

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

Denis Meuret, IREDU

On m'a invité à m'exprimer ici à cause d'une étude que j'ai faite sur les raisons que l'on pouvait trouver au fait que les performances des élèves français étaient inférieures à celles de huit pays, anglo-saxons, asiatiques ou scandinaves de l'échantillon de PISA 2000 (Meuret, 2003), étude qui reposait évidemment sur la supposition que les « biais culturels », le thème qui nous réunit, ne sont pas, s'ils existent, de nature à invalider ce type de comparaisons.

Je voudrais donc ici présenter cette étude (3) et plaider la légitimité (4) des comparaisons internationales à partir de PISA 2000. Auparavant, cependant, je voudrais situer cette discussion dans son contexte politique – dire pourquoi je tiens PISA pour une bonne nouvelle politique (1) - et technique – en ce sens qu'il est légitime et souhaitable de critiquer la façon dont PISA a été conduite, à condition toutefois de tenir compte des avancées réalisées (2).

Une avancée politique

Je tiens pour une bonne nouvelle que quarante trois pays¹ mesurent sur une échelle commune les compétences de leurs élèves à la fin de la scolarité obligatoire, et donc se soient accordés sur la nature des compétences souhaitables comme sur le processus qui devait les mesurer. On aurait tort de tenir pour un simple processus technocratique le fait que 43 gouvernements reconnaissent ensemble pour légitime une certaine conception de la compréhension de l'écrit, des sciences ou des mathématiques, et ceci plus encore s'il s'agit d'un compromis entre des traditions pédagogiques différentes. Dans un monde affecté par les conflits crypto-religieux que nous connaissons, cela me semble important. De plus, dans un monde conduit par des forces économiques et des volontés de puissance mal régulées, l'existence de ce processus raisonné de production d'un accord, quelles que puissent être ses imperfections, me semble, en soi, positif.

Du point de vue politique, PISA présente à mon sens deux caractéristiques importantes.

La première est le fait que la construction des items a été régulée à la fois par des experts, par des groupes nationaux et un groupe de représentants des ministères de l'éducation. Les représentants des pays ont validé la canevas théorique qui sous-tend la nature des épreuves ; sur cette base, les pays ont envoyé des propositions d'épreuves, parmi lesquelles un choix a été fait par des spécialistes de l'évaluation, puis ces épreuves ont été envoyées aux pays pour examen et soumises au groupe des experts internationaux de chaque domaine (écrit, maths, sciences) ; L'épreuve ainsi construite a été pré-testée dans chaque pays, ce pré-test conduisant à une liste d'items amendée en fonction des observations de chaque pays, soumise à nouveau aux représentants des pays². A quoi il faut ajouter la transparence des procédures (elles sont

¹ Y compris les douze pays de « PISA+ ».

² Le fait que 60 % des items soient originaires de pays anglo-saxons ou que 88 items sur 141 aient été soumis en anglais doit s'apprécier en tenant compte de ce que les pays ont eu trois occasions au moins de juger si ces items

explicités dans le rapport technique, téléchargeable gratuitement sur PISA.org) et la rapidité avec laquelle les données ont été mises à la disposition des chercheurs et des citoyens du monde entier, deux éléments qui ont une signification politique et pas seulement technique.

La seconde est la nature des compétences mesurées par PISA. La différence entre l'approche des études de l'IEA et celle de PISA n'est pas, elle non plus, une simple question technique. L'approche IEA est celle de la recherche du plus grand commun dénominateur entre les programmes d'étude des différents pays participants, tandis que celle de PISA ne consiste pas, comme l'en accuse le professeur Prais (2002) à « tester les capacités mathématiques requises par la vie quotidienne » - sous entendu, des compétences subalternes- , mais, comme rétorque Adams (2002), le responsable du consortium PISA, « la connaissance de concepts fondamentaux et la capacité de les utiliser pour traiter de problèmes réels », ce qui n'est pas tout à fait la même chose. Bref, dans le cas de PISA, des pays s'accordent pour considérer qu'ils visent tous, chacun à leur manière, un certain socle de compétences à la fin de la scolarité obligatoire, et ils définissent ce socle comme devant donner les moyens intellectuels d'une vie dans le monde, soit une définition profondément laïque (Meuret, 1998) des objectifs de l'éducation, particulièrement bienvenue dans un pays comme la France où il est bien vu de l'élite de considérer, par exemple, que « les enseignants ne se doivent qu'à la logique interne de leur discipline » (Régis Debray, 1998).

Des progrès techniques

De l'avis général, PISA a élevé le standard technique des évaluations internationales, en partie, sans doute, à cause de l'enjeu politique de ses résultats. Ce progrès a touché de nombreux aspects, par exemple, la définition de la population (une tranche d'âge au lieu d'un niveau scolaire), la rigueur des procédures d'échantillonnage, le taux de couverture (moins de jeunes de très bas niveaux ont été exclus que dans les évaluations précédentes, en particulier par la mise au point d'épreuves ne comprenant pas les items les plus difficiles ; l'exclusion d'un élève trop faible obéissait à des procédures strictes), les procédures de correction (il était demandé aux correcteurs de corriger une question après l'autre et non un élève après l'autre pour éviter les effets de halo qui font qu'un correcteur se laisse influencer par la qualité des réponses précédentes de l'élève ; de nombreux contrôles ont été effectués sur la fidélité entre correcteurs, qui ont mis en évidence une fidélité très élevée des corrections aux questions ouvertes, de 91 % en moyenne et de 80% dans le pays où elle a été la moins bonne).

Je voudrais ici m'attarder un peu sur deux de ces progrès, qui touchent de plus près notre objet. Le premier est la définition de la lecture retenue. Lafontaine (2001) a comparé la notion de lecture utilisée dans les premières évaluations internationales (à partir de 1971) et dans PISA. Elle montre comment PISA, puis PIRLS³, ont été les premières à asseoir la définition de la lecture sur un véritable cadre conceptuel, mobilisant les avancées de la recherche,

convenaient à leur culture éducative ou autre. Ces précautions expliquent que, malgré ce déséquilibre initial, le classement de la France se modifie à peine lorsqu'on classe les pays avec l'ensemble des items, seulement avec les items que la France estimait être ceux qui convenaient particulièrement bien pour évaluer ses élèves, ou encore seulement avec les items que d'autres pays, y compris anglo-saxons, estimaient être ceux qui convenaient particulièrement bien pour évaluer leurs élèves. Ainsi, la France est mieux classée avec les items privilégiés par la Nouvelle Zélande qu'avec ceux qu'elle-même privilégie (OCDE, 2001). Ce résultat n'exclut pas absolument l'existence de biais culturels, mais exclut qu'ils soient assez importants pour que les experts des différents pays les perçoivent à l'examen du contenu des items.

³ Progress in International reading literacy study, évaluation de la lecture en 2001 auprès des élèves de quatrième année de primaire, réalisée par l'IEA dans 35 pays. Voir une présentation des résultats de la France dans MEN-DEP, NI 03.22.

considérables dans les années 80. Ce cadre définit la lecture comme un « processus interactif entre un texte, un lecteur et un contexte : le lecteur, pour atteindre ses objectifs, déploie des stratégies et fait appel à sa culture pour élaborer le sens et répondre au texte ». De là les trois échelles (trouver l'information, interpréter, réfléchir sur les textes) et le fait qu'elles ne soient pas hiérarchisées. D'où, aussi, l'existence de questions auxquelles il fallait apporter une réponse construite. Ce n'était pas simple souci d'authenticité, mais volonté de mesurer la façon dont le lecteur met en résonance le texte et ses connaissances antérieures. Il s'ensuit qu'il est plusieurs bonnes réponses possibles à ces questions. Cela aboutit dans l'ensemble à des exercices de qualité. Je ne sais s'ils ont été soumis à des enseignants français. Ils l'ont été à deux groupes d'enseignants suisses, plutôt hostiles à PISA a priori, mais qui ont néanmoins été impressionnés, d'après ce qu'on m'a rapporté, par la qualité des items présentés.

Des progrès ont été faits aussi pour la traduction des épreuves dans les différentes langues (Grisay, 2002). Alors que les évaluations précédentes ont utilisé essentiellement la rétro-traduction (la traduction en allemand depuis l'anglais est retraduite en anglais et comparée au texte original), PISA utilise la double traduction depuis deux langages sources, l'anglais et le français. Par exemple, en Allemagne, un traducteur traduit depuis l'anglais, un autre depuis le français et un conciliateur compare les deux traductions et la version finale se fait par concertation des trois personnes. Ceci parce que la rétro-traduction ne décèle pas les erreurs dues à une traduction trop littérale de la version source, et pour garantir l'équivalence des versions cibles (entre la traduction en allemand et celle en japonais par exemple) par équivalence avec deux versions sources et non pas une, ce qui améliore l'équivalence sémantique et « diminue l'impact des formes culturelles véhiculées par le langage de référence lorsqu'il n'y en a qu'un seul ». En fait, seuls six pays ont procédé de la façon, coûteuse, préconisée par le Consortium PISA, ce qui a permis d'observer que la solution préconisée était effectivement plus fiable que la double traduction depuis un seul langage (Grisay, 2002, p 52). Cependant, malgré le soin apporté à cet aspect, il reste que dans pratiquement tous les pays, l'une ou l'autre imperfection de la traduction a entraîné des taux de réussite un peu plus bas ou plus élevés à quelques items. Ces erreurs, qui généralement se compensent sur l'ensemble des items, ne sont pas de nature à affecter significativement les différences moyennes entre pays.

La réussite des élèves français à PISA

On peut s'intéresser à la position de la France sur certains aspects de l'épreuve ou sur certaines sous-échelles, avec ce que la DEP appelle une lecture « pédagogique » qui s'intéresse au sens et à la taille des écarts entre les résultats des élèves français et la moyenne de l'OCDE. On aperçoit alors, par exemple, que :

- « Si les jeunes français s'avèrent des lecteurs pertinents lorsqu'il s'agit de prélever de l'information ou de l'interpréter, ils semblent éprouver des difficultés quand il faut exprimer son point de vue ou exercer son esprit critique » (Robin, 2002).
- Les élèves français sont moins habitués à ce qu'on leur présente en classe des textes non littéraires et y réussissent, relativement, plus mal qu'aux exercices basés sur des textes littéraires (MEN-DEP, 2002).
- Les élèves français ont un des plus forts taux de non réponse des pays de l'OCDE pour les questions appelant des réponses construites (MEN-DEP, 2002).

On peut aussi, c'est ce que j'ai fait (Meuret, 2003), prendre le risque de comparaisons avec des pays particuliers, mais avec des ensemble plus importants d'items, et tenter de trouver des

explications à l'écart des compétences moyennes entre les jeunes français et ceux des pays qui font mieux qu'eux dans les trois dimensions testées (Compréhension de l'écrit, maths et sciences) : deux pays asiatiques (Corée et Japon), deux pays scandinaves (Finlande et Suède) et quatre pays anglo-saxons (Canada, Australie, Nouvelle Zélande et Royaume Uni).

J'ai d'abord observé que cette supériorité demeurait lorsqu'on tenait sous contrôle les différentes caractéristiques des familles reliées à la réussite scolaire (richesse, éducation des parents, biparentalité, parler à la maison la langue de l'école, etc.). J'ai ensuite comparé la France et ces pays du point de vue de quatre grandes dimensions dont les études empiriques ont montré qu'elles étaient liées à l'efficacité de l'enseignement⁴. *L'organisation du système éducatif* semble plus favorable en Nouvelle Zélande et au Royaume Uni qu'en France, mais ce n'est pas le cas pour les six autres pays⁵. Pour trois (discussions culturelles, travail à la maison, ressources éducatives à la maison) des quatre variables qui décrivent *l'aide et la stimulation par la famille*, la situation est significativement plus favorable en France que dans ces huit pays et pour la dernière (aide apportée par les parents) elle est plus favorable que dans trois pays, moins favorable que dans les cinq autres. *La discipline et l'engagement des élèves* sont meilleurs en Corée et au Japon qu'en France mais moins bons dans les autres pays. Ce ne sont donc pas ces variables qui peuvent expliquer la moindre performance des élèves français, à l'exception de l'aide apportée par les parents au travail de leurs enfants et, pour les pays asiatiques, de la discipline des élèves. Reste la quatrième dimension, *l'efficacité de l'enseignement en classe*, mesurée par trois construits de PISA (pression pour la réussite, qualité des relations entre élèves et enseignants, soutien apporté aux élèves) qui se trouvent être aussi associés, toutes choses égales par ailleurs, à la progression des élèves en français et en mathématiques dans les collèges français (Grisay, 1997). La France est, pour ces trois variables, en significativement moins bonne position que les autres pays, sauf le Japon et, pour une variable seulement, que la Corée⁶. Il semble donc bien que l'infériorité des performances des élèves français avec les élèves des meilleurs pays scandinaves et anglo-saxons puisse s'expliquer en partie par les attitudes des enseignants en France comparée à celle des enseignants de ces pays.

Cette étude ne montre pas que les enseignants français sont moins «exigeants et attentifs aux difficultés» que leurs collègues de ces pays, elle montre qu'ils sont moins nombreux à l'être (un ordre de grandeur est : 75 % d'un côté, 55 % de l'autre), ce qui n'est pas pareil. Par ailleurs, elle ne montre pas que là réside la seule explication de l'infériorité des compétences mesurées par PISA : l'existence du redoublement, un enseignement moins orienté vers

⁴ Je me suis référé pour cela aux études classiques de recherche sur l'efficacité de l'enseignement et non aux relations que l'on peut observer dans PISA même entre les variables de processus et les performances parce que, faute de mesure longitudinale des compétences des élèves dans PISA, ces relations sont affectées d'effets suppressifs qui se traduisent par des absences de corrélation ou par des corrélations très faibles (OCDE, 2003) entre des variables entre lesquelles la littérature établit souvent des relations plus fortes.

⁵ La description de l'organisation des systèmes éducatifs a été trouvée dans OCDE, Regards sur l'éducation, 1998. Les trois dimensions suivantes ont été mesurées via le questionnaire « élèves » de PISA.

⁶ Les différences mesurées ici peuvent être fortes : 90 % des élèves britanniques contre 45 % des français répondent que « le professeur demande aux élèves de travailler beaucoup », 74 % des élèves britanniques contre 49 % des français répondent que « le professeur s'investit beaucoup pour aider les élèves », 74 % des élèves britanniques contre 57 % des français répondent que « le professeur continue à expliquer jusqu'à ce que les élèves aient compris (PISA data base, accessible à tous sur PISA.org). Ces résultats sont d'ailleurs cohérents avec le fait que les enseignants britanniques sont plus souvent que leurs collègues français persuadés de leur efficacité pédagogique, observé à la faveur d'une comparaison récente de l'enseignement en Angleterre, au Luxembourg et en France en fin du primaire (RERPESE., 2002).

l'acquisition de certaines des compétences mesurées par PISA⁷, d'autres facteurs certainement, y compris la moindre aide parentale, figurent sans doute parmi les explications. Elle permet seulement d'identifier un facteur négatif, et donc un domaine dans lequel des progrès sont possibles.

Je voudrais ici prolonger cette étude en soulignant qu'un enseignement moins fréquemment dispensé par des professeurs attentifs aux difficultés, cela me semble rejoindre certaines caractéristiques pointées dans les interprétations de la DEP (moindre motivation, taux plus élevé de non réponses aux questions appelant des réponses longues⁸, difficulté à argumenter, à exercer son esprit critique) comme d'ailleurs certains résultats de PIRLS (le fait que les élèves français se sous-estiment davantage que les autres). En gros, les élèves français, confrontés moins souvent que dans ces pays à des enseignants soucieux d'accompagner les élèves dans leurs apprentissages, moins souvent à des enseignants soucieux de leur faire comprendre l'utilité intrinsèque des compétences, moins souvent à des exercices proches de la vie extérieure, et donc, peut-on penser, exposés peut être à des cours de plus haut niveau –puisque telle est notre réputation - mais n'en assimilant que les parties les plus faciles, les élèves français acquerraient un peu moins de compétences et des compétences plus scolaires que dans ces pays, se sentiraient plus souvent incapables de donner leur avis, et seraient moins enclins à « travailler pour rien ». Ce tableau fait sens, ce qui me conduit à considérer, par exemple, le pourcentage de non réponses aux items complexes ou la moindre propension à l'esprit critique, non comme des sources ou des résultats de biais, susceptibles par conséquent d'invalider les scores globaux que j'ai utilisés, mais comme des résultats intéressants et qui, plutôt, rejoignent les résultats de mon analyse.

4. Pourquoi ne pas avoir considéré que, si la France avait des résultats seulement moyens à PISA, c'était à cause des « biais culturels » ?

D'abord, je voudrais souligner que les « biais culturels » sont un des multiples biais (d'échantillonnage, de définition de la population mère⁹, ...) qui peuvent affecter une évaluation internationale et qu'on peut sans doute interroger leur fortune dans la réception de PISA dans notre pays. Mais, puisque nous sommes ici pour cela, évoquons les biais culturels.

Si j'ai décidé de les tenir pour négligeable pour mon propos, c'est d'abord, mais je n'y reviens pas, à cause du soin mis dans PISA à éliminer les items qui auraient pu en être porteurs. C'est

⁷ La même étude pointe que les enseignants du primaire anglais privilégient, davantage que les français, l'enseignement de l'argumentation (RERPESE., 2002). On peut penser que cela laisse des traces à 15 ans et contribue à expliquer que les jeunes français réussissent moins bien, en particulier lorsqu'il faut faire preuve d'esprit critique.

⁸ Sur l'ensemble des items, le taux de non réponses (items manquants) en France est voisin du taux de non réponses dans la moyenne des pays de PISA. Il est supérieur au taux de non-réponse que l'on observe dans les huit pays qui nous dépassent. Cependant, les contrôles menés sur les phénomènes d'omission montrent une corrélation (négative) très élevée entre le taux d'omission ou d'items non atteints et le score que l'élève obtiendrait si on ne tenait compte que des items pour lesquels l'élève a donné une réponse. Cela est donc un aspect plutôt qu'une cause de notre faiblesse par rapport à eux. Toutefois, on compte un petit peu plus d'items manquants en France que dans les pays dont les performances ne sont pas significativement différentes du notre (Autriche, Belgique, Islande, Norvège, Etats-Unis, Danemark, Espagne) : 2,44 en France contre 2,01 pour la moyenne de ces pays pour la session 1, 2,99 contre 2,79 pour la session 2 (OCDE, 2002, p 157) La « peur de se tromper » - que l'on aurait pourtant bien vu dans un coin de ce tableau - est peut être un peu plus forte qu'ailleurs en France, mais pas de beaucoup.

⁹ Des biais de cette sorte ont affecté à la hausse les résultats la France dans Reading Literacy en 1991.

aussi parce qu'on a pu montrer que ces biais étaient minimes. Les mesures du fonctionnement différentiel des items – i.e. l'écart entre la réussite moyenne d'un pays à un item et celle qu'on attendrait compte tenu de sa réussite moyenne dans le domaine-, permettent de dire que les différences de difficulté d'un même item d'un pays à l'autre sont faibles et surtout que si certains items favorisent légèrement un pays, d'autres ont l'effet inverse de sorte que le score global- celui qui m'importait- n'est pas atteint.

A quoi il me semble qu'il faut ajouter que, si un item est mieux réussi dans un pays qu'il ne le devrait compte tenu du score général du pays, cela peut venir d'un biais culturel de l'item (une plus grande familiarité des élèves du pays avec le contenu du texte, par exemple), mais aussi de ce qu'il mesure des compétences particulièrement travaillées dans ce pays, auquel cas on n'a pas affaire à un biais mais à un résultat intéressant pour l'analyse du curriculum réel du pays. Il en va de même des supports utilisés : si, en sciences, on utilise des textes scolaires plus que des textes de presse pour présenter des problèmes scientifiques (MEN-DEP, 2002, p 159), on peut aussi bien reprocher aux items de PISA de ne pas épouser les habitudes pédagogiques françaises que s'interroger sur les conséquences de cette habitude en termes de capacité de traitement des données scientifiques. Dût-on estimer, in fine, ce coût supportable au regard des avantages de la pratique française, l'interrogation n'aurait pas été inutile. De même de l'habitude de poser dans des exercices de maths des questions de difficulté croissante, proposée comme explication, en effet possible, à certaines non-réponses (idem, p 146) : dussions nous – mais j'en doute- conclure à l'opportunité de cette pratique, permettre de s'interroger dessus est sans doute un bon exemple de ce à quoi peuvent servir, aussi, les comparaisons internationales. Il est parfois difficile de tracer une frontière nette entre un biais psychométrique et un résultat intéressant ; chacun la placera à un endroit qui dépend de la rigueur avec laquelle il estime que la situation française doit servir de norme¹⁰.

Reste la question de la longueur des textes (Grisay, 2002). Le chapitre sur la traduction du rapport technique de PISA montre en effet que les textes et les mots sont un peu plus longs en français qu'en anglais. Il semble donc que leur langue handicape légèrement les jeunes francophones¹¹, ce qui d'ailleurs n'a pas empêché les jeunes québécois d'atteindre des scores enviables¹² et que ce puisse donc être une (petite) partie de l'explication de notre score, surtout en compréhension de l'écrit. Il est donc légitime de penser que les caractéristiques mêmes de leur langue sont un petit handicap pour les français vis-à-vis de certains des pays qui ont de meilleures compétences qu'eux à PISA 2000 (pays anglophones, Finlande).

¹⁰ L'auteur de ces lignes était placé dans une situation proche lorsqu'il s'efforçait avec d'autres, dans les années 1980, de promouvoir l'utilisation d'une dotation globale dans les établissements scolaires : chaque établissement expliquait que ses spécificités exigeait qu'il fut traité différemment des autres, et critiquait la grossièreté et l'insensibilité de l'indicateur (H/E), et ce n'est qu'après avoir pensé longtemps qu'en effet cet indicateur était d'une grossièreté intolérable, que j'ai pensé que sa vertu résidait dans sa grossièreté même, qui permettait de faire apparaître le coût des fameuses « spécificités ». Dans l'intervalle, nous avons tenté de le « pondérer » de diverses manières, sentant confusément tout de même qu'à pousser trop loin les pondérations, nous risquions le sens même de ce que nous faisons.

¹¹ Le coefficient d'allongement des textes est de 15% par rapport à l'anglais. De même l'allemand, l'italien, l'espagnol, certaines langues scandinaves ont des coefficients d'allongement significatifs. Une des moins concises est le Coréen (30%). Que la difficulté de compréhension d'un texte ne dépende pas que de sa longueur explique que les effets de coefficients assez importants puissent être faibles et que la Corée puisse avoir une des performances les plus fortes en lecture... et que la corrélation entre les performances dans les trois domaines de PISA soit élevée, alors que ce problème de la langue les affecte différemment.

¹² 536 en compréhension de l'écrit, soit 10 points de moins seulement que la Finlande et 31 points de plus que la France, soit 0,3 écart-type : le score du québécois médian est supérieur au score de 62 % des français. Statistiques Canada, 2001, *A la hauteur : la performance des jeunes du Canada en lecture, en mathématiques et en sciences, étude PISA de l'OCDE*, 93p.

Toutefois, il faut s'interroger avec Grisay (2002) sur la réponse qu'appelle cette situation : faut-il donner à lire des textes un peu raccourcis aux élèves des langues moins concises ou s'y refuser, en considérant que ce serait précisément biaiser que de comparer les élèves des différentes langues sur des textes dont la charge sémantique serait différente ?

La solution idéale serait sans doute d'insérer la concision de la langue parmi les différents facteurs d'efficacité de l'enseignement de la compréhension de l'écrit. Il faut souligner d'ailleurs qu'une étude internationale qui comporterait un pré-test et un post-test et mesurerait l'efficacité d'un système éducatif une année donnée par l'estimation d'une « valeur ajoutée » éviterait cette difficulté, comme bien d'autres qui se posent lorsqu'il s'agit de mettre en regard les caractéristiques d'un système éducatif et les performances de ses élèves.

Denis Meuret
Université de Bourgogne
IREDU

Bibliographie

Adams, R.J., 2003, *Response to « Cautions on OECD's Recent Educational survey (PISA)*, Oxford Review of Education, 29(3), 377-389.

D'Haultefeuille, X., Murat, F. et Rocher, T., 2000, *La mesure des compétences, les logiques contradictoires des évaluations internationales*, Intervention aux VII^{ème} journées de méthodologie statistique, Paris, 37p.

Grisay, 2002, *Translation and cultural appropriateness of the test and survey material*, PISA technical report, 42-54.

Grisay, A. 1997 *Evolution des acquis cognitifs et socio-affectifs des élèves au cours des années de collège*, MEN-Direction de l'Evaluation et de la Prospective, Dossiers Education et formations, n°88.

Lafontaine, *Quoi de neuf sur la littéracie ? Regard sur trente ans d'évaluation de la lecture*, Cahiers du SPE, Université de Liège, n°7-8, 71-95.

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.

Meuret, D., 1998, *Intérêt, justice, laïcité*, 8p, in (Le) Télémaque, Presses universitaires de Caen, N° 14.

Prais, S. J., 2003, *Cautions on OECD'S Recent Educational Survey (PISA)* Oxford Review of Education, 29(2), 139-163

Robin, I., 2002, *L'enquête PISA sur les compétences en lecture des élèves de 15 ans : trois biais culturels en question*, Ville, Ecole, Intégration, Enjeux, n° 129, CNDP.

Men-DEP, 2001, *Les élèves de 15 ans, premiers résultats d'une évaluation internationale des acquis des élèves*, NI 01-52, 6p.

MEN-DEP, 2002, *Les élèves de CMI, premiers résultats d'une évaluation internationale en lecture (PIRLS)*, NI 01-52, 6p.

MEN-DEP, 2002, *Les compétences des élèves français à l'épreuve d'une évaluation internationale, Premiers résultats de l'enquête PISA 2000*, Dossier Education et formation 137, Paris, 182p.

OCDE, 2001, *Choice of assessment tasks and the relative standing of countries in PISA 2000, a first analysis*. Document pour la treizième réunion du conseil des pays participants, Paris, 10 p.

OCDE, 2002, *Pisa Technical Report*, Paris, 262p.

OCDE, 2003, *La lecture, moteur de changement, Performances et engagement d'un pays à l'autre, résultats de PISA 2000*, Paris, 279p.

Réseau européen des responsables des politiques d'évaluation des systèmes éducatifs, (RERPESE) 2002, *Enseignement et Compétences en lecture à la fin de la scolarité primaire dans trois pays européens*, MEN-DEP, 105p.