

# **24ème Congrès International du CIRIEC**

*Napoli Italia 30 settembre - 2 ottobre 2002*

## **PISA: L'INDICATEUR PENCHÉ? DIFFICULTÉS D'INTERPRÉTATION DU PROGRAMME OCDE POUR L'ÉVALUATION DES ACQUIS DES ÉLÈVES DE 15 ANS**

**Stefania Gabriele**

ISAE – Istituto di Studi e Analisi Economica

**Stefano Zolea**

Regione Campania

(Italie)

*Communication libre*

Le "Programme international pour le suivi des acquis des élèves", PISA dans l'acronyme, est une étude très ample, finalisée principalement à évaluer, dans une perspective comparative entre les pays, les capacités de compréhension de l'écrit et la familiarité avec les notions mathématiques et scientifiques des jeunes âgés de 15 ans qui fréquentent l'école. L'enquête vise aussi à explorer les possibles raisons des différences remarquées entre les jeunes.

Le programme a utilisé des ressources considérables et il a de grandes ambitions, en particulier il entend représenter un point de repère important pour ceux qui seront appelés à intervenir ou à dessiner les politiques pour l'école. Le premier chapitre du rapport "Connaissances et compétences: des atouts pour la vie. Premiers résultats du programme international de l'OCDE pour le suivi des acquis des élèves (PISA) 2000" affirme que la rigueur avec laquelle le travail a été mené, le concours d'experts, les contrôles de la qualité des traductions, de l'échantillonnage et du rassemblement des données assureraient " un niveau élevé de validité et de fiabilité".

La communication présente vise à éclairer quelques problèmes sous-jacents à l'essai de comparer les habiletés d'individus qui appartiennent à des communautés de différents langages et milieux culturels. Les doutes qui émergent suggèrent que les indicateurs calculés par le PISA doivent être considérés avec quelque prudence, surtout quand on veut raisonner sur les politiques à entreprendre.

## **LES CLASSEMENTS DES PAYS**

Le projet PISA a été réalisé pour la première fois en 2000 et il a intéressé 32 pays, dont 28 Membres de l'OCDE (ensuite il a été étendu à d'autres nations, mais les résultats ne sont pas encore disponibles). Plus de 250.000 étudiants ont été interviewés (parmi 17 millions d'élèves âgés de 15 ans). Les échantillons nationaux couvrent entre 95% et 97% de la population scolaire âgée de 15 ans<sup>1</sup>, si l'on fait exception de trois pays (Brésil, Luxembourg et Pologne).

L'objectif déclaré du programme de l'OCDE pour l'évaluation internationale des étudiants est de vérifier jusqu'à quel point les jeunes sont à même d'exploiter les connaissances et les habiletés acquises pour relever les défis de la vie réelle, et non dans quelle mesure ils ont assimilé les programmes d'enseignement. Le projet a pour but de considérer la compréhension des notions, la maîtrise des processus et la capacité de faire face à des situations différentes. Cette conception ample du savoir et des habiletés est synthétisée dans le mot "littératie" ("literacy" en anglais), qui se réfère à une notion plus vaste que ce qu'on entend par le mot "alphabétisation".

---

<sup>1</sup> Pour l'exactitude, on a interviewé des jeunes d'un âge compris entre 15 ans et 3 mois et 16 ans et 2 mois.

L'évaluation concerne trois domaines de compétences, la lecture avant tout, un aspect auquel une attention particulière a été attribuée dans cette première enquête, et puis les mathématiques et les sciences. Ces deux dernières disciplines seront analysées plus attentivement à l'occasion des prochaines études, qui seront réalisées respectivement en 2003 et en 2006.

La figure 1 montre la liste des pays examinés rangés selon la moyenne des points obtenus par les étudiants qui ont participé aux épreuves. La moyenne de tous les pays OCDE est de 500 points. La figure 2 présente un classement construit sur la base des pourcentages d'étudiants qui atteignent au moins le troisième niveau de compétence (ce qui correspond à un score entre 481 et 552 points), sur les cinq niveaux déterminés. Les figures 3 et 4 indiquent les classements des valeurs moyennes respectivement pour les mathématiques et les sciences. Dans toutes ces figures les barres qui représentent les pays sont colorées en blanc quand la langue parlée est l'anglais; en gris-bleu pour les autres langues indo-européennes; en noir pour les langues non indo-européennes. Avant d'expliquer pourquoi, on a mis en évidence les caractéristiques linguistiques, on va rappeler en bref les bases du projet PISA.

Figure 1 - Compréhension de l'écrit (performances moyennes)

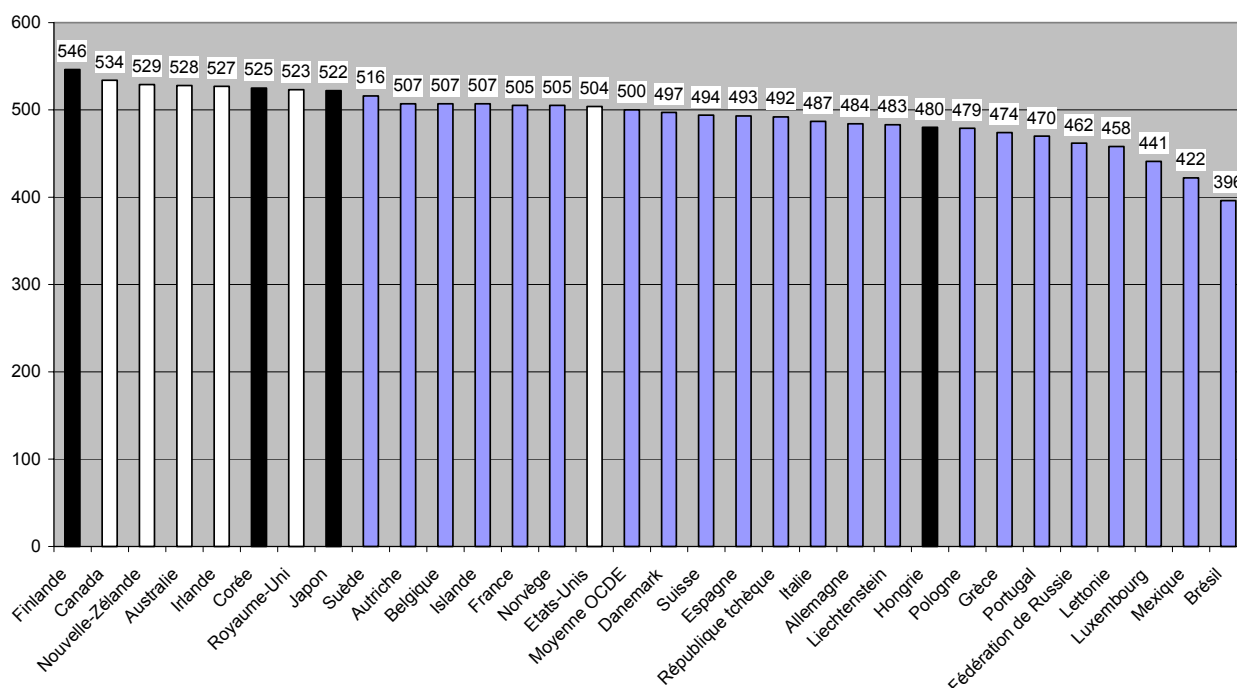


Figure 2 - Pourcentage d'élèves ayant atteint le niveau 3, 4 ou 5

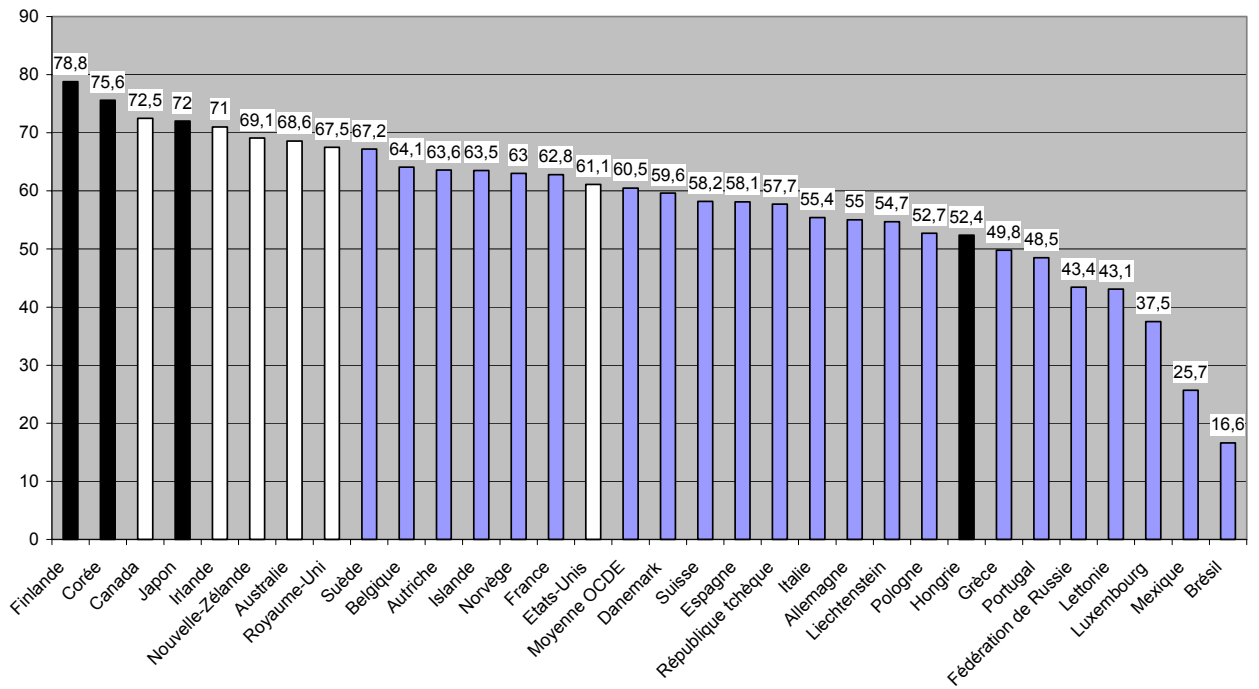


Figure 3 - Culture mathématique (performances moyennes)

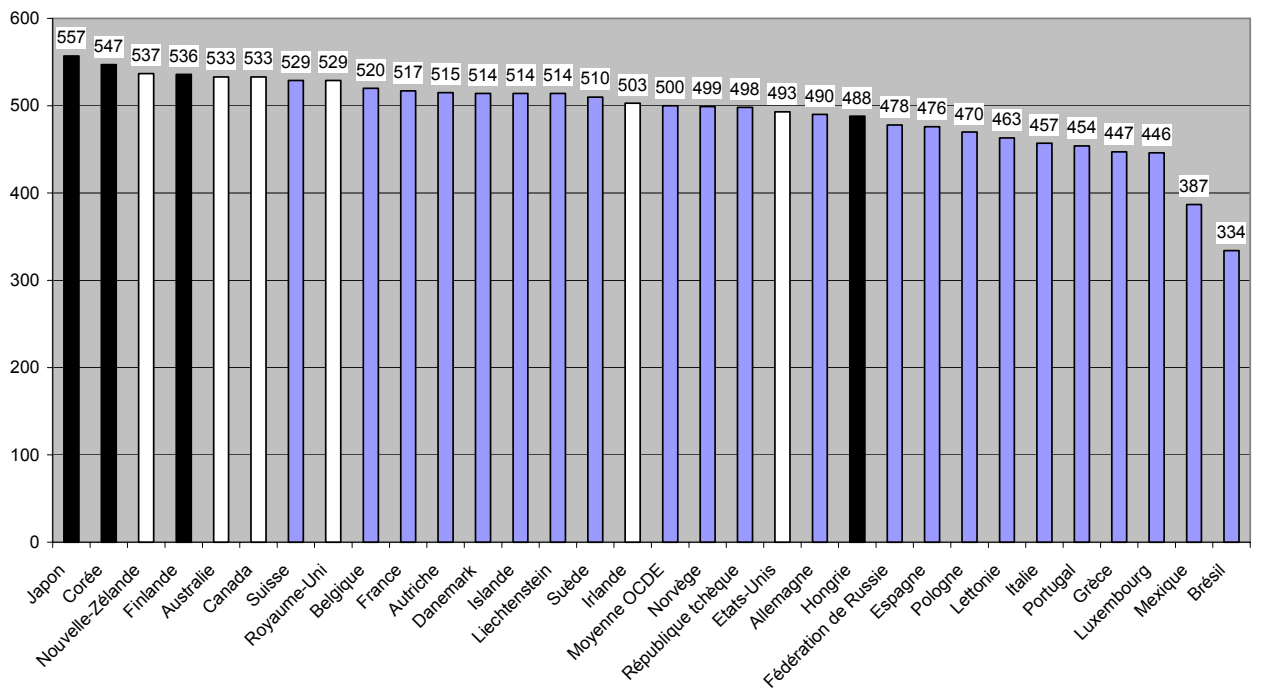
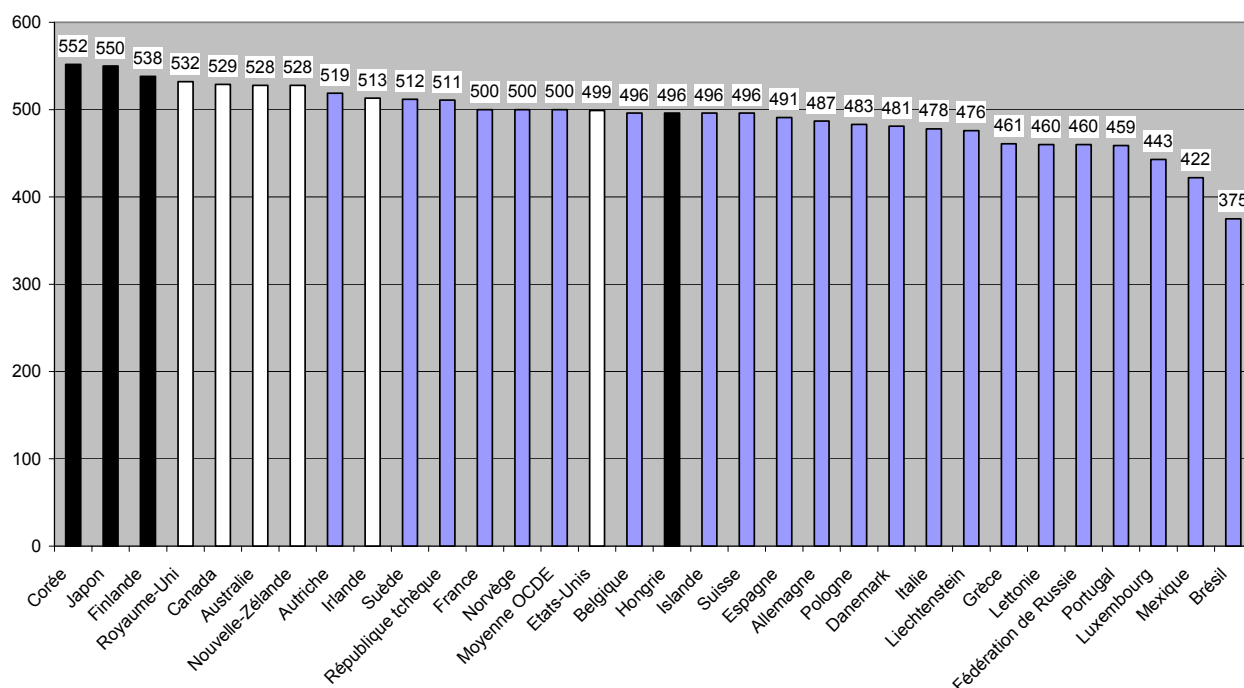


Figure 4 -Culture scientifique (performances moyennes)



## LES CARACTÉRISTIQUES DES TESTS

L'examen des capacités des jeunes a été réalisé en leur proposant quelques tests écrits et deux heures ont été accordées pour répondre aux questions. Dans l'ensemble, 140 épreuves ont été préparées, pour un total de sept heures de travail, mais des combinaisons différentes de tests ont été proposées à différents élèves. Les items ont été révisés par des spécialistes nationaux des disciplines concernées et par des experts en matière d'évaluation. Les pays intéressés ont confirmé l'applicabilité culturelle, la pertinence par rapport aux programmes scolaires et la justesse du degré de difficulté. Les épreuves ont été effectuées dans les écoles.

Les tests se présentent en quelques cas dans la forme de "textes continus" (parmi lesquels on peut distinguer les textes narratifs, descriptifs et argumentatifs), mais quelquefois ils comprennent des diagrammes, des graphiques, des formulaires, des listes; ils demandent donc la souplesse nécessaire pour pouvoir tirer des informations à partir de textes avec des caractéristiques différentes. On ne vise pas à vérifier les compétences les plus élémentaires liées à la capacité de lire, qui devraient avoir été assimilées longtemps avant l'âge de 15 ans, mais plutôt dans quelle mesure le jeune soit en mesure de trouver les informations, de comprendre un texte, de l'interpréter, de réfléchir sur ce qu'on a lu, d'évaluer et argumenter sa propre opinion. En outre, on demande de déterminer le contexte auquel le texte se rapporte, en distinguant les

documents de différentes typologies et finalités (une lettre privée ou un document officiel, *et cetera*). Quant aux mathématiques, on entend mesurer la capacité d'analyser, raisonner et communiquer des idées en sachant poser, formuler et résoudre des problèmes mathématiques en des contextes différents. Aux fins de l'évaluation, trois grandes dimensions ont été prises en considération: en ce qui concerne le contenu des mathématiques, les concepts de variations et relations et d'espace et formes ont été privilégiés; quant aux processus, on se rapporte à la reproduction de calculs simples ou de définitions de base, à la détermination des procédés pour résoudre des problèmes faciles, à la réflexion, c'est-à-dire au raisonnement mathématique, à la généralisation et à la compréhension en profondeur; enfin, pour la dernière dimension considérée, qui concerne les situations où les mathématiques sont utilisées, on a choisi des problèmes semblables à ceux qu'on rencontre dans la vie réelle. L'évaluation de la culture scientifique part d'une définition de cette dernière comme capacité d'utiliser les connaissances scientifiques, d'identifier les problèmes et de tirer des conclusions fondées sur l'évidence, pour comprendre le monde naturel et les changements apportés par l'être humain et adopter des décisions à ce propos. Trois dimensions ont été considérées pour les sciences aussi: les concepts scientifiques, se référant à la physique, à la chimie, à la biologie, aux sciences naturelles et spatiales, avec un égard particulier aux thèmes de la biodiversité, des forces et du mouvement et du changement physiologique; le processus scientifique, c'est-à-dire la capacité de constater, interpréter et utiliser les faits, et en particulier l'identification du problème, la découverte d'éléments probants, l'établissement de conclusions, la communication des mêmes conclusions et la manifestation de la compréhension des notions scientifiques; les situations scientifiques, c'est-à-dire les domaines d'application, qui sont constitués par les expériences de la vie réelle, et parmi lesquelles on a choisi des exemples tirés des sciences de la vie et de la santé, des sciences de la terre et de l'environnement, des sciences dans la technologie.

Outre les tests, un questionnaire sur les caractéristiques des élèves et de leurs établissements scolaires a été fourni. Dans plusieurs pays des questions sur l'attitude vis-à-vis de l'étude et les méthodes d'étude adoptées ainsi que sur la familiarité avec l'informatique ont été proposées. Un questionnaire a été soumis aussi aux chefs d'établissement pour approfondir quelques aspects liés au fonctionnement des écoles. Les informations recueillies ont été élaborées dans le but d'analyser les relations entre les résultats des tests et les autres variables tirées des questionnaires, comme par exemple le *background* familial, le genre, la méthode d'étude préférée.

## LES FINS “POLITIQUES” DU PISA

Comme on l’a déjà évoqué, le programme PISA a été réalisé avec l’ambition de fournir un important instrument politique: le rapport de présentation des premiers résultats, cité plus haut, déclare que l’objectif est de consentir aux gouvernements de tirer des enseignements politiques des évidences émergées de l’étude. Il est bien vrai que les différents rapports découlés du projet expriment çà et là une certaine circonspection, et invitent à la prudence sous plusieurs aspects. On reconnaît par exemple que le score d’un pays supérieur à celui d’un autre pays n’implique pas automatiquement que le premier jouisse d’un système d’enseignement meilleur. Au contraire, on affirme que l’influence de toutes les expériences d’apprentissage vécues à l’intérieur et à l’extérieur de l’école, avant et après l’âge scolaire, produit les résultats différents qui émergent du PISA. Néanmoins on croit que les analyses sorties de l’enquête peuvent “orienter les décisions politiques et l’affectation des ressources”, car elles permettent de déterminer les facteurs qui contribuent au développement des compétences. Les informations rassemblées pourraient être exploitées par les enseignants et les décideurs pour identifier les points faibles et les points forts de leur système, grâce aussi à l’analyse des situations individuelles des étudiants (genre, caractéristiques familiales, *et cetera*), de leur attitude vis-à-vis de l’étude (intérêt pour les sujets, inclination à la lecture, *et cetera*), des caractéristiques de l’école. On prétend en outre que le PISA, en visant aux compétences utiles dans la vie de tous les jours, ne fournit pas des indications bornées à l’efficacité intérieure de chaque système scolaire, mais il assure plutôt des résultats comparables entre les pays et entre les cultures. Le choix de condenser les résultats des tests proposés aux élèves par des classements de pays (relatés à la lecture, aux mathématiques et aux sciences) semble néanmoins fourvoyant, et il peut, quand même involontairement, suggérer une interprétation des classements comme étant connectés directement à l’efficacité des systèmes scolaires.

Il faut certainement partager l’opinion que l’extension du matériel mis à disposition, le grand nombre de rapports, de données et d’informations, l’ampleur des analyses proposées rendent le PISA un instrument précieux pour quiconque veut s’occuper des thèmes qu’il traite <sup>2</sup>. C’est toutefois l’ambition politique qui sollicite une vérification tous azimuts pour constater si les fondements du projet soient absolument à partager. Quelques observateurs sont en train de réfléchir sur les méthodologies statistiques utilisées soit par le PISA, soit par des études précédentes de comparaison internationale sur le “produit” de l’instruction et suggèrent d’adopter de la circonspection dans l’interprétation des résultats (Micklewright J., 2002, *Measuring Functional Literacy and Numeracy in the European Union*, *Politica Economica*, a. XVIII, n.1, avril). Mais l’on peut mettre en évidence une difficulté aussi d’un autre

---

<sup>2</sup> On a de la perplexité, au contraire, sur le manque d’informations à propos des résultats de l’enquête menée en Italie: non seulement le rapport-pays n’est pas encore disponible, mais le Ministère de l’Instruction ne fournit même pas, sur demande, d’exemples des tests en italien, bien que beaucoup d’exemples soient disponibles en d’autres langues.

point de vue, ou bien on peut se demander si une évaluation comparée des performances des élèves puisse être rigoureusement effectuée, et si l'on peut la considérer fiable jusqu'à penser de régler sur elle les interventions de réforme que les gouvernements voudraient entreprendre.

## SUR LA TRADUCTION

Les théories linguistiques les plus récentes affirment la neutralité essentielle des langues par rapport à l'expression de la pensée. Contrairement aux positions dominant jusqu'à il y a quelques décennies, la plupart des linguistes soutiennent que « **presque tout** ce que nous pouvons dire dans une langue donnée peut être traduit en n'importe quelle autre langue, en gardant inchangée la pensée exprimée dans la langue donnée . » (Jackendoff, R., 1990, *Linguaggio e natura umana*, Bologna, p.250).

Bref, on remarque un vaste accord entre les spécialistes « à l'égard de la conviction que les différences de pensée susceptibles d'être conditionnées par la différence de structure grammaticale sont **assez** superficielles » (ibid.).

Il est d'ailleurs bien connu que parfois un excellent traducteur de secteur (par exemple un traducteur scientifique ou un traducteur littéraire) rencontre dans quelques traductions de textes de remarquables limites d'efficacité dans le travail de transposition. Cette gêne peut aussi bien se vérifier dans la même personne. On sait que cela est arrivé à Nabokof, brillant écrivain de langue anglaise qui " s'entêtait à écrire toujours à l'aide de dictionnaires et de grammaires, il expliquait à propos de lui-même –Je pense comme un génie, j'écris comme un grand écrivain et je parle comme un enfant-" (Pinker S., 1997, *L'istinto del linguaggio*, Milano, p.283).

Il faut en outre tenir compte des contradictions des résultats scientifiques obtenus jusqu'à présent sur la base de la révélation de petites différences d'efficacité d'apprentissage et de raisonnement résultant de la diversité de structure entre les alphabets phonétiques occidentaux et ceux de l'Extrême Orient; et comment ne pas citer aussi la *vexata quaestio* de la possibilité de l'influence des différences sur la base des aires corticales cérébrales du cerveau entre hommes et femmes.

Même dans les plus tenaces défenseurs chomskiens de l'unicité et de l'universalité de la grammaire mentale des doutes quand même émergent sur l'absolue possibilité de superposer les différentes langues (dans la littérature consultée on trouve souvent des expressions du type : des différences **assez** superficielles entre les langues émergent ; **à peu près** toute forme de pensée peut être transposée d'une langue à l'autre, *et cetera*). Quiconque connaît la théorie de l'erreur sait qu'un petit « bias » de fond dans les bases d'un parcours scientifique peut multiplier l'erreur même à des niveaux incontrôlables dans l'interprétation des résultats finals du parcours. Etant donné que



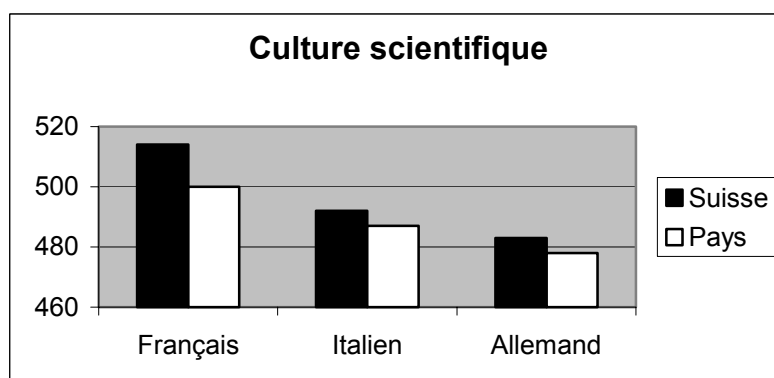
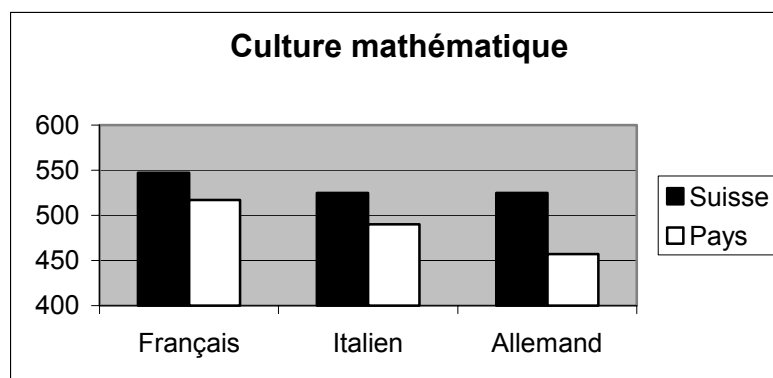
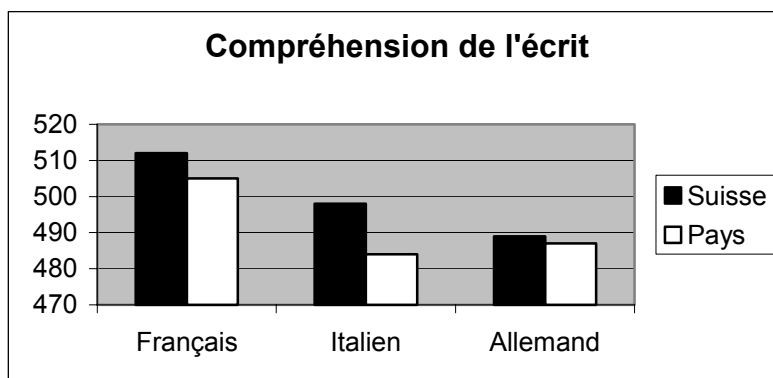
dans le cas du PISA il s'agit d'une quantification de résultats obtenus d'après des réponses à des tests traduits en plusieurs langues, on ne peut pas exclure une influence « linguistique » dans les réponses aux tests.

Il est incontestable que la méthode choisie par l'OCDE pour la dernière version du PISA vise à atténuer les éventuels biais de traduction puisqu'on a souvent réalisé une pratique de double traduction de confirmation pour d'autres langues à partir de l'anglais et du français, pratique ultérieurement soutenue par d'autres interventions d'experts des différentes langues pour minimiser les erreurs qui se sont produites dans des enquêtes analogues par le passé. Toutefois non toutes les pratiques semblent avoir eu la même rigueur et il reste l'impression de tests conçus dans la tête de savants de langue mère anglaise (ou quand même parlant couramment l'anglais).

De toute façon par le passé des critiques sévères ont déjà été conduites par Blum et d'autres (Blum A., Goldstein H., Guerin-Pace F., An analysis of international comparison of adult literacy, dans *Assessment in education : principles, policy and practice*, JULY 2001, volume 8, n° 2, pp. 225-246) à propos de la signification des tests d'enquête entre divers pays et entre diverses langues sur la *literacy* d'après l'International Adult Literacy Survey, IALS (OCDE, 1997, *Literacy skills for the knowledge society*). De cette enquête il résultait une basse évaluation de la capacité de « lire » les journaux ou d'écrire une lettre de la part des Français en comparaison des Britanniques, des Américains et des Hollandais. Blum, en étudiant soigneusement les versions anglaise et française, a constaté des sources de biais dans les traductions et une meilleure précision dans les tests en langue anglaise. Dans l'analyse de la structuration des tests on a remarqué des évidences au sujet d'une possible source de « bruit » dans ce qu'on appelle le cultural bias entre le français et l'anglais. De plus, on remarquait des corrélations statistiques selon les langues entre les réponses de nations qui avaient des communautés de langue mère différente (les Flamands de la Belgique s'assimilaient aux Hollandais, les francophones de la Suisse aux Français, les germanophones suisses aux Allemands de l'Allemagne) : ce phénomène existe aussi pour le PISA, ainsi que la figure 5 le montre, en suggérant une non indifférence du bias linguistique dans ce type d'enquêtes, quelle qu'en soit la motivation.

Un dernier détail qui mérite d'être relevé: l'excellente position de performance dans les tests du PISA pour ce qui concerne Corée et Japon, deux nations de langues très différents des langues indo-européennes (on a aussi les remarquables performances des Finnois qui parlent une autre langue agglutinante), révèle en tout cas la possibilité d'une influence du bias linguistique.

**Figure 5 - Classement du PISA pour le français, l'italien, l'allemand parlés dans les pays d'origine et en Suisse**



## DES PERPLEXITÉS À PROPOS DE QUELQUES TESTS DE CULTURE SCIENTIFIQUE (SCIENTIFIC LITERACY)

Comme on l'a déjà vu, l'évaluation de la culture scientifique entreprise par le PISA n'entend pas se borner seulement à la stricte capacité d'utiliser les connaissances scientifiques, mais encore s'étendre à la capacité d'approfondir quelques questions et de les actualiser dans la réalité quotidienne.

Il faut remarquer que dans une batterie de tests auxquels on doit répondre en un temps limité on a par conséquent exclu d'autres caractéristiques de la pensée et de l'action scientifique contemporaine qui sont considérées très importantes, comme le scepticisme, la nature probabiliste d'une grande partie de la connaissance scientifique et la nature « accumulative » de la pratique et de l'expérimentation scientifique, parfois contradictoire (M.Ageno, 1980, *Lezioni di Biofisica*, Bologna, p.17).

Il est bien connu aussi que les théories néodarwiniennes les plus répandues et les plus en vogue dans la communauté scientifique peuvent créer et créent des problèmes dans l'enseignement de quelques parties de la biologie dans les écoles de quelques pays à cause de la forte présence d'instances et d'objections créationnistes et lamarckiennes. Il est donc bien connu que dans les tests de culture scientifique il existe souvent une remarquable autocensure dans la préparation de questionnaires se rapportant au néodarwinisme : il en résulte ainsi un *vulnus* sur la vérification de l'un des « paradigmes » les plus importants de la pensée scientifique contemporaine.

On remarque enfin dans le PISA un exemple d'un test fournissant des évaluations à une réponse de nature politique et non de nature scientifique. C'est le cas de la réponse au test concernant la diminution de la couche d'ozone de la stratosphère. La question prévoit une réponse sur la base de la validité du Principe de précaution, qui est manifestement un principe absolument politique (Cameron J. (1995), *Addressing uncertainty. Law, policy and the development of the precautionary principle*, in Dente, B. (éd.), *Environmental Policy in Search of New Instruments*, Dordrecht, pp.95-142 ; Godard O. (1997), *Le principe de précaution*, Paris, p.63).

Il faut partant prendre acte que quelques perplexités émergent au sujet de ce qu'on essaie d'évaluer au moyen du PISA pour ce qui concerne la culture scientifique.