

---

## LES MATHÉMATIQUES FACE AUX ÉVALUATIONS NATIONALES ET INTERNATIONALES

---

*De la première étude menée en 1960<sup>1</sup>  
aux études TIMSS<sup>2</sup> et PISA<sup>3</sup>... en passant  
par les études de la DEP<sup>4</sup> et d'EVAPM<sup>5</sup>*

Antoine BODIN  
Irem<sup>6</sup> de Franche-Comté

*Résumé : L'article cherche à faire le point sur la place que les mathématiques occupent dans les études et évaluations nationales et internationales depuis les années 60 où de telles études ont été initiées. Il présente rapidement l'évolution des cadres théoriques utilisés pour ces études. Il s'intéresse surtout à la question de la validité de ces études, à la signification qu'il convient de donner aux résultats publiés et à ce que ces études peuvent apporter, en particulier aux enseignants. Il présente quelques observations que ces études permettent de faire sur l'évolution du savoir mathématique des élèves au cours des dernières années. Il pose enfin la question de la relative absence de la France dans les études internationales.*

### Présentation

Le présent article répond à une commande : il s'agissait de faire le point sur la façon dont les mathématiques, en France, se situaient face aux évaluations internationales et nationales. Cela, dans le cadre d'un séminaire portant plus généralement sur l'évaluation des systèmes éducatifs<sup>7</sup>. L'article aborde donc, de façon assez générale, les questions suivantes :

- Ce que les évaluations internationales et nationales cherchent à évaluer et ce qui est vraiment évalué.

- Les spécificités de l'évaluation du domaine mathématique.
- Les obstacles rencontrés.
- Les cadres théoriques, les méthodes et les instruments d'évaluation.
- Les principaux enseignements des évaluations.
- Les effets et conséquences de ces évaluations.

Nous laisserons de côté les questions qui sont relativement indépendantes du ou des domaines concernés : questions méthodologiques

---

1 Étude pilote sur 12 pays, commanditée par l'UNESCO et menée par l'IEA.

2 Trends in International Mathematics and Science Study

3 PISA : Program for International Student Assessment - OCDE

4 DEP : Direction de l'Évaluation et de la Prospective du Ministère de l'Éducation Nationale.

5 EVAPM : Observatoire des programmes de mathématiques du secondaire, des acquis des élèves et du contexte

---

d'enseignement développé depuis 1986 par l'APMEP (Association des Professeurs de Mathématiques) et l'INRP.

6 Institut de Recherche sur l'Enseignement des Mathématiques.

7 Communication faite dans le cadre d'un séminaire de l'EHESS (École des Hautes Études en Sciences Sociales - sociologie et économie de l'éducation) - Paris 1er décembre 2005.

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 fidélité, 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).

## 1.— L'objet de l'évaluation

### 1.1. *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'usage<sup>8</sup> 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 certai-

<sup>8</sup> 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.

ne 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 constitueraient :

- un langage universel, et qu'elles seraient donc 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 fautive de la discipline.

### 1.2. *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.

#### A - Les mathématiques pratiques

Il s'agit des mathématiques reconnues comme utiles à tous, pour la vie de tous les jours, ... celles du citoyen-consommateur-employable ? (littéracie mathématique au sens de PISA<sup>9</sup>).

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 6<sup>ème</sup> 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'exer-

<sup>9</sup> Le mot anglais « literacy » a été malencontreusement traduit en français par « culture », ce qui est source de profonds malentendus.

cice 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 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 (voir §5).

Quelques remarques avant d'aller plus loin :

- Aucune évaluation n'est possible si l'on ne définit pas au préalable un référentiel d'évaluation<sup>10</sup>, 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, habilité, 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'hui<sup>11</sup>,... 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 la 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

10 Cadre de référence, en anglais : framework.

11 En anglais : « realistic math », « authentic assessment »

l'impression d'inutilité et d'artificialité des mathématiques.

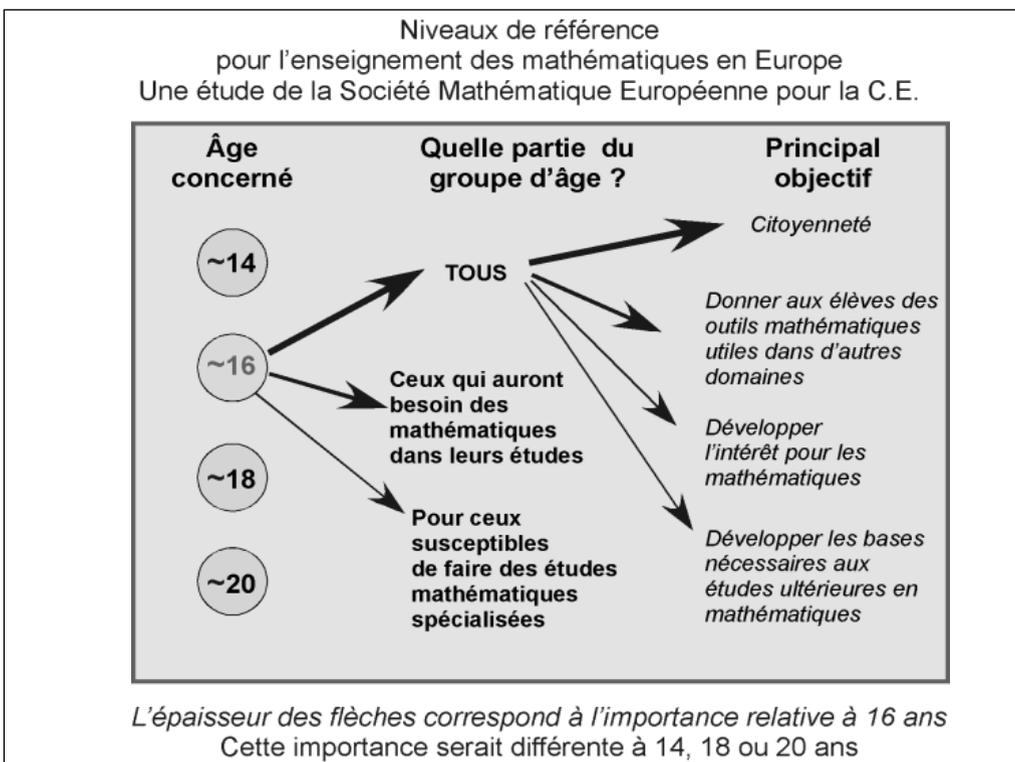
### 1.3. Les sujets de l'évaluation

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.

Les niveaux habituellement visés sont les 4ème, 8ème et 12ème années d'enseignement (en France, CM1, Quatrième et classes terminales).

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.

La figure ci-dessous est empruntée à une étude conduite pour la Commission Européenne, par la Société Mathématique Européenne (EMS)<sup>12</sup>. L'étude s'est attachée à définir des niveaux de référence en mathématiques



<sup>12</sup> Voir en fin d'article l'adresse Internet où l'on peut consulter les documents de cette étude.

à 16 ans, valables dans l'ensemble de l'espace européen. L'évaluation proprement dite de ces niveaux n'a pas été faite, mais leur définition serait une étape essentielle pour une évaluation de ce type. La figure cherche à articuler les différentes dimensions de l'objet à évaluer avec l'âge des élèves concernés.

## 2.— Les études

### 2.1. Études internationales portant sur les mathématiques

Plusieurs organisations internationales sont en concurrence plus ou moins ouverte :

- l'IEA (International Association for the Evaluation of Educational Achievement),
- l'ETS (Educational Testing Service),
- l'OCDE (Organisation de Coopération et de Développement Économiques).

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 (« *accountability* »), qui, partie des USA 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.

Je ne développerai pas cet aspect de la question, me contentant de renvoyer au rapport très documenté établi par N. Bottani et P. Vri gnault pour le Haut Conseil de l'Évaluation de l'École (cf. Références).

A propos d'unidimensionnalité du domaine, on lit en particulier dans ce rapport :

*« On peut ... dire que la plupart des enquêtes internationales privilégient l'unidimensionnalité. Cette option est en accord complet avec la théorie psychométrique et les modèles de mesure utilisés. Ceci a des conséquences importantes sur la présentation des résultats qui privilégieront l'ordre des performances des pays donc leur classement. »*

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.

Il convient de préciser que l'évaluation est devenue un marché mondial, âprement disputé dans le cadre des organisations citées plus haut par des instituts publics et privés. Dans cette compétition, la France ne fait malheureusement pas le poids, et cela en dépit de :

- La réputation, dans le monde, de l'enseignement français, et particulièrement de

l'enseignement français des mathématiques, qui reste bonne, même si des signes contraires commencent à se faire sentir.

- La haute estime dans laquelle est tenue l'École mathématique française, dont, à tort ou à raison, une partie des mérites est attribuée à la qualité des enseignements primaires et secondaires.
- L'intérêt porté aux recherches menées en France dans le cadre de la didactique des mathématiques.
- Les compétences méthodologiques en matière d'enquêtes, développées par la recherche en didactique des mathématiques, dans certains instituts (INRP, IREDU, certains IREMS,...) et par la DEP.

De nouvelles études sont en cours de préparation ; en particulier TIMSS 2007 sur les niveaux 9 ans et 13 ans et TIMSS 2008 qui apportera des informations comparées essentielles sur les acquis des futurs scientifiques à la fin de leurs études secondaires. Il serait dommage que la France n'y participe pas.

## 2.2. Études nationales

En France...

### Les études de la DEP

Entre 1980 et 1989 : sous des noms variables, la direction correspondante du Ministère de l'Éducation Nationale<sup>13</sup>, a mené des études bilans (en fin d'année scolaire), sur échantillons. En mathématiques, ces études ont essentiellement porté sur les niveaux CM2, Cinquième, Troisième, Seconde. Elles étaient discrètes et modestes, mais renseignaient

13 SIGES, SPRESSE, DEP, DPD, puis à nouveau DEP en attendant un nouveau nom qui pourrait faire référence à l'observation des Performances.

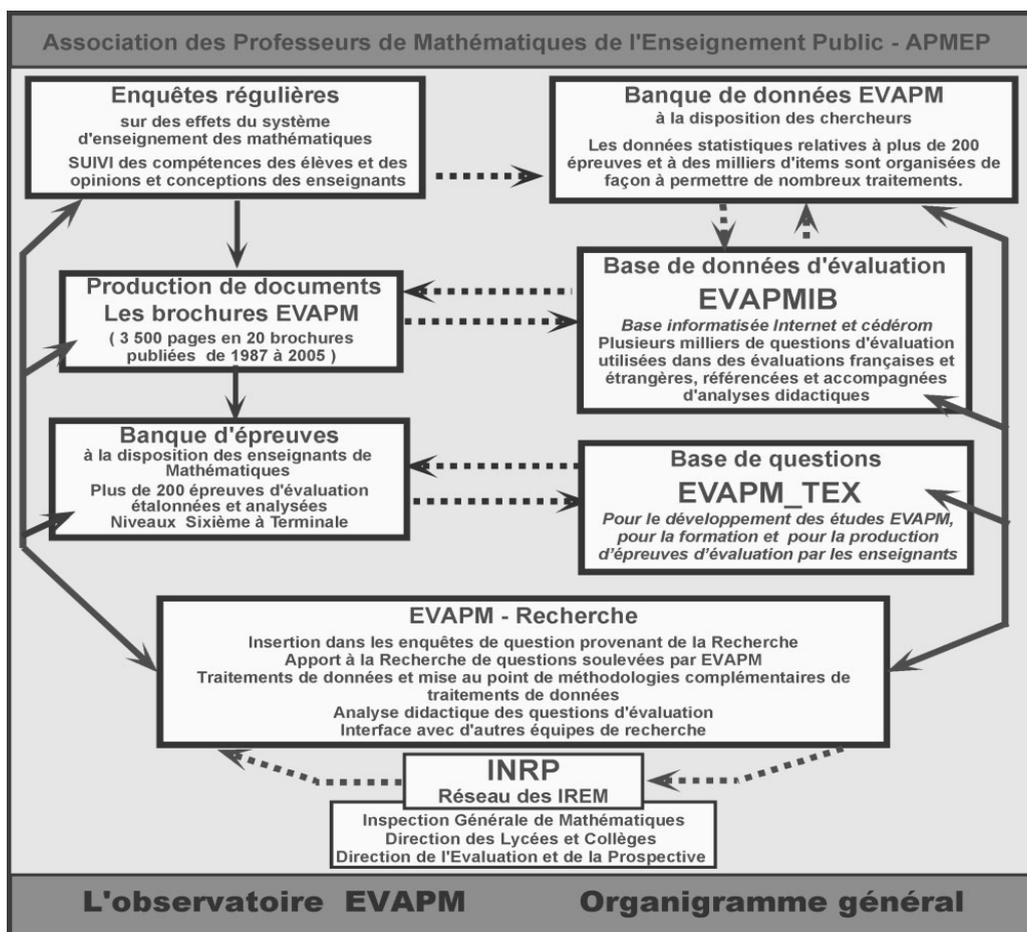
efficacement les responsables sur les acquis des élèves. Par contre, les enseignants n'étaient que très rarement au courant de ces études et de leurs résultats.

A partir de 1989 ces études ont été remplacées par des évaluations diagnostiques effectuées en début d'année scolaire. Il s'agit d'évaluations portant d'abord sur les élèves, tous les élèves d'une classe d'âge : CE2, Sixième, Cinquième, Seconde, selon les années... Ces actions d'évaluation, construites pour répondre à leur fonction pédagogique, peuvent éventuellement produire des informations utiles sur les acquis des élèves. Toutefois, le diagnostic ne concerne, à juste titre, que les points les plus sensibles, ceux dont la non-maîtrise serait susceptible d'entraver les apprentissages ultérieurs. Les questions sont en général calibrées pour être réussies par environ 70% des élèves. Utiliser ensuite les résultats, comme cela est régulièrement fait, comme indicateur global des acquis des élèves est une erreur méthodologique confinant à la malhonnêteté intellectuelle.

### Le cas d'EVAPM

Développé par L'APMEP et l'INRP, l'Observatoire EVAPM organise, depuis 1987, des études à grande échelle de la sixième aux classes terminales.

Les études portent, de façon aussi exhaustives que possible, sur les acquis des élèves, mais aussi sur le contexte d'enseignement et sur les conceptions des enseignants. Depuis sa création, l'observatoire a accumulé une masse importante de données statistiques qui lui permet de suivre l'évolution dans le temps des compétences acquises à un niveau donné et de suivre l'évolution des compétences à travers les différents niveaux de la scolarité.



Dans ce cadre, plusieurs bases de données ont été développées :

La base EVAPMIB, qui regroupe toutes les questions utilisées par EVAPM depuis 20 ans avec des questions provenant d'autres études nationales et internationales. La base garde trace des passations de chaque question, des résultats obtenus et des analyses faites.

La base EVAPM\_TEX\_PDF, qui est utilisée pour l'élaboration de nouvelles évaluations au niveau national ou local. L'utilisation du langage Tex a permis d'obtenir des éditions de questions et de textes mathématiques de qualité professionnelle (supérieur à ce qui a pu être observé jusqu'à présent dans les diverses études). De nombreux documents produits par EVAPM sont à la disposition des enseignants

et autres personnes intéressées, en particulier sur Internet (voir adresse en fin d'article).

Dans d'autres pays

On peut citer NAEP aux USA (qui a généré IAEP), mais aussi des études menées au Royaume-Uni (APU, CATS,...), en Italie, etc. En fait, la plupart des pays cherchent à se doter de dispositifs d'évaluation des acquis des élèves, très souvent en relation ou en complément des études internationales.

### 3.— Les outils de l'évaluation

Si la tendance dans certains milieux est de considérer les mathématiques comme génératrices 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 NLSMA<sup>14</sup> utilisée pour SIMS 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 :

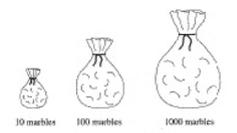
- Nombres et algèbre
- Géométrie
- Fonctions et Analyse
- Statistiques et probabilités

Une mise à plat de ce type a en particulier été élaborée pour l'étude TIMSS<sup>15</sup>. 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 (notons que, dans certains pays, les « mathématiques scolaires » constituent quasiment une discipline autonome).

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 intercurriculaires 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.

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

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



10 marbles      100 marbles      1000 marbles

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 %

14 NLSAM : New Longitinal Study of Mathematical Abilities

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

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 évident 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 est vrai pour TIMSS 1995 ne l'est plus pour PISA 2003).

**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 %

Par exemple, la question ci-dessus, où pourtant le terme « probabilité » est utilisé dans l'énoncé de la question, était réussie par 54 % des élèves 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 »<sup>16</sup>. 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)<sup>17</sup> :

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

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.

<sup>16</sup> Voir les problématiques de l'APMEP élaborées sous l'égide de Régis Gras.

<sup>17</sup> Voir sur le site de l'APMEP (Informations sur PISA) une description de ces problématiques.

### 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 Guttman<sup>18</sup> — le rêve de l'édu-mètre !).

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 définir<sup>19</sup>.

- 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 constitutif sont eux-mêmes des vecteurs appartenant à des espaces respectivement générés par :

- Les contenus (syllabus),
- Les types de compétences (démarches sollicitées et produits attendus),
- 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. Gras<sup>20</sup>.

18 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.

19 L'analyse implicite 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.

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

- Les évaluations nationales de la DEP 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.

### 3.3. Les types de questions

#### QCM et questions à choix forcés

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 en bas à gauche (TIMSS 95) est typique de cette catégorie de questions. Les études EVAPM utilisent des QCM plus riches, à « *pattern* » de réponse complexe, autorisant en plus la réponse « Je ne sais pas ». Cette ouverture nous semble en effet plus conforme au souci de rigueur qui devrait accompagner la formation mathématique que l'incitation à répondre au hasard lorsque l'on sait que l'on ne sait pas.

Par exemple, la question de la page ci-contre, posée en Première scientifique, admet pour bonne réponse le « *pattern* » FFVVFV. L'information recueillie par une telle question est beaucoup plus importante que si la question était posée en QCM simple. Il en est de même pour les possibilités de traitements qu'elle offre.

Les remarques qui précèdent ne nient cependant pas l'intérêt des questionnaires de type « Quiz » où la rapidité fait partie des qualités testées et où la prise de risque est valorisée (sélection rapide d'une réponse qui apparaît seulement plus vraisemblable que les autres), mais de tels questionnaires sont mieux adaptés à l'entraînement des connaissances et à l'évaluation diagnostique qu'aux évaluations bilans.

**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 ?

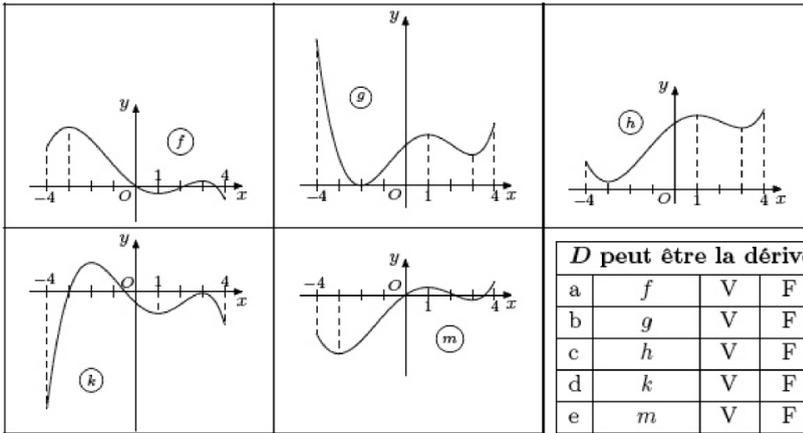
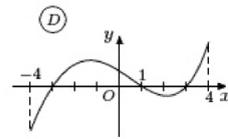
Réponse : Hélène a payé : .....

**Résultats niveau Quatrième (grade 8)**

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

Exemple de question en QCM à "pattern" de réponse complexe  
EVAPM - Première S 2005 - Question ANA105Q

Soit  $D$  la fonction dont la courbe représentative sur l'intervalle  $[-4 ; 4]$  est dessinée ci-contre.  
On propose ci-dessous les courbes représentatives de cinq fonctions :  $f, g, h, k, m$ .  
On cherche celles pour lesquelles  $D$  peut être la fonction dérivée.



$D$ peut être la dérivée de :				
a	$f$	V	F	Jnsp
b	$g$	V	F	Jnsp
c	$h$	V	F	Jnsp
d	$k$	V	F	Jnsp
e	$m$	V	F	Jnsp

Résultats Première S

<b>Réussite conjointe (TOUT)</b>
<b>29 %</b>

Item a	item b	Item c	Item d	Item e
60 %	59 %	52 %	59 %	53 %

QROC (Questions à Réponses  
Ouvertes et Courtes)<sup>21</sup>

En France, les évaluations n'utilisent le format QCM que de façon marginale. La plupart des questions utilisées sont de type QROC. A 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 page suivante). Il reste donc peu de place pour la rédaction, l'expression des

21 Short answer questions.

**Exemple de QROC - PISA 2003 - Question M702**

**OPINIONS FAVORABLES AU PRÉSIDENT**

**Question 1 : OPINIONS FAVORABLES AU PRÉSIDENT**

M702Q01 - 0 1 2 9

En Zedlande, des sondages d'opinion ont été menés pour déterminer la cote de popularité du président en vue de la prochaine élection. Quatre éditeurs de journaux ont chacun mené leur propre sondage d'opinion à l'échelle nationale. Les résultats des quatre sondages sont les suivants :

Journal 1 : 36,5 % (sondage effectué le 6 janvier sur un échantillon de 500 citoyens ayant le droit de vote, tirés au hasard) ;

Journal 2 : 41,0 % (sondage effectué le 20 janvier sur un échantillon de 500 citoyens ayant le droit de vote, tirés au hasard) ;

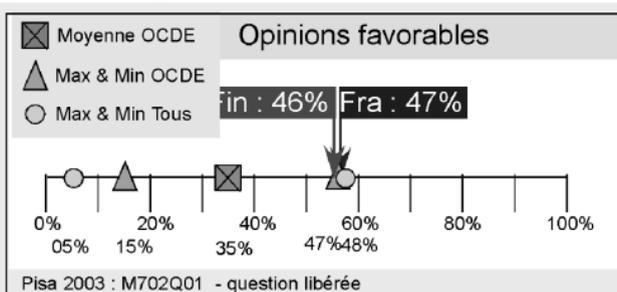
Journal 3 : 39,0 % (sondage effectué le 20 janvier sur un échantillon de 1 000 citoyens ayant le droit de vote, tirés au hasard) ;

Journal 4 : 44,5 % (sondage effectué le 20 janvier, sur 1 000 lecteurs qui ont appelé la rédaction pour voter).

Quel est le journal qui fournit probablement le résultat le plus fiable pour prédire le taux d'opinions favorables au président si les élections se tiennent le 25 janvier ?  
Donnez deux raisons pour justifier votre réponse.

**Résultats à 15 ans**

FRA = France  
FIN = Finlande



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.

### QRR (Questions à Réponses Rédigées)<sup>22</sup>

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.

#### Exemple de QRR - EVAPM Seconde 2003 - Question REC015

Deux nombres ont pour somme 456.

De combien augmente leur produit si l'on ajoute 7 à chacun d'eux ?

(il est précisé que toute réponse doit être complètement justifiée)

Résultats Seconde : réponse exacte 27%  
Noter que la question a donné lieu à 9 prises d'information.

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. A 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 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 éducatif (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 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. – Principaux enseignements des évaluations internationales et nationales

### 4.1. 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) peu-

<sup>22</sup> Long answer ou extended answer questions.

vent 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.

Il convient ici de noter 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, cache le fait que ces valeurs sont des positions sur une échelle normale de moyenne 500 et d'écart-type 100. Ainsi, la différence constatée est 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%<sup>23</sup>).

Il est important de garder à l'esprit que les valeurs de ces indicateurs ne sont en aucun cas des mesures d'un niveau absolu, mais tout au plus d'un niveau relatif. 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 sou-

vent 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.

Dans le cas de la comparaison France-Finlande, par exemple, les amplitudes des différences, selon les items étudiés, s'évaluent 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 savoirs<sup>24</sup>.

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

<sup>23</sup> Valeur approximative estimée – un calcul plus précis reste à faire.

<sup>24</sup> Ce point de vue fait écho aux échanges que nous avons eus avec des collègues finlandais lors du colloque franco-finlandais d'octobre 2005.

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 (voir 5.1).

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'évaluer<sup>25</sup>. 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 ont ainsi été laissés dans l'ombre.

<sup>25</sup> C'est l'« écrasement des objectifs sur l'évaluation » dénoncée en France par Guy Brousseau depuis les années 80.

- 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.

#### 4.2. 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 beaucoup<sup>26</sup>, 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'efficacité, 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 : A côté de ces cas extrêmes, il reste la plupart des pays d'Europe, l'Amé-

<sup>26</sup> 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.

rique 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.

- 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 pro-

babilités 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-cinquièmes 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.

#### 4.3. Les acquis des élèves en France

Dans un rapport récent de l'IGEN<sup>27</sup> 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,

<sup>27</sup> Inspection Générale de l'Éducation Nationale.

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 (DEP, mais aussi EVAPM), on sait en effet déjà beaucoup de choses sur les acquis des élèves en mathématiques. Il manque sans doute :

- 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 DEP, 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 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 DEP (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 direc-

tement 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 la 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 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.

A 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

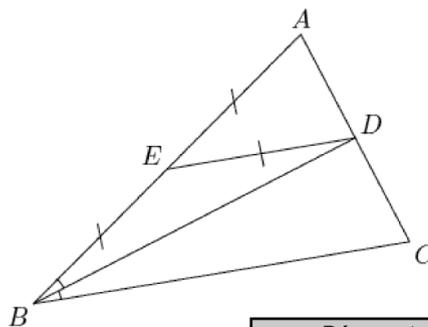
les observations que l'on peut faire sur les acquis des élèves (cf. EVAPM terminale 1999).

- 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. Vitgosky).

- 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

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

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



Démontrer que le triangle  $ABC$  est isocèle.

**Démonstration correcte**

Ensemble des élèves de seconde : 10%

Élèves orientés en première scientifique : 26%

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 sans doute de nature à relativiser la faiblesse des résultats, non à les nier<sup>28</sup>. 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 en bas de la page précédente, 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 prises 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 !

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

Nous présentons page suivante 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 dix dernières années en ce qui concerne les acquis des élèves en mathématiques nous renvoyons à un texte publié dans

<sup>28</sup> 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.

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

**Développer et réduire chacune des expressions suivantes :**

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

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

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

**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%.

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.

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

EVAPM SECONDE 2003 - SYNTHÈSE DES RÉSULTATS COMPARAISONS AVEC LES RÉSULTAT d'EVAPM SECONDE 1991					
		EVAPM2003	EVAPM91	Différence	Nombre d'items
	<b>ENSEMBLE</b>	<b>31%</b>	<b>37%</b>	<b>-6%</b>	<i>57</i>
	<b>Question abordée</b>	<b>75%</b>	<b>76%</b>	<b>-1%</b>	
<b>DOMAINE</b>	<b>Domaine numérique</b>	<b>30%</b>	<b>42%</b>	<b>-13%</b>	<i>12</i>
	<b>Domaine géométrique</b>	<b>30%</b>	<b>36%</b>	<b>-5%</b>	<i>42</i>
	<b>Statistiques</b>	<b>36%</b>	<b>27%</b>	<b>8%</b>	<i>3</i>
<b>NIVEAU TAXONOMIQUE</b>	<b>A : connaissance - reconnaissance</b>	<b>46%</b>	<b>50%</b>	<b>-4%</b>	<i>6</i>
	<b>B : Analyse</b>	<b>30%</b>	<b>41%</b>	<b>-11%</b>	<i>11</i>
	<b>C : Compréhension</b>	<b>34%</b>	<b>41%</b>	<b>-7%</b>	<i>28</i>
	<b>D : Synthèse et créativité</b>	<b>16%</b>	<b>17%</b>	<b>-1%</b>	<i>12</i>

Le tableau met en évidence les différences moyennes de réussite constatée, pour les 57 items communs à EVAPM 2003 et EVAPM 1991  
Voir la taxonomie complète dans les documents d'EVAPM (sur le site).  
A. Bodin - 20/04/04

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.

## 5. — Quels effets ?

### 5.1. *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) relevées 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 socié-

tés 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. Freudhenthal, 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 sont 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 par-

lons ne sont pas étrangères à cette évolution. Il ne faut toutefois pas trop penser à des effets directs mais plutôt à une sorte de percolation<sup>29</sup>, 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 d'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.

#### 5.2. En France

Comparé à ce qui se passe dans d'autres pays, peu d'effort 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 DEP (voir références). Là encore, il manque des synthèses et des productions en direction des enseignants et des formateurs d'enseignants<sup>30</sup>. Il manque aussi des analyses du contenu même des questions 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 DEP 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 dis-*

<sup>29</sup> Selon une expression de Guy Brousseau qui nous paraît particulièrement adaptée ici.

<sup>30</sup> Voir cependant les articles d'Yves Olivier et de Claire Dupé (cf. références)

*ciplines, 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 DEP, 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.

## 6. — 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, et en premier chef, ici, l'APMEP, 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 a trop longtemps été la règle en matière d'évaluation du système 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 l'institution devait avoir sa 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 notre système. 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.

De ce simple point de vue, il serait hautement souhaitable que notre pays reprenne sa participation aux études TIMSS et, en particulier, qu'il participe, enfin, aux études portant sur l'élémentaire.

### Références

- Adams, R.J., 2003, Response to « Cautions on OECD's Recent Educational survey (PISA), Oxford Review of Education, 29(3),
- Adams, R.J., 2003, Response to « Cautions on OECD's Recent Educational survey (PISA), Oxford Review of Education, 29(3),
- 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 : 2003, Comment classer les questions de mathématiques ? Communication au colloque international du Kangourou, Paris 7 novembre 2003. Article à paraître.

- 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. & 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 Academics Publishers, Dordrecht.
- 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'Éducation 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) : 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 conférences 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'Illettrisme 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
- 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](http://www.apmep.asso.fr/rubrique.php3?id_rubrique=114)

Niveaux de référence pour l'enseignement des mathématiques en Europe : [http://www-irem.univ-fcomte.fr/Presentation\\_ref\\_levels.HTM](http://www-irem.univ-fcomte.fr/Presentation_ref_levels.HTM) and <http://www.emis.de/projects/Ref/>

Antoine Bodin : [bodin.antoine@nerim.fr](mailto:bodin.antoine@nerim.fr)

**ANNEXE 1****Les études internationales  
concernant les mathématiques de 1960 à 2008.**

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	Mathematical literacy – Secondary topic			X		32	OCDE
2000	TIMSS VIDEO STUDY	Étude portant sur les pratiques d'enseignement	?	X			5	IEA
2003	PISA 2003	Mathematical literacy – Main topic			X		41	OCDE
2003	TIMSS 2003	Trends in International Mathematics and Science Study 2003	X	X			55	IEA
2006	PISA 2006	Math literacy - Secondary topic			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		IEA

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
<b>A</b>	<b>Connaissance et reconnaissance</b>	<b>A1</b>	des faits
		<b>A2</b>	du vocabulaire
		<b>A3</b>	des outils
		<b>A4</b>	des procédures
<b>B</b>	<b>Compréhension</b>	<b>B1</b>	des faits
		<b>B2</b>	du vocabulaire
		<b>B3</b>	des outils
		<b>B4</b>	des procédures
		<b>B5</b>	Des relations
		<b>B6</b>	Des situations
<b>C</b>	<b>Application</b>	<b>C1</b>	Dans des situations familières simples
		<b>C2</b>	Dans des situations familières moyennement complexes
		<b>C3</b>	Dans des situations familières complexes
<b>D</b>	<b>Creativité</b>	<b>D1</b>	Utiliser dans une situation nouvelle des outils et des procédures connus
		<b>D2</b>	Émission d'idées nouvelles
		<b>D3</b>	Création d'outils et de démarches personnelles
<b>E</b>	<b>Jugement</b>	<b>E1</b>	Production de jugements relatifs à des production externes
		<b>E2</b>	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	<b>Reproduction</b>	Les compétences classées dans ce groupe impliquent essentiellement la reproduction de connaissances déjà bien exercées	<b>Reproduction</b>
2	<b>Connexions</b>	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.	<b>Mathématisation simple</b>
3	<b>Réflexion</b>	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).	<b>Mathématisation complexe</b>