profondes remises en cause et à des mesures spécifiques. Pour s'en convaincre, sur le seul plan mathématique, il suffit de consulter les sites des associations de spécialistes et, dans nombre de cas, des sociétés mathématiques nationales.

En France, jusqu'à une période récente, après avoir fait quelques jours la une des journaux, PISA n'a pas intéressé grand monde, du moins officiellement. Les responsables ont tenté de minimiser la pertinence de l'étude en invoquant les biais évoqués plus haut. Les commentaires officiels ont surtout cherché à rassurer le public et les enseignants.

"En culture mathématique ...[les élèves] font preuve d'une relative aisance dans les activités qui reposent sur des supports « scolaires ». Ils savent néanmoins tirer parti de l'enseignement théorique dispensé ...pour affronter des exercices qui ne sont généralement pas pratiqués dans le cadre de l'école française."

"Les scores français se situent de façon significative au dessus du score moyen des pays de l'OCDE." DEPP (Ministère de l'éducation nationale) 04/12/2005

Le rapport public présentant une analyse assez complète de l'étude 2003 n'a été publiée qu'en 2007 ! (Dossier DEPP 180 – 2007). Ce rapport ne fait que confirmer les premières déclarations officielles et, creusant un peu les différents champs de l'étude, écrit :

Variation et relations : très bons résultats (!)

"...une performance particulièrement solide en Variations et Relations, champ où les élèves français montrent leurs compétences en matière de lecture, d'interprétation et d'exploitation de documents graphiques (courbes, tableaux), ou encore d'application de relations mathématiques comme la proportionnalité.

Le prélèvement d'information sur des supports divers est un point fort des élèves français, ce qui est probablement dû au fait qu'il est pratiqué dans plusieurs disciplines, dès le collège."

Espace et forme : bon niveau (!)

“Dans le champ Espace et Formes, les élèves français montrent également un bon niveau de compétence sur l’interprétation des configurations, sur des calculs d’aires et de périmètres ou l’appréhension de figures dans l’espace.”

Quantité : réussite relativement plus faible

“Les performances sont plus moyennes dans le champ Quantité qui fait appel au travail sur les nombres et au calcul”

Incertitude : hors curriculum

Il est surtout remarqué que les probabilités ne faisaient pas partie du curriculum français de l'école obligatoire (alors même qu'un regard sur les questions et une prise en considération du concept d'incertitude montrerait que les probabilités ne sont pas vraiment concernées).

Dans l'ensemble, le rapport se contente de pointer quelques points faibles :

“Les “points faibles” des élèves français semblent résider dans la capacité à effectuer des généralisations (par exemple, établir une formule) et, de façon générale, à prendre des initiatives sans se référer à un schéma connu, ou encore à faire des essais avant de répondre.”

Malgré cette apparente indifférence le système a cherché à s'adapter :

- Publication d'un décret définissant un socle commun de connaissances et de compétences précisant ce que tous les jeunes doivent maîtriser à la fin de leur scolarité obligatoire. Le décret se réfère explicitement à PISA.
- Modification plus ou moins profonde des questions d'examen (Brevet des collèges et baccalauréat). En particulier les questions du Brevet cherchent à être moins formelles et davantage ancrées sur des situations de la "vie réelle". Toutefois, l'essentiel est oublié, à savoir la dévolution au candidat des démarches à mettre en œuvre. On continue à baliser l'activité par un questionnement du type I. a), I, b) ... en n'oubliant pas de préciser : en utilisant tel théorème, en appliquant telle procédure,... ce qui s'inscrit en totale contradiction avec les conceptions développées et évaluées par PISA.
- Révision des programmes du collège pour y inclure une dose d'aléatoire.
- Révision des programmes du primaire, pour les élèves de 6 à 11 ans, en se référant aux mauvais résultats enregistrés par PISA entre 15 et 16 ans et sur la base, à mon avis (et pas seulement !), d'une totale incompréhension des objectifs privilégiés par PISA.

Les résultats de PISA 2006 ont été accompagnés d'un changement radical dans les déclarations officielles, changement que les variations réelles des taux de réussite entre 2003 et 2006 expliquent moins que les récents changements politiques. Quoi qu'il en soit, un certain consensus existe maintenant pour affirmer que les choses ne vont pas aussi bien que ce qui avait pu être annoncé et que des mesures doivent être prises (mais bien sûr le consensus n'existe pas en ce qui concerne ces mesures).

Le discours officiel a donc changé. Par exemple, le ministre de l'éducation nationale interrogé par France Culture le 3 novembre 2007 déclarait :

“Il n'est pas normal - nous allons le voir bientôt dans une enquête PISA à propos de la compétence mathématique en fin de collège - que la France ne cesse de baisser. Nous sommes très en-dessous de la moyenne européenne”

Ces mauvais résultats devaient conduire, selon le ministre, à "remettre de l'école dans l'école" et à "recentrer les missions de l'école primaire sur l'acquisition des savoirs fondamentaux".

Toujours selon le ministre (12 décembre 2007) :

“... la parution, au cours des dernières semaines, des deux enquêtes internationales PISA et PIRLS, a livré un constat alarmant sur l'état de notre système scolaire....

L'enquête PISA menée par l'OCDE auprès des élèves âgés de quinze ans montre que les résultats obtenus vers la fin de la scolarité obligatoire sont à la fois médiocres pour la culture scientifique, où la France se situe à peine dans la moyenne des pays de l'OCDE, inquiétants pour la compréhension de l'écrit, où la part des bons élèves recule et celle des élèves en difficulté régresse, et alarmants pour les mathématiques où les résultats de la France régressent et où la part des élèves les plus faibles augmente de 37%.”

Les informations diffusées dans les médias emboîtent le pas.

Le quotidien "Le Monde", qui n'est pas le plus acide, n'hésite pas à écrire (16 décembre 2007) :

“Alerte sur le « niveau » scolaire. Des signaux négatifs s'accumulent sur les performances de l'École en France.”

Certes, le problème est réel et il est heureux que l'on s'en préoccupe, mais il convient de noter que comparativement, cette situation n'est pas plus catastrophique que celles de la plupart des pays comparables (Allemagne, Royaume-Uni, USA, sans parler de l'Espagne, de l'Italie ou du Portugal). Tous les pays, y compris ceux qui figurent en tête du palmarès ont de quoi être inquiets.

En tout cas, si ces exagérations conduisent à prendre des mesures pertinentes et efficaces, et non à démoraliser davantage les enseignants, les alertes provoquées ou renforcées par PISA n'auront pas été inutiles.

On le sait, PISA n'est pas, en France, la seule source d'indicateurs sur les acquis mathématiques des élèves. Sur de nombreux points, PISA ne fait que confirmer d'autres études ou observations.

Malheureusement les sources d'information indépendantes sont rares et partielles. La France ne participe qu'avec réticence aux études internationales. Elle est sortie de TIMSS après 1995 et de ce fait n'a pas participé à l'étude de 2008 qui a répliqué l'étude menée en 1995 sur les acquis des futurs scientifiques en fin d'études secondaires. On peut toujours évoquer les insuffisances de ce type d'étude, mais d'une part cela nous prive d'indicateurs importants et de sources de réflexion utiles, et, d'autre part, cela nous fait regarder, dans le monde, comme des mauvais joueurs, qui préfèrent s'exclure des comparaisons possibles, chaque fois qu'elles risquent de tourner à leur désavantage.

Parmi les réflexions internes, en 2003, sous le titre "alerte aux mathématiques ?", l'association des professeurs de mathématiques (APMEP) a largement diffusé un texte résumant les résultats d'une étude à grande échelle (EVAPM) conduite à la fin de la classe de seconde. Cette alerte ne faisait que prolonger les inquiétudes manifestées après les études menées au niveau des classes terminales en 1999.

Les mathématiciens prennent leur part dans la dénonciation des insuffisances du système.

Ainsi, Laurent Lafforgue (Médaille Field) écrit dans "Le Figaro" du 4 décembre 2004 :

“... Les gens l'ignorent, mais on assiste à un naufrage ! On imagine encore que les petits Français sont bons en maths. Mais, ils sont désormais mauvais, et cela pose un problème aux universités et aux écoles supérieures, qui ont déjà commencé à baisser le niveau de leurs programmes. ...”

D'une façon plus modérée, mais tout aussi ferme, Jean-Pierre Bourguignon (directeur de l'institut des hautes études scientifiques - IHES) écrit dans le Monde (4/12/2007) :

"On est obligé de reconnaître que le système scolaire français ne réussit pas à monter tout le monde à un niveau convenable. On peut craindre que, de ce point de vue, les écarts ne se soient récemment creusés encore.

Il serait facile de multiplier les citations.

8. Impact possible de ces études sur les pays francophones

On observe dans la francophonie des proximités dans les conceptions et les curriculums, mais aussi de grandes différences en ce qui concerne l'enseignement des mathématiques. Il s'en suit une grande variabilité de l'impact que les études internationales ont, ou peuvent avoir, dans ces pays.

Nous avons vu l'impact important que finalement l'étude PISA a en France (socle commun de connaissance et de compétence, modification des programmes d'étude,...). Des effets de même type peuvent être observés en Belgique et en Suisse.

Le cas du Québec mériterait une étude particulière (voir la communication de Lucie Debois à EMF2009). Cette province obtient d'assez bons résultats dans les études internationales et l'enseignement des mathématiques semble y être un genre de compromis de fait entre celui que l'on rencontre habituellement dans les pays de langue anglaise et celui observé dans d'autres pays francophones.

Bien que le français soit l'une des langues officielles de l'OCDE, l'influence de la langue française, comme celle des cultures francophones, dans les études internationales, est très faible. Les recherches en didactique, qui pourraient avoir une influence dans l'élaboration des cadres de référence comme dans l'interprétation des résultats des études, sont méconnues. Cela est dû, bien sûr, à l'absence de francophones représentant de ce champ de recherche dans les instances tant techniques que décisionnelles qui encadrent les études. La primauté est donnée aux instances gouvernementales qui n'ont que faire des recherches et aux psychométriciens qui, bien que maîtrisant des techniques intéressantes et utiles, manifestent une certaine naïveté devant la complexité des domaines qu'il s'agit d'évaluer.

La tentation française, de se désintéresser de ces études a fait long feu. En France comme ailleurs dans la francophonie, à juste titre, ou de façon abusive, dans les pays qui y participent comme dans ceux qui n'y participent pas encore, les études internationales sont devenus des déterminants importants des politiques éducatives.

Pour le moment, dans nos pays, l'impact se fait surtout par le biais d'injonctions technocratiques qui donnent peu la parole à la recherche et aux acteurs des systèmes. Nous avons déjà dit l'intérêt que nous portons à ces études et aux enseignements que l'on peut en tirer, mais pour que les effets positifs l'emportent sur les effets négatifs prévisibles, il conviendrait que la communauté mathématique s'investisse davantage à tous les niveaux de leur élaboration et de leur traitement.

Un peu partout dans les pays francophones, l'enseignement des mathématiques est assez formel et prend mal en compte les élèves qui éprouvent des difficulté devant le formalisme et l'abstraction. Cet enseignement réussit plus ou moins bien avec une partie des élèves, mais ne parvient pas à entraîner les autres. On dénonce alors un manque de motivation puis un manque de bases chez les apprenants et l'on observe, en fin de compte, un rejet de la discipline chaque fois que c'est possible.

Le premier mérite des études internationales est de ramener l'attention sur les mathématiques pour tous, sur le nécessaire accès de tous à un ensemble de compétences où les mathématiques jouent un

rôle important. À nous d'agir pour que ce repositionnement de l'enseignement des mathématiques se fasse sur des objectifs à la fois socialement pertinents et épistémologiquement valides.

9. Conclusion

Qu'elles soient nationales ou internationales, les évaluations ne sont pas sans défaut et nous en avons souligné quelques-uns. Toutefois, à condition de ne pas les prendre comme des mesures intrinsèques de la qualité des enseignements, de la qualité des programmes, ou de la qualité du système lui-même, à condition donc de considérer leurs résultats comme des indicateurs qu'il est possible et souhaitable de questionner, ces études présentent une utilité indéniable.

Les évaluations ne sont pas des mesures au sens de la physique et il faudrait en finir avec la mystification qui conduit à publier des scores qui n'ont souvent que peu de signification. Toute évaluation ne produit qu'une image partielle et imparfaite de la réalité dont elle est censée rendre compte. De ce fait, toute évaluation doit rester ouverte à l'analyse et aux interprétations plurielles, pour, ensuite ouvrir la voie à l'action. Cela est bien sûr valable pour les décideurs, mais ça l'est tout autant pour les divers acteurs du système éducatif : professeurs, parents, élèves, associations, ainsi que les autres acteurs du champ social. Évidemment, tout le monde n'a pas la même compétence pour traiter de la question, et les compétences des uns et des autres doivent être respectées, mais tout le monde, est concerné par l'enseignement et tout le monde doit avoir accès à l'intégralité de l'information.

La politique du domaine réservé qui est souvent la règle en matière d'évaluation du système, en France comme dans la plupart des autres pays, correspond, dans le meilleur des cas, à un manque de confiance dans l'intelligence et la bonne foi des différents acteurs. L'inquiétude correspondante n'est pas, il est vrai, dénuée de tout fondement, mais ce n'est qu'en libérant l'information et en favorisant recherches et débats, que ces évaluations pourront contribuer à l'amélioration de l'enseignement des mathématiques et, ce qui va de pair, à une meilleure professionnalisation des enseignants de cette discipline.

Au niveau des études internationales, cette politique d'ouverture est d'ailleurs celle qui est suivie officiellement par l'OCDE comme par l'IEA (TIMSS) mais les relais nécessaires au niveau national ont parfois tendance à se faire discrets.

D'une façon générale, les différentes évaluations sont complémentaires : évaluations nationales officielles ou non, évaluations internationales PISA et TIMSS, etc. Des analyses prenant en compte ces diverses évaluations devraient être conduites par des équipes où les institutions devraient avoir leur place, sans pour autant avoir le monopole de l'interprétation. Des recherches indépendantes devraient être conduites sur les études déjà faites. Des enquêtes complémentaires devraient être suscitées et menées au niveau national et au niveau européen.

Les études internationales sont une bonne occasion de s'interroger sur la solidité et sur la qualité de nos systèmes. Elles fournissent des indicateurs objectifs indépendants dont il serait bien dommage de se priver. Elles sont aussi une occasion d'échanges et de partage à travers pays et cultures qui ne peut qu'enrichir notre propre vision de l'enseignement des mathématiques comme, d'une façon plus générale, de la culture de notre temps.

10. Références

- Adams, R.J., 2003, Response to « Cautions on OECD's Recent Educational survey (PISA), Oxford Review of Adams, R.J., 2003, Response to « Cautions on OECD's Recent Educational survey (PISA), Oxford Review of Education, 29(3)
- Hopmann, S., Brinek, G, Retzl, M., (éds) (2007) : PISA according to PISA. Wien: Lit Verlag.
- Anderson, W. A. : 2001, A taxonomy for learning, teaching, and assessing; a revision of Bloom's taxonomy of educational objectives. Longman.
- Bardi, A. M. & al. (2005) : Les acquis des élèves, pierre de touche de la valeur de l'école ? Inspection Générale de l'Éducation Nationale.
- Beaton, A. E. et al. (1999) : The benefits and limitations of International achievement studies. IIEP – UNESCO.
- Bodin A. : 1997, L'évaluation du savoir mathématique - Questions et méthodes. *Recherches en Didactique des Mathématiques*, Éditions La Pensée Sauvage, Grenoble.
- Bodin, A ; Straesser, R. ; Villani, V.: 2001, Niveaux de référence pour l'enseignement des mathématiques en Europe - Rapport international - Reference levels in School Mathematics Education in Europe - International report.
- Bodin, A : 2003, Comment classer les questions de mathématiques ? Communication au colloque international du Kangourou, Paris 7 novembre 2003. Article à paraître.
- Bodin, A (2009) : L'étude PISA pour les mathématiques. Résultats français et réactions. Gazette des mathématiciens N°120 (Société Mathématique de France).
- Bodin, A. (1992), Les mathématiques en fin de Troisième générale - évolution des compétences observées chez les élèves au cours des années 80. -In Rapport à Monsieur le Ministre de l'Education Nationale établi par Monsieur Claude Thélot , Directeur de la DEP (1992). Publié dans le dossier Education et formations n°17 d'octobre 1992 (DEP).et dans le bulletin de l'APMEP Février 1993
- Bodin, A. (2005) : Alerte aux maths ? Bulletin Grande Vitesse de l'APMEP – Septembre 2005
- Bodin, A. (2005) Classification des questions d'évaluation et cadre de référence des études PISA pour les mathématiques – présentation commentée. (document de travail - site web)
- Bodin, A. & Capponi, B. : 1996, Junior Secondary School Practices, *International Handbook of Mathematics Education*, Chapter 15, Teaching and learning Mathematics, A. Bishop & C. Laborde (eds), pp 565-613, Kluwer Academic Publishers, Dordrecht.
- Bodin, A.. (2005) : En français : Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français. Communication faite à la conférence Franco Finlandaise sur PISA. Aussi en anglais : "What does PISA really assess? What it doesn't? A French view." Sites Web de la SMF et de l'APMEP
- Bottani, N.& Vrignaud, P. (2005) : La France et les évaluations internationales. Haut Conseil de l'Évaluation de l'École.
- Clarke, D. 2003, International comparative Research in Mathematics Education : Of What, By Whom, for What, and How. Second international Handbook on Mathematics education, Kluwer academic Publishers.
- Cytermann, J.R., Demeuse, M. (2005) : La lecture des indicateurs internationaux en France. Haut Conseil de l'Évaluation de l'École.
- Demonty, I. & Fagnant, A. (2004) : Évaluation de la culture mathématique des jeunes de 15 ans (PISA). Ministère de la Communauté Française. Bruxelles.
- Dupé, C. & Olivier, Y. (2005) : Ce que l'évaluation PISA 2003 peut nous apprendre. Bulletin de l'APMEP N°460 - octobre 2005

- Freudenthal, H : 1975, Pupils' achievements internationally compared - The IEA. In Educational Studies in Mathematics - Vol 1975.
- Gras R. : 1977, *Contributions à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques* - Thèse- université de RENNES.
- Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., Jocelyn, L. (2004). International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results From the U.S. Perspective. (NCES 2005-003). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Lie, S. & al (2003) : Northern lights on PISA. Unity and diversity in the Nordic countries in PISA 2000. University of Oslo, Norway
- Livingston, K. ; McCall, J. (2005) : Evaluation : judgemental or developmental? European Journal of Teacher Education.
- Meuret, D. 2003 Considérations sur la confiance que l'on peut faire à PISA 2000. Intervention au colloque international de l'Agence Nationale de Lutte Contre l'Illetrisme sur l'évaluation des bas niveaux de compétences, Lyon, 5 novembre 2003
- Meuret, D. 2003 Pourquoi les jeunes français ont-ils à 15 ans des performances inférieures à celles des jeunes d'autres pays ? Revue française de Pédagogie, n°142, 89-104.
- Mullis, I & al. (2000) : TIMSS 1999 International Mathematics Report. TIMSS International Study Center.
- Mullis, I & al. (2004) : TIMSS 2003 International Mathematics Report. TIMSS & PIRLS International Study Center.
- Note DPD 04.12 (décembre) - Les élèves de 15 ans Premiers résultats de l'évaluation internationale PISA 2003
- OECD (2004), Problem Solving for Tomorrow's World: First measures of Cross-Curricular Competencies from PISA 2003
- OECD (2004), Technical report.
- OECD 2004, First results from PISA 2003. Executive summary.
- OECD 2004, Learning for Tomorrow's World: First results from PISA 2003
- OECD 2004, PISA 2003 Assessment Framework - Mathematics, Reading, Science and Problem Solving Knowledge and Skills
- Orivel, F. (3003) : De l'intérêt des comparaisons internationales en éducation.
- Robert A : 2003, Tâches mathématiques et activités des élèves : une discussion sur le jeu des adaptations introduites au démarrage des exercices cherchés en classe de collège. Petit x N°62
- Robitaille & al. (2000) : The impact of TIMSS on the Teaching & Learning of Mathematics & Science. Pacific Educational Press - Vancouver
- Robitaille D.F. & al : 1993, TIMSS Third International Mathematics and Science Study, Monography n°1, Curriculum Frameworks for Mathematics and Science. Pacific Educational Press, U.B.C, Vancouver
- Steen, L. A. (ed) (1990) On the shoulders of the giants - new approaches to numeracy. National Academic Press (Washington)
- Varcher, P. (2002), Évaluation des systèmes éducatifs par des batteries d'indicateurs du type PISA : vers une régression des pratiques d'évaluation dans les classes.

Adresses et contacts

Note officielle française : <http://www.educ-eval.education.fr/pisa2003.htm>

Cadre de référence et rapports internationaux : <http://www.pisa.oecd.org/>

Sur le site de l'APMEP, article, diaporamas et présentation des questions libérées avec des résultats

(en français et en anglais) http://www.apmep.asso.fr/rubrique.php3?id_rubrique=114

Antoine Bodin : antoinebodin@mac.com

site internet : <http://web.me.com/antoinebodin/pro/>

Annexe 1 : Les études internationales concernant les mathématiques de 1960 à 2012.

Année de passation	NOM de l'étude		Primaire 3-4	École moyenne 7-8	15 y.o.	Pre-université	Nb de pays(*)	Instigateur
1960	PTCS	Étude pilote sur 12 pays, commanditée par l'UNESCO		X			12	IEA/ Unesco
1964	FIMS	Première étude internationale sur l'enseignement des mathématiques.		X		X	12	IEA
1982	SIMS	Seconde étude internationale sur l'enseignement des mathématiques et des sciences.		X		X	19	IEA
1988	IEAP1	International Assessment of Educational Progress 1		X			6	ETS
1991	IEAP2	International Assessment of Educational Progress 2	X	X			20	ETS
1995	TIMSS	Troisième étude internationale sur l'enseignement des mathématiques et des sciences.	X	X		X	40	IEA
1999	TIMSS-repeat	RéPLICATION de TIMSS 1995		X			40	IEA
2000	PISA 2000	Littéracie mathématique : sujet secondaire			X		32	OCDE
2000	TIMSS VIDEO STUDY	Étude portant sur les pratiques d'enseignement.	?	X			5	IEA
2003	PISA 2003	Littéracie mathématique : sujet principal			X		41	OCDE
2003	TIMSS 2003	Trends in International Mathematics and Science Study 2003	X	X			55	IEA
2006	PISA 2006	Littéracie mathématique : sujet secondaire			X			OCDE
2007	TIMSS 2007	Trends in International Mathematics and Science Study 2007	X	X				IEA
2008	TIMSS Advanced 2008	Replication of TIMSS 1995 Advanced Mathematics and Physics (Pop 3)				X	10	IEA
2009	PISA 2009	Littéracie mathématique : sujet secondaire					66	OCDE
2011	TIMSS 2011	Trends in International Mathematics and Science Study 2007	X	X			60	IEA
2012	PISA 2012	Littéracie mathématique : sujet principal			X		?	OCDE

En gras : études auxquelles la France a pris part.

(*) Nombre approximatif dans certains cas – lorsqu'il y a lieu, le nombre de pays ayant participé dépend des niveaux concernés

Annexe 2 : Taxonomie des demandes cognitives pour la construction et l'analyse de tâches mathématiques – organisée par niveaux intégrés de complexité.

Version simplifiée, voir la taxonomie complète sur Internet

Catégorie générale		Sous catégorie	
A	Connaissance et reconnaissance	A1	des faits
		A2	du vocabulaire
		A3	des outils
		A4	des procédures
B	Compréhension	B1	des faits
		B2	du vocabulaire
		B3	des outils
		B4	des procédures
		B5	Des relations
		B6	Des situations
C	Application	C1	Dans des situations familiaires simples
		C2	Dans des situations familiaires moyennement complexes
		C3	Dans des situations familiaires complexes
D	Creativité	D1	Utiliser dans une situation nouvelle des outils et des procédures connus
		D2	Émission d'idées nouvelles
		D3	Création d'outils et de démarches personnelles
E	Jugement	E1	Production de jugements relatifs à des production externes
		E2	Auto-évaluation

Taxonomie adaptée par Antoine Bodin avec toute reconnaissance pour le travail initial et les conseils de R. Gras, ainsi que pour l'influence ultérieure de W. A. Anderson.

Annexe 3 : Classes de compétences

Selon PISA – voir description complète dans le cadre de référence de PISA

Niveau		Définitions de l'OCDE	
1	Reproduction	Les compétences classées dans ce groupe impliquent essentiellement la reproduction de connaissances déjà bien exercées	Reproduction
2	Connexions	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.	Mathématisation simple
3	Réflexion	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-problème qui contiennent plus d'éléments que celles du groupe connexions, et qui sont plus « originales » (ou peu familières).	Mathématisation complexe