

PISA : une enquête bancale ?

Julien GRENET

Des enquêtes PISA sur les acquis des élèves, les médias ne retiennent d'ordinaire que les classements à sensation, peu favorables à la France ; or c'est là l'aspect le plus fragile de ces enquêtes, voire même absurde du point de vue statistique. Tandis que leurs qualités méthodologiques permettent de mettre en évidence les forces et les faiblesses spécifiques de l'école française – pourvu qu'on apprenne à les lire.

La publication au mois de décembre dernier des résultats de l'enquête 2006 du Programme international pour le suivi des acquis des élèves (PISA), réalisée tous les trois ans par l'Organisation de coopération et de développement économiques (OCDE) pour mesurer les acquis des élèves de 15 ans dans 57 pays, a confirmé le classement médiocre de la France tant en compréhension de l'écrit qu'en culture mathématique et scientifique. Ces mauvais résultats rejoignent les conclusions des éditions précédentes de l'enquête, rendues publiques en 2000 et en 2003.

Bien que les performances très moyennes des élèves français aient suscité cette année un certain nombre d'analyses dans la presse¹, les réactions sont toujours restées relativement discrètes en France si on les compare avec le « choc » créé par PISA au Royaume-Uni ou en Allemagne. Chez nos voisins britanniques, la publication en 2000 des résultats de la première vague d'enquête, dans laquelle le Royaume-Uni se classait parmi les dix premiers pays, fut accueillie avec un soulagement teinté d'incrédulité, le *Times* allant jusqu'à barrer la Une de son édition du 5 décembre 2000 d'un facétieux « *Are we not such dunces after all ?* » (« Et si nous n'étions pas des cancre après tout ? »). Au même moment, les très mauvaises performances des élèves allemands provoquèrent un véritable séisme médiatique et politique

¹ Cf. « La France paralysée devant ses mauvais résultats scolaires » in *Le Monde* du 4 décembre 2007.

outré-Rhin : le constat sévère dressé par la *Frankfurter Allgemeine Zeitung* daté du 4 décembre 2000 (« [Miserable Noten für deutsche Schüler](#) », qu'on peut traduire par « Des notes lamentables pour les élèves allemands ») fut suivi par une kyrielle d'articles et de rapports qui occupèrent le devant de l'actualité allemande pendant plusieurs semaines.

Si l'enquête PISA ne s'est jamais vraiment retrouvée au cœur du débat éducatif français, c'est que les comparaisons internationales en matière d'éducation ont longtemps été considérées avec scepticisme, car jugées biaisées et difficiles à interpréter : on se souvient que la France avait décidé en 1995 de suspendre sa participation à l'enquête internationale sur la « littératie » des adultes ([IALS](#)), après avoir pris connaissance de son très mauvais classement, en soulignant les réelles limites méthodologiques de l'enquête. Il semblerait néanmoins que cette méfiance traditionnelle laisse aujourd'hui la place à une attitude plus constructive, le ministre de l'Éducation nationale n'hésitant plus à faire référence aux études internationales lorsqu'il s'agit de tracer des pistes de réforme pour le système éducatif français.

Que faut-il donc penser de ces enquêtes ? Sont-elles fiables ? Que révèlent-elles réellement ? Pour anticiper sur les conclusions de l'analyse qui suit, on peut souligner un paradoxe : ce que les médias en retiennent le plus souvent est de loin la dimension la plus fragile de ces études (en particulier les classements), et ce à quoi ils s'intéressent le moins en constitue au contraire la dimension la plus riche et sans doute la plus robuste. Les qualités méthodologiques de l'enquête PISA permettent en effet de mettre en évidence les forces et les faiblesses spécifiques de l'école française. On a donc de bonnes raisons de prendre ces enquêtes au sérieux, à condition de les utiliser correctement et d'apprendre à les lire...

I/ PISA : un modèle d'enquête

Un demi-siècle d'enquêtes internationales de suivi des acquis des élèves

L'enquête PISA n'est pas une création *ex nihilo*. Les enquêtes internationales sur le suivi des acquis des élèves existent en effet depuis près d'un demi-siècle : la première étude de ce type fut pilotée par l'[Institut international de l'éducation](#) en 1961, une institution fondée en 1952 par l'UNESCO et dont l'objectif initial était d'évaluer les systèmes d'enseignement en mesurant les acquis d'un échantillon de près de 12 000 élèves de 13 ans issus de 12 pays différents. En 1961, il fut décidé de confier à l'[IEA](#) (*International Association for the*

Evaluation of Educational Achievement), dont le siège fut fixé à Hambourg, le monopole de la réalisation d'enquêtes internationales de grande envergure sur les acquis des élèves.

Critiquées sur le plan scientifique et sans grand impact sur les politiques éducatives des pays participants, les grandes enquêtes de l'IEA², ont été progressivement éclipsées par le programme PISA, qui fut lancé en 1997 par les ministères de l'Éducation des pays de l'OCDE pour obtenir une source régulière de données sur les résultats de l'enseignement capable d'alimenter l'ensemble d'indicateurs internationaux que l'OCDE avait commencé à produire en 1992 ([cf. Bottani & Vrignaud, 2005](#)). L'opération débuta en 1998 et la première enquête fut réalisée en 2000.

L'originalité de PISA

Bien qu'elle s'inscrive dans le droit fil des enquêtes internationales sur les acquis des élèves menées par l'IEA, l'enquête PISA présente trois grandes nouveautés par rapport à celles qui l'ont précédée.

La principale originalité de l'enquête PISA tient à la nature des acquis des élèves qu'elle cherche à mesurer : contrairement aux enquêtes antérieures, il ne s'agit pas d'évaluer l'acquisition de connaissances fixées par les programmes scolaires, mais les compétences ou aptitudes jugées nécessaires pour mener une vie d'adulte autonome. L'évaluation des acquis des élèves ne s'effectue donc plus en référence à un hypothétique dénominateur commun des programmes scolaires nationaux, mais par rapport à la notion de « littératie », qui désigne un ensemble de compétences génériques identifiées comme nécessaires pour pouvoir comprendre le monde, assurer son développement personnel et participer pleinement à la vie collective. Pour les besoins de l'enquête, l'OCDE a isolé trois grands domaines de compétence : la « compréhension de l'écrit », la « culture mathématique » et la « culture scientifique »³.

² Troisième enquête internationale sur les mathématiques et les sciences ([TIMSS](#)) et Progrès dans la compréhension de la lecture à neuf ans ([PIRLS](#)).

³ La compréhension de l'écrit est définie comme étant « *La capacité de comprendre, d'utiliser et d'analyser des textes écrits afin de pouvoir réaliser ses objectifs, développer ses connaissances et son potentiel et jouer un rôle actif dans la société* ». La culture mathématique mesure quant à elle « *l'aptitude d'un individu à identifier et à comprendre les divers rôles joués par les mathématiques dans le monde, à porter des jugements fondés à leur propos, et à s'engager, en fonction des exigences de sa vie présente et future en tant que citoyen constructif, impliqué et réfléchi* ». Enfin, la culture scientifique est définie comme « *la capacité à utiliser des connaissances scientifiques, à identifier les questions et à tirer des conclusions fondées sur des faits, en vue de comprendre les changements qui y sont apportés par l'activité humaine l'essentiel étant d'être en mesure d'adopter une pensée scientifique à l'égard des faits rencontrés* ».

La seconde nouveauté de l'enquête PISA réside dans la définition de la population testée : contrairement aux études précédentes, l'évaluation ne porte pas sur l'ensemble des élèves d'un niveau scolaire donné (les élèves de 5^e, par exemple), mais sur l'ensemble des élèves de 15 ans indépendamment de leur classe (l'enquête 2006 porte par exemple sur les individus nés en 1990). Le choix de ce groupe d'âge présente deux avantages : d'une part, il permet de mesurer les compétences des élèves arrivant en fin de scolarité obligatoire, l'obligation scolaire dans la plupart des pays de l'OCDE étant fixée à 15 ou 16 ans ; d'autre part, il autorise une comparaison des acquis des élèves plus fiable qu'une comparaison par niveau d'enseignement, dont la composition par âge est très sensible à la pratique du redoublement. En France, par exemple, les classes de 3^e accueillent en majorité des élèves de 15 ans, mais aussi des élèves de 16 et 17 ans, ainsi que quelques élèves en avance âgés de 14 ans. En revanche, dans les pays qui pratiquent le passage automatique dans la classe supérieure, presque tous les élèves de 15 ans sont dans des classes de même niveau d'enseignement.

Une dernière spécificité de PISA tient à sa périodicité : réalisée tous les trois ans, l'enquête a pour objectif d'assurer un suivi régulier des performances des élèves de 15 ans de l'ensemble des pays de l'OCDE afin d'effectuer des comparaisons dans le temps. Le programme d'évaluation est établi de telle manière que chaque enquête met l'accent sur l'une des trois compétences évoquées précédemment, en y consacrant les deux tiers des questions environ : la compréhension de l'écrit fut le domaine principal testé en 2000 ; en 2003, ce fut la culture mathématique, en 2006, la culture scientifique et en 2009, ce sera à nouveau au tour de la compréhension de l'écrit.

Le nombre de pays participants n'a cessé de croître depuis le lancement du programme PISA : de 32 pays membre de l'OCDE en 2000, on est passé à 42 pays (dont 12 pays partenaires) en 2003 pour arriver à 57 pays (dont 27 pays partenaires) en 2006, représentant près de 90 % de l'économie mondiale. Lors de la dernière vague d'enquête, ce sont au total près de 400 000 élèves, représentatifs des 20 millions de jeunes de 15 ans scolarisés dans les pays participants, qui ont été sélectionnés de manière aléatoire pour participer au cycle PISA 2006.

Une méthodologie rigoureuse

Reconnaissons-le d'emblée : l'enquête PISA est d'une très grande qualité et offre peu de prises à une critique purement méthodologique. Toutes les étapes de sa construction ont en effet été conduites avec un soin extrême, depuis l'élaboration des questionnaires jusqu'au calcul des scores individuels, en passant par la constitution des échantillons nationaux.

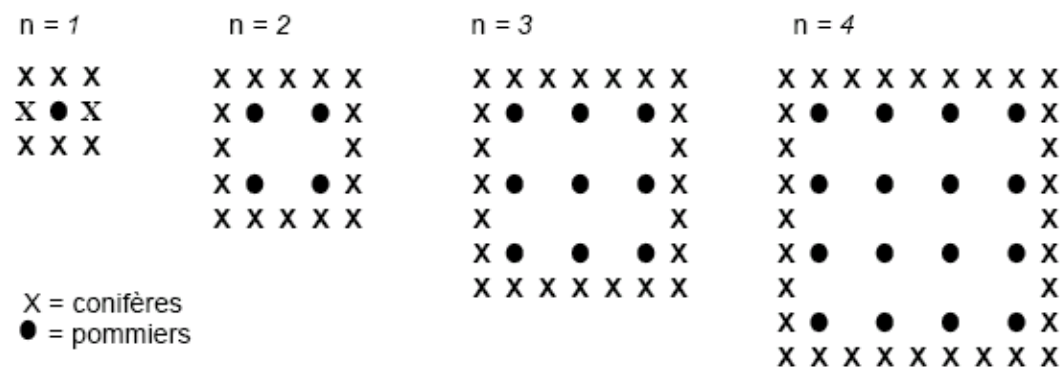
L'organisation de chaque enquête est subdivisée en trois phases d'un an : la première année est consacrée à la définition des objectifs et à la mise au point des instruments de mesure ; la deuxième année donne lieu à une expérimentation de ces instruments et des procédures de collecte ; enfin, la troisième année fait place à l'évaluation proprement dite. Un consortium composé de divers instituts de recherche, dirigé par le Conseil australien pour la recherche en éducation ([ACER](#)) est chargé de l'exécution du projet en coopération avec des représentants de chaque pays, sous l'égide de l'OCDE.

L'étape-clé de la phase préparatoire de l'enquête est la construction des instruments d'évaluation : ceux-ci consistent pour l'essentiel en « items » mis au point par des groupes d'experts internationaux et par le consortium, à partir d'un cadre de référence établi au niveau international. Ces items, au nombre de 160 environ, sont constitués à égalité de questions à choix multiples et questions à réponse construite (« fermée » ou « ouverte »). Ils sont rédigés en anglais, puis traduits en français, deuxième langue officielle de l'OCDE. Ces deux versions sources sont ensuite utilisées conjointement par les pays non anglophones et non francophones pour effectuer la traduction dans leur propre langue. La qualité des traductions est assurée par une retraduction dans la langue source des versions élaborée par les différents pays. L'année précédant l'enquête, les différents items sont testés au cours de la phase expérimentale de l'enquête. La figure 1 reproduit l'un des items proposés dans la partie « culture mathématique » de l'enquête PISA 2000 (d'autres exemples sont consultables sur [cette page](#) du Portail de l'Évaluation du ministère de l'Éducation nationale) :

POMMIERS

Un fermier plante des pommiers en carré. Afin de protéger ces arbres contre le vent, il plante des conifères tout autour du verger.

Vous pouvez voir ci-dessous un schéma présentant cette situation, avec la disposition des pommiers et des conifères pour un nombre (n) de rangées de pommiers :



Question 1 : POMMIERS

M136Q01- 01 02 11 12 21 99

Complétez le tableau :

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

Le recueil des données est soumis à un protocole particulièrement rigoureux. En France, la conduite des opérations est confiée à la Direction de l'évaluation, de la prospective et de la performance (DEPP) du ministère de l'Education nationale. Le tirage de l'échantillon d'élèves de 15 ans se déroule en deux temps : un certain nombre d'établissements scolaires sont tirés au hasard puis, au sein de chacun de ces établissements, une sélection aléatoire d'élèves est effectuée. À titre d'exemple, l'enquête 2006 comporte un échantillon représentatif de 4 700 élèves français nés en 1990 et issus de 187 établissements distincts, soit une trentaine d'élèves par établissement. Les normes PISA prévoient divers motifs d'exclusion d'élèves : certains établissements peuvent être exclus parce qu'ils sont situés dans des régions reculées, parce qu'ils sont difficilement accessibles ou parce qu'ils ne se prêtent pas à l'administration des épreuves pour des raisons de taille, d'organisation ou de mise en

œuvre. Toutefois, afin de garantir la comparabilité des échantillons, l'OCDE a imposé que le taux global d'exclusion de la population cible reste inférieur à 5 % dans chaque pays. En France, la population visée par l'enquête PISA 2006 couvre ainsi 95 % de la génération des jeunes de 15 ans en cours de scolarité.

Afin d'éviter tout biais lié au déroulement des épreuves, ces dernières sont étroitement encadrées : dans chaque établissement scolaire, un « coordinateur scolaire » est responsable de la mise en œuvre de l'opération et un « administrateur de test » est chargé de l'exécution des épreuves. En France, ces deux personnes sont choisies au sein de l'équipe administrative de l'établissement. Les élèves passent deux heures d'évaluation dans les trois domaines de compétence évoqués plus haut : afin de couvrir un nombre important d'items sans que les élèves aient à répondre à tous, les questions sont réparties dans une dizaine de cahiers différents avec des « blocs » communs permettant de relier les résultats obtenus. Les élèves doivent également remplir un questionnaire « contextuel » portant sur leur famille, leur rapport à la lecture, leurs opinions, etc.

À l'issue des épreuves, les copies sont envoyées au centre national PISA pour correction et codage. Afin de garantir la précision et la cohérence du codage d'un pays à l'autre, des consignes de correction détaillées sont envoyées aux correcteurs et un sous-échantillon de réponses d'élèves est prélevé dans chaque pays pour être corrigé de manière centralisée par quatre correcteurs différents.

Le score des élèves dans chacun des trois domaines évalués est calculé à partir de modèles statistiques dits de « réponse à l'item » reposant sur l'hypothèse que la réussite à un item ne dépend que du niveau de compétence de l'élève, qui est inobservé mais que l'on peut estimer statistiquement. Ces scores sont utilisés pour construire une « échelle » continue de performance qui permet de situer le niveau de chaque élève dans chaque domaine évalué. Afin de faciliter l'interprétation des scores attribués aux élèves, les échelles ont été conçues de manière à ce que le score moyen des pays de l'OCDE soit égal à 500 points et que deux tiers environ des élèves des pays de l'OCDE obtiennent entre 400 et 600 points (ce qui correspond à un écart-type de 100). Cette échelle est elle-même subdivisée en plusieurs niveaux de compétences. L'échelle de compétence en compréhension de l'écrit en comporte cinq, allant du niveau 1 et inférieur (élèves incapables de mettre couramment en œuvre les connaissances et les compétences les plus élémentaires que l'enquête PISA cherche à mesurer) jusqu'au niveau 5 (élèves capables de mener à bien des tâches complexes de lecture).

Dès le lancement du programme PISA, l'OCDE a tenu à assurer la diffusion la plus large possible aux résultats des vagues successives d'enquête, tant en direction des chercheurs que des gouvernements nationaux, afin que leurs enseignements puissent influencer directement l'orientation des réformes éducatives menées dans les pays participants. De ce point de vue, l'enquête PISA constitue un modèle de transparence : les bases de données sont intégralement [téléchargeables sur le site de l'OCDE](#), ainsi qu'une documentation exhaustive sur la méthodologie de l'enquête et les principales analyses auxquelles elle a donné lieu.

II/ Des résultats à interpréter avec prudence

Si nous avons choisi de présenter les grandes lignes de l'approche méthodologique adoptée par PISA, c'est que ce détour est indispensable pour bien comprendre ce que mesure cette enquête et les précautions à prendre lorsqu'on en vient à interpréter ses résultats.

Que mesure PISA ?

Comme il a été indiqué plus haut, le programme PISA ne cherche pas à évaluer la maîtrise par les élèves de leur programme scolaire, mais à mesurer des connaissances et des aptitudes que les individus ont avantage à posséder pour leur vie future. Le rapport préliminaire de l'OCDE consacré aux résultats de l'enquête réalisée en 2006 définit ainsi l'objectif assigné à l'enquête : « *L'enquête PISA cherche non seulement à évaluer la capacité des élèves à reproduire ce qu'ils ont appris, mais aussi à déterminer dans quelle mesure les élèves sont capables de se livrer à des extrapolations à partir de ce qu'ils ont appris et d'utiliser leurs connaissances dans des situations familières ou originales et dans des contextes en rapport ou non avec l'école* » ([OCDE, 2006](#)). Les compétences évaluées par PISA reposent donc davantage sur la compréhension globale de concepts que sur l'accumulation de connaissances spécifiques.

La compréhension de l'écrit est par exemple définie comme « *la capacité à comprendre, à utiliser et à analyser des textes écrits afin de pouvoir réaliser ses objectifs, développer ses connaissances et son potentiel et jouer un rôle actif dans la société* ». Cette définition va donc bien au-delà de la simple maîtrise de la lecture : il s'agit d'évaluer la capacité des élèves à comprendre et à interpréter des textes de types divers (textes descriptifs,

narratifs, graphiques, tableaux, formulaires, publicités, etc.) et à leur donner une signification en les reliant aux contextes dans lesquels ils sont proposés. PISA évalue séparément trois compétences de lecture qui sont jugées indispensables pour participer à la société contemporaine : trouver des informations dans des matériaux variés (« s'informer »), interpréter ce qu'on lit (« interpréter »), réfléchir à ce qu'on vient de lire et y réagir (« réagir »).

Cette définition normative des compétences que sont supposés maîtriser les élèves de 15 ans ne va pas de soi, dans la mesure où elle ne prend en effet nullement en compte les objectifs fixés par les systèmes éducatifs nationaux. On peut même dire qu'elle s'inscrit dans une certaine mesure *contre* eux. Certains observateurs ([Prais, 2003](#)) ont fait remarquer qu'une telle approche était susceptible de pénaliser mécaniquement les approches pédagogiques fondées sur la construction par « briques » d'un savoir dont le terme est fixé à la fin de l'enseignement secondaire (lorsque les élèves atteignent l'âge de 17 ou 18 ans) plutôt que la constitution d'un « socle commun de connaissances » supposé atteint la fin de la scolarité obligatoire. On peut en effet considérer que ce type de pédagogie laisse davantage de côté la mobilisation concrète des connaissances pour résoudre des problèmes de la vie quotidienne qu'une pédagogie plus « pragmatique », fondée sur l'idée qu'un élève arrivant en fin de scolarité obligatoire doit disposer d'un bagage intellectuel suffisamment complet et autonome pour affronter la vie d'adulte. Or rien ne permet de conclure à la supériorité de l'une sur l'autre dans le long terme, bien que la seconde soit sans doute plus propice que la première à la réussite des élèves de 15 ans aux épreuves de PISA.

Faut-il pour autant considérer que l'enquête PISA est viciée par un « biais culturel » qui donnerait un avantage structurel aux systèmes éducatifs des pays anglo-saxons ? À l'appui de cette thèse, certains auteurs ([Robin, 2002](#)) invoquent le fait que la langue source des items est l'anglais et que certains formats (les QCM notamment) ou supports (les articles de presse, par exemple) sont plus ou moins pratiqués dans le cadre scolaire selon le pays d'origine des élèves.

Un examen attentif des résultats des vagues successives de l'enquête PISA indique que si un biais culturel il y a, il n'est susceptible d'expliquer qu'une très faible part des écarts constatés entre les pays participants. Il apparaît en effet que d'une part, les élèves qui obtiennent les meilleurs résultats appartiennent à des aires culturelles très différentes (Finlande, Canada, Corée du Sud) alors que les élèves scolarisés dans des pays culturellement

proches peuvent manifester des performances très inégales (les États-Unis et le Canada, par exemple) ; que d'autre part, un soin extrême a été apporté à la sélection définitive des questions soumises aux élèves : les items détectés comme « biaisés » car induisant des écarts trop importants entre la réussite moyenne d'un pays à un item et celle qu'on attendrait compte tenu de sa réussite moyenne dans le domaine ont été éliminés à l'issue de la phase expérimentale de l'enquête. Enfin, il faut noter que les responsables de l'enquête PISA ont procédé à une ré-estimation des performances des élèves à en se limitant aux seuls items considérés par les pays eux-mêmes comme particulièrement bien adaptés à leurs élèves, sans que cela modifie significativement la position des différents pays. La seule forme de « biais culturel » réellement identifiée provient de la longueur relative des items d'une langue à l'autre, certaines langues étant moins concises que d'autres : le coefficient d'allongement des textes est par exemple de 15 % pour le français par rapport à l'anglais. Les rapports techniques produits par l'OCDE indiquent néanmoins que les biais imputables à la traduction des items ne sont susceptibles d'expliquer qu'une part infime des écarts de performance⁴.

Les écueils des classements internationaux

Le caractère vraisemblablement marginal des écarts de performance liés à l'existence d'un hypothétique « biais culturel » ne doit pas toutefois nous dispenser d'une grande prudence lorsqu'on en vient à interpréter les classements internationaux produits par l'enquête PISA.

Il faut d'abord garder à l'esprit que malgré le soin mis à la conception et à la conduite de cette enquête, un certain nombre de biais statistiques sont susceptibles de gonfler ou de minorer artificiellement les résultats de certains pays. La principale source de biais provient de la représentativité de la population d'élèves couverte par les échantillons nationaux. Par définition, la population nationale cible inclut uniquement les jeunes de 15 ans qui sont scolarisés dans un établissement, à l'exclusion des élèves qui ont quitté le système éducatif. Si cette définition ne pose guère de problèmes dans la plupart des pays membres de l'OCDE, où la quasi-totalité des élèves est scolarisée à l'âge de 15 ans, c'est loin d'être le cas pour tous, et *a fortiori* pour les pays partenaires : alors que l'échantillon utilisé représente près de 97 % des individus âgés de 15 ans, 95 % au Royaume-Uni ou 90 % en France, ce taux n'est que de 55 % au Brésil, 54 % au Mexique (qui fait partie des pays membres de l'OCDE) ou encore

⁴ L'allongement des questions dû à la traduction en français ne semble pas avoir particulièrement pénalisé les jeunes québécois, qui ont atteint dans le domaine de la compréhension de l'écrit des scores supérieurs de 30 points aux élèves français.

53 % en Indonésie. Dans ces conditions, confronter les performances de pays dont les échantillons ne couvrent pas la même fraction d'une génération est un exercice hautement périlleux. Les comparaisons devraient donc se limiter aux pays membres de l'OCDE disposant d'échantillons réellement représentatifs.

Une source mineure de biais statistique provient de l'inégale couverture des élèves scolarisés à l'âge de 15 ans. Comme indiqué plus haut, l'OCDE a autorisé les pays participants à exclure de leurs échantillons certains établissements (en général parce qu'ils sont situés dans des régions reculées) et certains élèves (parce qu'ils sont atteints d'un handicap intellectuel ou qu'ils ne maîtrisent pas suffisamment la langue du test) dans la limite de 5% de la population cible. Si ce taux d'exclusion a été respecté dans l'immense majorité des cas, les normes d'échantillonnage ont été violées dans un certain nombre de pays : au Royaume-Uni, par exemple, les taux de réponse étaient si bas lors des cycles PISA 2000 et 2003 que les données de ce pays ont été jugées invalides du point de vue de la comparabilité internationale et ont, dès lors, été exclues de la plupart des comparaisons.

Au-delà des biais statistiques qui viennent d'être évoqués, la comparaison des résultats d'un pays à l'autre comporte une marge d'erreur importante et inhérente à toute enquête fondée sur l'utilisation d'échantillons représentatifs plutôt qu'exhaustifs. Dans l'enquête PISA, les scores moyens d'un pays sont en général affectés d'une marge d'erreur de plus ou moins cinq points, ce qui signifie que l'intervalle de confiance correspondant à un score national de 523 points en lecture est en réalité compris entre 518 et 528 points, soit dix points environ. L'existence de telles marges d'incertitude rend impossible l'établissement d'un classement univoque des pays : le rang d'un pays donné est en effet susceptible de varier beaucoup selon la valeur retenue dans l'intervalle de confiance. Tout au plus les résultats de l'enquête permettent-ils d'établir des « plages de classement » plus ou moins larges selon les pays. Dans l'enquête PISA 2006 par exemple, la France se situe entre la 18^e et la 28^e place (sur 56) sur l'échelle de compréhension de l'écrit. Un grand nombre d'écarts de performance entre pays ne sont donc pas statistiquement significatifs : dans le domaine de la compréhension de l'écrit en 2006, la performance française n'est pas significativement différente de la performance de 13 pays sur 55 ! L'habitude médiatique consistant à assigner une position déterminée à un pays dans le classement PISA constitue donc une absurdité du point de vue statistique.

La réification du classement PISA est d'autant plus regrettable qu'elle conduit bien souvent à exagérer les écarts de performance constatés au sein des pays de l'OCDE.

Il existe plusieurs manières d'appréhender l'ampleur de ces écarts. Une première approche consiste à comparer la variabilité des performances moyennes au sein d'un groupe de pays par rapport à la variabilité totale dans l'ensemble des pays participants. On constate alors que trois groupes de pays aux performances relativement homogènes peuvent être distingués : un premier ensemble de 24 États regroupe les pays anglo-saxons, les pays d'Europe du Nord (France, Allemagne et Benelux inclus), le Japon et des « dragons » asiatiques⁵ ; un deuxième ensemble de 20 pays comprend les États d'Europe de l'Est, la Russie et les pays d'Europe du Sud⁶ ; enfin, un troisième ensemble est constitué de 13 pays asiatiques, d'Afrique du Nord et d'Amérique latine⁷. La variabilité des performances à l'intérieur de chacun de ces trois groupes est étonnamment faible, puisqu'elle est égale à 5 % de la variabilité totale pour les pays du premier groupe et à 10 % de la variabilité totale pour les pays des deuxième et troisième groupes. L'écart de performance qui sépare deux pays comme la France et l'Australie (32 points dans le domaine de la culture scientifique) est donc particulièrement faible lorsqu'on le compare à l'écart séparant l'Australie et l'Argentine (136 points) ou à celui qui sépare l'Australie et le Kirghizistan (205 points). Malgré les apparences, les performances des pays du premier groupe sont donc généralement situées dans un mouchoir de poche, cette faible variabilité expliquant que les performances d'un pays comme la France ne soient pas significativement différentes de celles de la moitié des pays du premier groupe.

Pour se faire une idée de la réalité des écarts de performance, on peut également les comparer aux niveaux de compétences définis par l'enquête PISA à partir des scores individuels des élèves. Selon le domaine considéré, l'échelle de compétence comporte cinq ou six niveaux distincts. En 2006, le niveau 1 de l'échelle de culture scientifique correspond à des scores compris entre 335 et 409 points et caractérise des élèves « *dont les connaissances scientifiques sont tellement limitées qu'ils peuvent uniquement les appliquer dans un petit nombre de situations familières* » ; au sommet de l'échelle, le niveau 6 de la culture

⁵ France, Allemagne, Autriche, Pays-Bas, Belgique, Suisse, Liechtenstein, Luxembourg, Islande, Finlande, Suède, Norvège, Danemark, Royaume-Uni, Irlande, États-Unis, Canada, Australie, Nouvelle-Zélande, Japon, Corée du Sud, Taipei chinois, Hong-Kong (Chine), Macao (Chine).

⁶ Portugal, Espagne, Italie, Slovénie, Croatie, Serbie, Monténégro, Grèce, Turquie, Roumanie, Bulgarie, Hongrie, République Tchèque, Slovaquie, Pologne, Lettonie, Estonie, Lituanie, Fédération de Russie, Azerbaïdjan.

⁷ Kirghizistan, Thaïlande, Indonésie, Tunisie, Israël, Jordanie, Qatar, Chili, Uruguay, Brésil, Colombie, Argentine, Mexique.

scientifique correspond aux scores supérieurs à 708 et caractérise des élèves « *capables d'identifier, d'expliquer et d'appliquer des connaissances en sciences et des connaissances à propos des sciences dans un éventail de situations complexes qui s'inspirent de la vie réelle* ». L'écart de score moyen entre deux niveaux successifs est de l'ordre de 75 points. Dans le domaine de la culture scientifique, le différentiel de performance moyen entre le premier et le dernier pays est égal à 241 points, soit un écart correspondant à trois niveaux de performance, ce qui est loin d'être négligeable. En revanche, au sein des pays appartenant au premier groupe, l'écart de performance maximum ne dépasse pas 70 points, c'est-à-dire moins d'un niveau de compétence, l'écart de performance typique étant quant à lui de l'ordre de 20 points. Lorsqu'on sait que l'OCDE estime qu'une année d'études représente en moyenne 38 points sur l'échelle PISA de compétence, on ne peut que relativiser la portée du classement international des acquis des élèves pour les pays les plus développés.

L'importance des contextes institutionnels

Au-delà de la connaissance des marges d'incertitude et de l'ampleur réelle des écarts de performances entre pays, une lecture avertie des résultats de l'enquête PISA nécessite qu'une attention particulière soit portée aux contextes éducatifs des pays participants.

Certaines caractéristiques institutionnelles des systèmes éducatifs nationaux sont en effet susceptibles d'influencer fortement les résultats obtenus par les élèves aux tests PISA. Bien que ces caractéristiques ne puissent en aucune manière être assimilées à des biais statistiques faussant les comparaisons internationales, elles n'en demeurent pas moins essentielles lorsqu'il s'agit d'interpréter ces comparaisons.

On aurait tort en particulier de croire que la population scolaire des élèves de 15 ans est homogène dans tous les pays : les règles qui régissent la définition des cohortes scolaires, l'orientation scolaire et la pratique du redoublement expliquent que la ventilation de cette population entre les différents niveaux d'enseignement varie énormément d'un pays à l'autre. En France, par exemple, les élèves âgés de 15 ans sont scolarisés dans plus de six classes différentes : le tableau qui suit indique qu'à peine plus de la moitié des élèves de la population visée est « à l'heure » ou en avance dans leur scolarité, c'est-à-dire au moins au niveau de la classe de 2^{nde}, alors près de 40 % d'élèves sont « en retard » parce qu'ils ont redoublé au moins une fois et sont scolarisés en classe de 3^e ou de 4^e.

Répartition des élèves français de 15 ans ayant participé au cycle 2003 de PISA

Classe fréquentée	Répartition
1ère générale et technologique	2,20% } 2% "En avance"
2nde générale et technologique	49,60% } 57% "À l'heure"
2nde professionnelle	7,40% }
3e générale	26,80% } 40% "En retard"
3e autre (SEGPA, Techno, Insertion)	7,70% }
4e	5,20% }
Autres niveaux scolaires	1,10% }
Ensemble	100%

Source : MEN-DEPP (2007)

Lecture : 49,6% des élèves français de 15 ans ayant participé au cycle PISA 2003 étaient scolarisés en classe de 2nde générale et technologique au moment de l'évaluation.

L'échantillon des élèves Français de l'enquête PISA n'est donc pas représentatif d'un niveau scolaire particulier, à la différence d'autre pays qui pratiquent le passage automatique comme la Finlande, le Royaume-Uni ou la Suède.

Les écarts de performances observés entre pays ne sont sans doute pas totalement indépendants de ces caractéristiques institutionnelles. On peut en effet penser que les élèves français de 15 ans qui ont redoublé au cours de leur scolarité sont structurellement désavantagés par rapport aux élèves « à l'heure » parce qu'ils ont répété le même programme scolaire deux années de suite et n'ont pas suivi les enseignements de la classe de 2^{nde}. Ce handicap structurel vient s'ajouter aux effets de stigmatisation et de perte de confiance en soi que suscite la pratique du redoublement ([MEN-DEPP, 2005](#)). On constate d'ailleurs empiriquement que les pays qui pratiquent le passage automatique affichent globalement de meilleurs résultats que les pays qui ont recours au redoublement, même si cette régularité empirique n'est pas suffisante pour prouver l'existence d'un lien de causalité entre cette pratique et la performance des élèves.

III/ Les performances des élèves français : un constat mitigé

L'intérêt de l'enquête PISA réside moins dans le classement de la performance éducative des États membres et partenaires de l'OCDE que dans l'identification des forces et faiblesses d'élèves scolarisés dans des contextes éducatifs nationaux très différents.

Des performances moyennes dans tous les domaines

Même si, comme nous venons de l'expliquer, il faut relativiser l'importance des écarts de performances entre les pays les plus développés, les résultats des élèves français n'en demeurent pas moins médiocres lorsqu'on les compare aux scores moyens affichés par certains de leurs voisins européens. Quel que soit le domaine de compétence ou l'année considérés, le score moyen obtenu par les élèves de 15 ans en France se situe presque systématiquement au niveau de la moyenne des pays de l'OCDE.

Le tableau suivant indique le score moyen obtenu par les élèves français dans chacun des trois domaines de compétence évalués par le programme PISA (compréhension de l'écrit, culture mathématique et culture scientifique) lors des trois vagues de l'enquête :

Score moyen des élèves Français dans les trois domaines de compétences évalués par PISA

Enquête PISA	Score moyen		
	Compréhension de l'écrit	Culture mathématique	Culture scientifique
PISA 2000 [intervalle de confiance à 95 %]	505 [500 ; 510]	517 [512 ; 522]	500 [494 ; 506]
PISA 2003 [intervalle de confiance à 95 %]	496 [491 ; 501]	511 [506 ; 516]	511 [505 ; 517]
PISA 2006 [intervalle de confiance à 95 %]	488 [480 ; 496]	496 [490 ; 502]	495 [488 ; 502]

Source : OCDE (2001, 2004 et 2007).

Lecture : en 2000, le score moyen des élèves Français aux évaluations de PISA dans le domaine de la compréhension de l'écrit était égal à 505, ce score ayant 95 % de chance d'être compris entre 500 et 510. Les valeurs en gras indiquent les scores correspondant au domaine de compétence privilégié par chaque enquête. Pour chaque domaine, le score moyen des élèves de l'ensemble des pays de l'OCDE est arbitrairement fixé à 500 avec un écart-type de 100.

Dans la moitié des cas environ, le score moyen des élèves français n'est pas significativement différent du score moyen dans l'ensemble des pays de l'OCDE qui est arbitrairement fixé à 500 points. Les résultats en mathématiques sont un peu supérieurs aux résultats en compréhension de l'écrit ou en culture scientifique et se situent en général au-dessus de la moyenne des pays de l'OCDE. La comparaison des scores moyens dans le temps

indique que les résultats des élèves français sont relativement stables, même si on note une légère dégradation des performances en compréhension de l'écrit et en mathématiques.

Les pays qui affichent des résultats significativement supérieurs à ceux de la France sont les pays d'Europe du Nord (Finlande, Danemark, Pays-Bas), une partie des pays anglo-saxons (Canada, Irlande, Australie, Nouvelle-Zélande) et certains pays asiatiques (Japon, Corée du Sud). En revanche, les performances de la France ne sont pas en général significativement différentes de celles de l'Allemagne, du Royaume-Uni, de l'Autriche, de la Belgique, de la Suisse ou encore des États-Unis. Enfin, le score moyen des élèves français est presque toujours significativement supérieur à celui des pays d'Europe de l'Est et du Sud, ainsi que des pays partenaires de l'OCDE.

Pour comparer la dispersion des résultats des élèves français avec celle des autres pays, on peut se référer à l'écart-type des scores individuels, indicateur classique de dispersion et qui est fixé arbitrairement à 100 dans les pays de l'OCDE pris dans leur ensemble. Le calcul de cet indicateur révèle que la France présente une moindre dispersion des résultats que la moyenne des l'OCDE en compréhension de l'écrit (92) et en culture mathématique (92) mais une dispersion équivalente en culture scientifique (101). La dispersion des performances des élèves français est cependant loin d'être la plus faible des pays de l'OCDE : l'écart-type du score des élèves finlandais (qui obtiennent presque systématiquement les meilleurs résultats aux évaluations PISA) est de 89 en compréhension de l'écrit, 83 en culture mathématique et 86 en culture scientifique.

Points forts et points faibles des élèves français

Les performances moyennes des élèves français dans les divers domaines de compétences évalués par l'enquête PISA masquent une grande variabilité des taux de réussite selon les items proposés. On s'aperçoit notamment que le succès ou l'échec à un item dépend en partie du format des questions, du support utilisé et de la nature des compétences testées. D'une manière générale, les élèves français semblent plus compétents lorsqu'il s'agit de prélever des informations dans un document ou de restituer des connaissances que lorsqu'on leur demande de mobiliser leurs connaissances et d'exercer leur esprit critique pour affronter des situations qui sortent du cadre scolaire.

Dans le domaine de la compréhension de l'écrit, les élèves français obtiennent de meilleurs résultats aux épreuves qui visent à évaluer les compétences « s'informer » (score moyen de 515) et « interpréter » (score de 505) qu'aux épreuves qui mesurent la compétence « réagir » (score de 496) et qui nécessitent en général le développement d'une opinion personnelle sous la forme de réponses écrites « ouvertes ». Par ailleurs, les scores sont plus élevés lorsque les items ont pour support des textes littéraires, des graphiques ou des tableaux que lorsque les élèves sont confrontés à des supports moins familiers (notices d'utilisation, articles de presse, textes à caractère publicitaire, etc.). Enfin, on constate que les élèves français sont relativement plus à l'aise lorsqu'ils ne sont pas sollicités dans le domaine de l'écrit mais qu'on leur propose une réponse sous forme de cases à cocher. Le taux de non réponse des jeunes français aux questions appelant des réponses construites est d'ailleurs l'un des plus élevés parmi les pays développés : 15,7 % contre 12,9 % en moyenne dans les pays de l'OCDE lors de l'enquête PISA 2000.

En mathématiques, les élèves français réussissent beaucoup mieux les items correspondant à une formulation proche du contexte scolaire ou dont les contenus se rapprochent de l'enseignement mathématique dispensé en France : exercices sur des supports géométriques, lecture de graphiques, application et établissement de formules. Ils éprouvent en revanche des difficultés à construire une relation entre des variables et obtiennent des résultats décevants lorsqu'il s'agit de travailler sur des nombres entiers et décimaux. Ainsi que le notent les rédacteurs du rapport de la DEPP consacré aux résultats de l'enquête PISA 2003 ([MEN-DEPP, 2007](#)) : « *les points faibles des élèves français semblent résider dans la capacité à effectuer des généralisations (par exemple, établir une formule) et, de façon générale, à prendre des initiatives sans se référer à un schéma connu* ». Les mauvais résultats obtenus dans le domaine des probabilités s'expliquent quant à eux en partie par le fait que ces dernières ne sont pas étudiées en France avant le lycée.

Dans le domaine de la culture scientifique, les résultats de la dernière enquête PISA révèlent que les résultats des élèves français sont supérieurs à la moyenne lorsqu'il s'agit de prélever des informations dans des supports habituellement utilisés dans l'enseignement scientifique (graphiques, tableaux, croquis) mais que les jeunes français de 15 ans ont des difficultés à mobiliser leurs connaissances pour expliquer des phénomènes de manière scientifique dans des situations de la vie courante non évoquées en classe. On constate également que leurs résultats varient considérablement selon le type de connaissances évaluées : le score moyen aux questions destinées à évaluer leurs connaissances *en sciences*

(qui font référence à la connaissance du monde naturel tel qu'il se définit à travers les disciplines scientifiques) est inférieur au score moyen des élèves de l'OCDE, tout particulièrement dans la catégorie « Systèmes de la Terre et de l'univers » (463 points contre 500 en moyenne dans l'OCDE). Les performances françaises sont néanmoins supérieures à la moyenne (score moyen de 507) pour les connaissances *à propos* de la science, qui font référence aux moyens utilisés par la science (démarche scientifique) et à ses objectifs (explications scientifiques).

Quelques tentatives d'explications

L'examen attentif des taux de réussite aux items proposés dans les versions successives de l'enquête PISA permet non seulement d'identifier les forces et faiblesses des élèves français, mais également de formuler un certain nombre d'hypothèses permettant de mieux comprendre le positionnement international de la France.

La pratique massive du redoublement en France constitue un premier facteur explicatif. Ainsi qu'il a été évoqué précédemment, près de 40 % des élèves français âgés de 15 ans sont « en retard » parce qu'ils ont redoublé une fois ou plus au cours de leur scolarité. Or, comme le signale la DEPP dans le rapport qu'elle a consacré à ce thème ([MEN-DEPP, 2005](#)), le retard scolaire est en France le facteur le plus « explicatif » des variations de performances entre élèves : sur l'échelle de l'évaluation des compétences en compréhension de l'écrit de l'enquête PISA 2000, les élèves français « à l'heure » à 15 ans et scolarisés en 2nde générale et technologique obtenaient un score de 560 points qui les situait au niveau du score moyen obtenu par les élèves finlandais, alors que les élèves ayant un an de retard obtiennent un score moyen de 430 points qui les plaçait au niveau des pays occupant la dernière place du palmarès international. S'il est difficile de déterminer l'influence précise du redoublement, on peut néanmoins penser qu'il tend à affaiblir les performances aux épreuves de PISA des élèves « en retard », en pesant négativement sur leur sentiment de performance scolaire (les évaluations bilan de fin de collège indiquent que les élèves en retard ont tendance à sous-estimer leur niveau réel de compétence) et en limitant leur accès aux savoirs qui ne sont pas enseignés avant le lycée (les probabilités, par exemple).

Un second facteur explicatif est à chercher du côté des programmes et des méthodes d'enseignement français. On peut en effet se demander si les performances relativement médiocres des élèves français dans les exercices exigeant la formulation d'hypothèses, la mise

en relation d'informations disparates ou encore l'expression d'un jugement critique ne proviennent pas en partie d'une approche pédagogique trop exclusivement axée sur l'accumulation de savoirs reçus passivement par les élèves, sans que l'utilité intrinsèque des compétences leur soit suffisamment expliquée. Moins souvent confrontés à des exercices s'inspirant de situations rencontrées dans la vie quotidienne, rarement sollicités pour débattre oralement, les élèves français ne sont sans doute pas suffisamment initiés au débat contradictoire, à l'élaboration de réponses argumentées et se retrouvent plus souvent démunis lorsqu'on fait appel à leur avis personnel ou à leur expérience propre. Ce phénomène pourrait expliquer le taux plus élevé de non réponse aux questions appelant des réponses longues et la difficulté à envisager un document d'un point de vue critique.

Un troisième facteur à prendre en compte pour expliquer le niveau globalement moyen des résultats des élèves français renvoie au rôle des enseignants et à leurs relations avec les élèves. Un certain nombre d'indices suggèrent en effet que les élèves français ne se sentent pas suffisamment accompagnés par les enseignants dans leur apprentissage ni suffisamment encouragés à progresser. C'est ce qui ressort des réponses fournies dans le cadre de l'enquête PISA 2000 par les élèves français à propos de leurs relations avec les enseignants. Elles sont reproduites dans le tableau suivant :

Comparaisons des réponses des élèves français et de des élèves de l'ensemble des pays de l'OCDE aux questions relatives aux relations entre élèves et enseignants (enquête PISA 2000)

Question	France				OCDE			
	Jamais	Parfois	Souvent	Toujours	Jamais	Parfois	Souvent	Toujours
L'enseignant(e) veut que les élèves travaillent beaucoup	8,9%	45,0%	31,5%	14,5%	6,5%	24,6%	29,4%	39,4%
L'enseignant(e) dit aux élèves qu'ils pourraient mieux faire	4,8%	45,4%	33,3%	16,5%	8,6%	33,0%	29,3%	29,0%
L'enseignant(e) est mécontent lorsque les élèves rendent un travail négligé	7,5%	52,2%	25,4%	14,9%	16,2%	33,0%	23,9%	26,8%
L'enseignant(e) s'intéresse à l'apprentissage de chaque élève	10,4%	33,4%	29,7%	26,5%	10,4%	27,8%	26,3%	35,5%
L'enseignant(e) donne aux élèves l'occasion d'exprimer leurs opinions	9,8%	28,0%	27,4%	34,8%	6,6%	23,4%	27,7%	42,2%
L'enseignant(e) aide les élèves avec leur travail	10,4%	38,4%	28,6%	22,6%	10,6%	30,8%	28,4%	30,2%
L'enseignant(e) continue d'enseigner jusqu'à ce que les élèves comprennent	9,6%	32,5%	28,5%	29,4%	8,1%	28,6%	28,7%	34,6%
L'enseignant(e) s'investit beaucoup pour aider les élèves	13,8%	37,4%	28,8%	20,0%	8,5%	29,9%	30,2%	31,4%
L'enseignant(e) aide les élèves dans leur apprentissage	16,7%	39,9%	27,5%	15,9%	7,8%	27,6%	30,1%	34,5%
L'enseignant(e) vérifie les devoirs des élèves	11,8%	43,1%	26,9%	18,3%	8,5%	31,5%	29,0%	30,9%
	Pas du tout d'accord	Pas d'accord	D'accord	Tout à fait d'accord	Pas du tout d'accord	Pas d'accord	D'accord	Tout à fait d'accord
Les élèves s'entendent bien avec la plupart des enseignant(e)s	7,3%	28,5%	56,9%	7,2%	6,8%	23,2%	57,3%	12,7%
La plupart des enseignant(e)s s'intéressent au bien-être de leurs élèves	7,0%	27,8%	56,3%	8,8%	6,4%	22,3%	54,7%	16,7%
La plupart des enseignant(e)s sont réellement à l'écoute de ce que j'ai à dire	7,1%	31,2%	50,9%	10,7%	6,9%	25,6%	53,3%	14,2%
Si j'ai besoin d'aide supplémentaire, mes enseignant(e)s me l'apporteront	5,7%	18,4%	60,3%	15,6%	5,4%	17,1%	58,0%	19,5%
La plupart des enseignant(e)s me traitent avec justice	7,0%	20,3%	59,3%	13,4%	5,9%	16,3%	57,1%	20,6%

Source : Calculs de l'auteur à partir de la base PISA 2000.

Lecture : À la question « l'enseignant(e) veut-il que les élèves travaillent beaucoup ? », 8,9% des élèves français interrogés lors du cycle PISA 2000 répondent « jamais », 45% « parfois », 31,5% « souvent » et 14,5% « toujours ».

À la question « l'enseignant(e) aide-t-il les élèves dans leur apprentissage ? », seuls 43,4 % des élèves français déclarent que c'est souvent ou toujours le cas, alors que cette proportion atteint 64,6 % en moyenne dans les pays de l'OCDE. De même, seule une minorité d'élèves français interrogés (48,8 %) considère que « l'enseignant(e) s'investit beaucoup pour aider les élèves » contre 61,6 % dans l'ensemble des pays de l'OCDE. Il est probable que ce sentiment de moindre investissement des enseignants affecte négativement la motivation des élèves français au cours de leurs études en général et lors du passage des épreuves de PISA en particulier. Lors du cycle PISA 2003, les élèves ont été invités à évaluer sur une échelle de 10 l'effort qu'ils avaient fourni pour répondre aux différents items : avec un effort moyen de 7 sur 10, la France s'est classée avant-dernière parmi l'ensemble des 41 pays ayant participé à l'enquête...

Loin d'invalider le score global des élèves français, l'identification des facteurs que nous venons de passer en revue permet de procéder à une analyse plus lucide des résultats de l'enquête PISA que la lecture naïve du classement international. Les leçons que la France pourra tirer de cet ambitieux programme seront d'autant plus utiles que l'impact de ces différents facteurs aura été précisément évalué.

Conclusion : intérêt et limites de l'éducation comparée

Le programme PISA ne prétend pas fournir une évaluation de la qualité globale des systèmes éducatifs à travers le monde car il se limite à un aspect particulier : leur capacité à former à des élèves de 15 ans capables d'utiliser leurs connaissances de manière relativement autonome pour résoudre efficacement les problèmes qu'ils sont susceptibles de rencontrer dans leur vie d'adulte.

Malgré la grande rigueur méthodologique de cette enquête, l'importance de la marge d'erreur qui affecte les scores nationaux ainsi que la faiblesse des écarts de performances qui séparent les pays les plus développés rendent relativement vaine l'ambition de classer de manière univoque les pays participants sur une échelle de compétence commune.

Le principal mérite de ce programme réside plutôt dans la richesse des données recueillies et dans sa périodicité régulière : la combinaisons d'informations détaillées sur les performances des élèves, leur environnement social et familial et sur les caractéristiques

institutionnelles des établissements scolaires fréquentés constitue une avancée majeure dans la connaissance des déterminants de la réussite scolaire à l'échelle internationale.

Texte paru dans laviedesidees.fr, le 8 février 2008

© laviedesidees.fr