

Evaluations, mesures ou classements ? Débattre des enquêtes PISA.

Marie Duru-Bellat, Professeur Emérite à Sciences Po-Paris, Membre de l'Observatoire sociologique du changement et de l'Institut de recherche sur l'Education.

Résumé

Lancées au seuil des années 2000, les enquêtes PISA (Programme international sur les acquisitions des élèves) s'inscrivent dans un contexte européen spécifique marqué par l'objectif d'instaurer une politique européenne en matière d'éducation, et caractérisé par la montée d'une gouvernance dite « douce » par le biais du benchmarking, soit une gouvernance se voulant suggestive, fondée sur des indicateurs comparatifs censés alimenter une saine émulation. Après avoir décrit brièvement ce contexte, le texte rappellera les principales caractéristiques techniques de ces enquêtes, avant de développer une discussion sur ce qu'elles apportent, ainsi que sur les dérives possibles : peuvent-elles prétendre constituer aujourd'hui un instrument sans faille pour des comparaisons internationales des systèmes éducatifs, et/ou un levier pour des transformations et des politiques visant ces systèmes ?

Mots clefs : évaluation, comparaisons internationales, systèmes éducatifs.

Assesment, measurement or ranking? Debating about PISA Surveys

Abstract

Launched at the beginning of the 21rst century, PISA surveys (Programme for international student assessment) are embedded in a specific European context, whose purpose is to establish a European policy in the education field, and are characterized by the development of a soft governance through benchmarking, that is a supposed suggestive one only, based on comparative indicators, supposed to foster a sound emulation. After having rapidly described this context, the paper will recall the main technical characteristics of these surveys, before developing a discussion about what they bring about, and also about their possible abuses: can they claim to make up perfect tools for international comparisons of educational systems, and/or an efficient lever for some changes or policies concerning these systems?

Keys words: assessment, international comparisons, educational systems.

Les comparaisons internationales ont toujours été jugées potentiellement heuristiques par les chercheurs qui s'intéressent aux modes de fonctionnement des systèmes éducatifs et aux modalités de leur impact sur les performances des élèves, dès lors que les expérimentations sont impossibles dans ces domaines et que les sciences sociales s'efforcent pourtant de rechercher des régularités (voire des « lois ») dans les fonctionnements sociaux. Aux yeux des responsables politiques, un ensemble de facteurs est venu conforter la pertinence de cette perspective comparatiste: les institutions internationales n'ont de cesse d'invoquer une inévitable (mais censée stimulante) compétition internationale dans une « économie de la connaissance » ouverte, et dans un contexte de contrainte budgétaire étendant l'exigence d'efficacité à tous les services publics. Joue peut-être aussi, en filigrane, une incertitude quant aux objectifs intrinsèques des politiques éducatives : se donner comme objectif, en matière d'éducation, le « toujours plus » ou le « les autres font... » fait passer au second plan la question du pourquoi... Une chose est sûre, le choix des indicateurs et les comparaisons qui seront engagées sur cette base revêtent une dimension politique autant sinon plus que scientifique, dans un contexte où l'influence des instances internationales dans les politiques éducatives s'est subrepticement accrue.

Depuis les objectifs promus à Lisbonne en 2000, l'Europe s'est mise d'accord sur une liste d'objectifs, d'abord définis à l'horizon 2010, puis renouvelés ensuite, qui ont marqué la politique éducative française (tels que élever le pourcentage de jeunes au niveau du bac, lutter contre les sorties précoces, baisser le taux d'élèves faibles en lecture, etc.) ; ils se sont traduits par une série de repères chiffrés (les « benchmarks ») évalués régulièrement par la Commission européenne¹. Au-delà de l'Europe, l'OCDE est en première ligne, tant du fait de son obsession pour la croissance économique que de sa conviction quant au caractère stratégique crucial de l'éducation à cet égard. Prolige en matière d'études sur les compétences et les apprentissages, l'OCDE s'est d'emblée engagée (notamment financièrement) pour soutenir les enquêtes PISA (et plus généralement d'autres enquêtes à grande échelle du même type comme les enquêtes IALS, sur la « littératie » des adultes), gérées opérationnellement par un consortium de spécialistes en évaluation, docimologie, et psychométrie. D'emblée tournées vers les décideurs publics, ces enquêtes se donnent pour vocation explicite d'orienter le pilotage des systèmes éducatifs.

1. Qu'appréhendent les enquêtes PISA ?

Rappelons en quelques mots que les enquêtes PISA -*Programme for International Student Assessment*- visent, depuis 2000 et tous les 3 ans (la dernière date de 2018, et ses résultats sont annoncés pour la fin 2019), à évaluer la compréhension de l'écrit, la culture mathématique et la culture scientifique des jeunes de 15 ans, et ceci sur la base de situations de la vie réelle, l'objectif étant d'évaluer la maîtrise de *skills for life* –de compétences pour la vie- et non des connaissances disciplinaires. De plus, ont été ajoutées des questions sur la qualité des relations entre professeurs et élèves, ainsi que sur les attitudes de ces derniers face au travail scolaire. Ces enquêtes, qui pour la dernière vague (2018) concernent 80 pays (de l'OCDE, avec, au-delà, un ensemble de pays volontaires), approfondissent tour à tour l'un des trois domaines de compétences visés, en ajoutant, selon les années, d'autres

¹ Cf. notamment Commission of the European Communities, *Progress Towards the Lisbon Objectives in Education and Training. Indicators and benchmarks*, Brussels, 30/03/2011. Une note récente de la DEPP est consacrée au bilan de la stratégie européenne en matière d'éducation (Note 19.09 de mars 2019). Voir aussi les publications annuelles de l'OCDE, *Education at a Glance*, ou le document plus complet, traduit en Français sous le titre *Regards sur l'éducation*.

explorations plus spécifiques que les pays qui le souhaitent peuvent inclure : la « compétence en résolution collaborative de problèmes »², en 2015, ou encore les connaissances en matière financière pour 2018, ainsi que la notion de « compétence globale », présentée comme la capacité à appréhender et à discuter des questions aussi vastes que le changement climatique, la malnutrition dans le monde, l'égalité hommes/femmes... Depuis 2015, ces épreuves se passent en classe sur ordinateur. Soulignons qu'elles ont comme spécificité de concerner les élèves âgés de 15 ans, parce que c'est l'âge auquel prend fin la scolarité obligatoire dans nombre de pays, et non ceux d'un niveau scolaire donné. C'est là un changement par rapport aux enquêtes TIMSS testant les mathématiques et les sciences³, ou PIRLS (lecture-écriture)⁴, calées sur les programmes nationaux et donc imparfaitement comparables d'un pays à l'autre, sachant que la focalisation des secondes sur le niveau primaire limite quelque peu l'ampleur de cette difficulté, puisqu'à ce stade, on se centre sur des acquis de base en lecture et calcul. Notons d'emblée que le fait de caler les évaluations sur un âge donné et non un niveau scolaire va affecter de manière importante les résultats des élèves dans les pays (minoritaires) où le redoublement existe : on imagine aisément que les performances des élèves seront meilleures s'ils sont scolarisés à un niveau correspondant à la première année du lycée (la 2nde en France, où sont scolarisés seulement 73% des élèves testés, et plus précisément 59% en 2nde générale ou technologique) que s'ils sont encore dans une structure de type collège...

Cette double caractéristique des enquêtes PIA –tester les élèves d'un âge donné, et se focaliser sur les compétences jugées nécessaires pour s'intégrer dans la vie, a été l'objet de critiques, particulièrement en France. Au-delà, peut-être, de la conviction implicite diffuse selon laquelle nos programmes scolaires sont évidemment les meilleurs, on argue que les savoirs de nos élèves sont mal valorisés par ce type de questionnement (notamment parce qu'environ la moitié des items sont sous forme de QCM), ou encore on critique la notion même de compétence ne serait-ce que parce qu'elle apparaît par trop anglo-saxonne (Duru-Bellat, 2015). Une chose est sûre, ce qui est testé dans PISA n'est pas directement relié aux programmes scolaires de chaque pays : il s'agit d'évaluer la capacité des élèves à comprendre des textes variés quant à leur forme et à leur contenu (depuis des textes littéraires jusqu'à des articles de presse), et non de connaître certaines œuvres précises, forcément différentes d'un pays à l'autre. On demande aux élèves d'y repérer des informations, de les interpréter, de réfléchir pour répondre aux questions posées à leur propos. En d'autres termes, et c'est en cela que la notion de « compétence pour la vie » est mobilisée, il s'agit de voir s'ils sont capables de mettre en œuvre ce qu'ils ont appris en classe dans le contexte de la vie quotidienne. Même si toute évaluation inclut inévitablement une part d'arbitraire et de flou –en l'occurrence, quelles seront les exigences de la vie d'adulte des élèves testés ?-, il est difficile d'écarter d'un revers de main cette notion de « savoir en acte » et de conclure que l'école n'a pas à s'y intéresser !

² On désigne ainsi la capacité à s'engager efficacement avec autrui dans la résolution d'un problème, en mettant en commun ses connaissances, ses compétences et ses efforts.

³ Les enquêtes TIMSS (Trends in mathematics and science study) portent d'une part sur les performances des élèves achevant leur 4^{ème} année de scolarité obligatoire, soit le CM1 pour la France, d'autre part sur les performances en mathématiques des élèves finissant leur scolarité secondaire dans une filière scientifique, soit la Terminale S pour la France ; voir respectivement la Note d'Information 16.33 et 16.35 du Ministère de l'Éducation.

⁴ Pour une présentation des enquêtes PIRLS (Progress in reading literacy) évaluant la lecture chez les élèves achevant leur 4^{ème} année de scolarité obligatoire, soit le CM1 pour la France, voir la Note d'Information 03.22 du Ministère de l'Éducation.

Parmi les particularités des enquêtes PISA, la volonté de transparence doit être notée, puisque dès le départ, les tests et les modalités techniques des enquêtes sont accessibles sur internet (à l'adresse ocde.org/PISA). Pour donner une image plus précise de ce qu'apportent ces enquêtes, examinons brièvement les principaux résultats qu'elles produisaient dans la vague 2015 (puisque les résultats de la vague 2018 ne sont pas encore disponibles).

2. Qu'ont acquis, à cette aune, les élèves français de 15 ans ?

Alors que les premières vagues des enquêtes PISA ne suscitaient guère de réaction officielle en France (à la différence d'autres pays voisins comme l'Allemagne), faisant d'autant plus l'objet d'un déni qu'elles n'étaient guère flatteuses, elles sont à présent scrutées par les politiques et les médias. Ainsi, dès la diffusion des résultats de la vague 2015 (seuls connus à ce jour), le Ministère de l'Education s'est montré particulièrement réactif en publiant deux « Notes d'Information » aisément accessibles puisque téléchargeable (n°37 et 38, décembre 2016) pour présenter les principaux résultats nationaux. La presse s'est faite relativement discrète, sans doute parce que la stabilité dominait, quant aux performances des élèves français, telles qu'observées dans PISA, notamment depuis 2006 : on ne pouvait pas cette fois pointer la dégradation du niveau moyen ni notre dégringolade dans le classement, notées lors des vagues antérieures.

En effet, que ce soit en sciences (matière phare de l'enquête 2015), en mathématiques ou en compréhension de l'écrit, les résultats sont stables par rapport aux enquêtes précédentes et proches de la moyenne des pays de l'OCDE. Mais l'examen des classements dans les trois domaines testés n'en mérite pas moins un détour et il décelez quelques surprises : si l'on tient compte des marges d'erreur, rien ne nous distingue, en compréhension de l'écrit, du Danemark, de la Belgique, du Royaume-Uni ou des Etats-Unis ; mais ce peloton des pays moyens se disloque quand on se penche sur les résultats en mathématiques : cette fois, si nous restons proche du Royaume-Uni, le Danemark et la Belgique font significativement mieux que nous, tandis que les Etats-Unis sont un des pays les plus faibles du classement. En Sciences, le Royaume-Uni passe cette fois en tête (certes, après le groupe de tête stable que sont les pays asiatiques, le Canada et la Finlande), et nous nous retrouvons, dans la moyenne, avec les Etats-Unis.

Ce que ces constats suggèrent, c'est que, comme le pointent souvent les interprétations courantes, les structures éducatives (depuis l'organisation au niveau collège, jusqu'au statut des enseignants) ne font pas tout, car si tel était le cas, on devrait observer des performances plus homogènes d'une matière à l'autre. Il faut donc plutôt invoquer ici d'autres facteurs, non pris en compte dans PISA comme les programmes scolaires nationaux, voire l'environnement culturel (on connaît la qualité des documentaires scientifiques britanniques !). Toujours est-il que notre pays est moyen certes, mais homogène, qualité en général louée sur les bulletins scolaires... Un autre point positif est que la baisse des performances moyennes observée entre 2000 et 2003 semble bel et bien enrayerée puisque depuis 2006 c'est la stabilité qui domine.

Mais il est tout aussi important d'examiner, derrière la stabilité des moyennes, la dispersion des résultats et son évolution. Et en la matière, le constat est plus préoccupant : car en compréhension de l'écrit comme en mathématiques, on assiste à une augmentation du pourcentage d'élèves faibles ; ce facteur de baisse est compensée en compréhension de l'écrit par une hausse des élèves les plus forts, tandis que ce dernier chiffre est stable en mathématiques (où c'est le poids relatif des élèves moyens qui a un peu augmenté). Globalement, nous restons

l'un des pays (avec notamment la Belgique) où les écarts de performance entre élèves sont les plus marqués. Une raison importante est l'existence, dans notre pays, du redoublement : ainsi en sciences, la performance des élèves de 15 ans scolarisés en classe de 2nde (qui n'ont donc jamais redoublé) est de 545, ce qui les place dans le haut du classement (la moyenne des tests étant calée sur 500), tandis que celle des élèves ayant un an de retard scolarisés en 3^{ème} est de 407, ce qui les place à l'autre extrémité du classement. Globalement, si on ne tenait compte que des résultats des élèves « à l'heure », la France figurerait aisément dans le peloton de tête des pays évalués ; il y a bien dans notre pays un problème spécifique avec les élèves les plus faibles.

A cette forte variabilité des performances est associée une forte influence du milieu social : la France est l'un des pays où celles-ci sont les plus affectées par le milieu familial de l'élève, avec, pour prendre le cas de la culture scientifique (domaine pas moins inégalitaire que les performances face à l'écrit), un écart de 118 points entre élèves issus de milieu très défavorisé et très favorisé, contre 88 points en moyenne dans les pays de l'OCDE. Cet écart est de 103 points en Allemagne, pays qui était, lors de la première enquête PISA, un des pays les plus inégalitaires de l'OCDE, mais où cela a déclenché un vif sursaut ; il est de 90 points aux Etats-Unis, où les inégalités sociales (entre les familles) sont de, manière générale, bien plus fortes que dans les pays socialement plus égalitaires. On remarque aussi qu'une spécificité française est que le temps de travail à la maison (autre information recueilli dans PISA) est associé statistiquement aux inégalités sociales de compétence : en d'autres termes, une part de l'influence du milieu social de l'élève sur ses performances passe par le fait que le temps de travail à la maison est d'autant plus long que cet environnement est favorisé. Autre spécificité, le climat scolaire (caractérisé par des questions sur le bien-être à l'école, le sentiment d'être soutenu par les enseignants, etc.) se révèle assez médiocre dans notre pays, ce qui pourrait expliquer, pour une part, l'importance des inégalités sociales, tant les recherches montrent que les élèves les plus fragiles sont aussi plus sensibles à leur environnement scolaire (Meuret et Morlaix, 2006 ; Le Mener et Meuret, 2017).

Pour en revenir aux résultats en culture scientifique (les plus détaillés dans la vague 2015), on note que les écarts de performance entre garçons et filles ne sont pas significatifs en France (c'est aussi vrai pour ce qui est de la culture mathématique), alors qu'en moyenne, dans les pays testés, les garçons font légèrement mieux. C'est l'inverse pour ce qui est de la compréhension de l'écrit : les filles s'y montrent meilleures, mais dans ce domaine, leur avantage sur les garçons est comparable à ce qui est observé en moyenne ailleurs.

Les résultats livrent également des informations de nature pédagogique, que seul un examen précis des items soumis aux élèves permet de véritablement interpréter. Par exemple, on observe que quand il s'agit d'expliquer des phénomènes de manière scientifique, en proposant et en évaluant des thèses, les élèves français sont un peu moins performants que la moyenne, alors que face à une tâche d'interprétation de données scientifiques (quand des thèses ou des arguments sont proposés), ils font plutôt un peu mieux. Dans le champ de la littéracie, les élèves français sont plus à l'aise quand on leur demande de prélever des informations que quand il s'agit d'analyser un texte de manière autonome ou de répondre à des questions générales et d'exprimer un point de vue critique. De manière générale, ils manifestent une tendance plus marquée que dans les autres pays à s'abstenir de répondre aux questions, notamment les plus faibles d'entre eux, soit qu'ils craignent d'être stigmatisés par leurs erreurs, soit que l'élaboration d'une réponse soit pour eux par trop difficile. Peut-être aussi, et ceci vaudrait pour tous les élèves cette fois, les élèves français sont-ils peu familiers avec le type de questions couramment posées dans PISA, qui

relèvent de situations de la vie quotidienne (lecture d'une notice de médicament, par exemple), ayant du mal à faire le lien avec les approches plus abstraites prévalant en cours.

3. Quel intérêt des enquêtes PISA pour les politiques, les chercheurs et tous les acteurs ?

Dans une optique davantage focalisée sur les résultats du système en tant que tel, les enquêtes PISA sont également précieuses : elles ont le mérite de mettre en évidence une variété dans ce qui est parfois considéré avec un certain fatalisme, qu'il s'agisse du caractère inévitable des inégalités sociales par exemple, ou du fait qu'il y a nécessairement un pourcentage d'élèves très faibles, ou encore du prétendu incontournable « arbitrage efficacité/équité » (selon lequel se préoccuper des performances de tous reviendrait à sacrifier les plus forts). Les enquêtes PISA bousculent aussi nombre d'idées reçues comme le lien entre taille des classes et performances des élèves -les classes sont très chargées dans les pays asiatiques les plus performants- ou encore l'effet bénéfique des dépenses par élève –modérées à la fois dans certains pays performants comme la Finlande et d'autres qui le sont moins comme la Grèce.

En d'autres termes, ces enquêtes permettent de rechercher des corrélations entre les performances enregistrées dans les pays, d'une part, et de l'autre certaines caractéristiques des systèmes éducatifs et certaines réformes, même si chacun sait qu'une corrélation est insuffisante pour établir des relations causales : qui soutiendrait que c'est précisément du fait de leurs classes très chargées que les pays asiatiques se montrent si performants ? Plus près de nous (et plus comparable), si l'Allemagne voit ses performances s'améliorer depuis quelques années, c'est peut-être plus par l'aiguillon qu'a constitué dans ce pays le « choc PISA » des premières enquêtes (soit la révélation de résultats médiocres et inégalitaires) que par les réformes précises qui ont été lancées dans ce pays pour y réagir (cf. Le Donne, 2014). Le seul fait que les enquêtes aient été prises très au sérieux et qu'un consensus national ait émergé pour faire front est déjà sans doute très important, et le contraste est frappant avec le relatif scepticisme observé dans notre pays face à ces données objectives, mâtiné d'un certain fatalisme eu égard au taux élevé d'élèves faibles et aux inégalités sociales⁵. Les enquêtes PISA incarnent une conception « expérimentale » de l'éducation à laquelle les pays sont plus ou moins perméables : on évalue les élèves, certes toujours imparfaitement mais sur une base qui se veut objective et transparente, et on fait le pari que les choses peuvent s'améliorer, ce qui introduit un souffle d'air positiviste et comparatif dans les débats scolaires volontiers englués dans des considérations idéologiques et corporatistes (Lafontaine et Baye, 2012).

Il reste que, pour interpréter ces corrélations entre telle ou telle caractéristique des systèmes éducatifs et les performances des élèves (et les inégalités qui les marquent), ainsi que pour dépasser certaines conclusions de prime abord paradoxales, il est nécessaire de mobiliser ce qu'on sait par ailleurs des systèmes éducatifs (Rey, 2011). C'est vrai notamment des corrélations, livrées de manière récurrente par les enquêtes PISA, entre l'ampleur des inégalités sociales et les caractéristiques des systèmes. De prime abord, le système français devrait être un des moins inégalitaire, puisque comprenant nombre des caractéristiques associées le plus souvent à une égalité, scolaire et sociale, des performances. Nous avons un enseignement pré-élémentaire très développé et surtout nous avons depuis longtemps instauré un tronc commun au collège, un facteur *a priori* favorable à l'égalité, puisque les

⁵ Il y a peut-être là un effet pervers de la diffusion, sous une forme vulgarisée, des thèses bourdieusiennes, selon lesquelles la « fonction » du système éducatif serait précisément de reproduire les inégalités.

inégalités sont en général plus fortes dans les systèmes qui prévoient des filières dès l'âge de 14 ans (ou même avant) : l'âge à la première sélection dans le cursus scolaire a même un impact plus fort sur les inégalités sociales de performance que l'existence d'un choix de l'école par les parents (régulièrement associé à plus d'inégalités). Mais nous continuons à faire redoubler les élèves en difficulté, plus souvent de milieu social défavorisé, ce qui tire, on l'a vu, le niveau moyen vers le bas. Les inégalités entre les élèves dépendent aussi de l'importance de la ségrégation sociale et scolaire entre les établissements et en la matière, notre situation s'est dégradée avec notamment l'assouplissement de la carte scolaire, qui favorise, du fait de la fuite des élèves les moins défavorisés, la constitution d'établissements très populaires aux performances d'autant plus médiocres (Merle, 2012). On observe de manière générale que dans les systèmes les moins inégalitaires, le choix de l'école est régulé par les autorités, la décentralisation est encadrée et la part du secteur privé est faible, alors qu'à l'inverse, les systèmes plus orientés vers des logiques de marché scolaire valorisant la concurrence entre écoles débouchent plus souvent sur une ségrégation à la fois scolaire et sociale, qui accentue les inégalités entre élèves.

D'autres facteurs jouent sans doute également pour rendre compte de la spécificité française: notre système de grandes écoles –verrou de l'accès à l'élite- qui durcit la compétition entre élèves dès le secondaire, et des contenus de formation très académiques ; encore plus important peut-être, il faut compter avec une formation des enseignants peu développée (notamment la formation continue), dans un contexte où l'on méprise volontiers les questions pédagogiques, avec, non sans rapport sans doute, des relations entre maîtres et élèves plus mauvaises qu'ailleurs, nous l'avons évoqué. Il faudrait également prendre en compte le contexte socio-économique plus large et à cet égard, l'articulation forte qui existe dans notre pays, entre diplômes et emplois peut jouer un rôle dans cette importance des inégalités sociales scolaires : en effet, la réussite des études constitue un enjeu d'autant plus prégnant, et dans cette compétition alors durcie, certains élèves ont plus d'atouts que les autres ; volontiers perçue comme un gage de méritocratie, l'emprise des diplômes peut jouer contre la justice (Duru-Bellat, 2019)... Effectivement, les comparaisons internationales montrent que plus l'emprise des diplômes sur les emplois et les salaires est marquée, plus les inégalités sociales à l'école tendent à l'être également (Dubet *et al.*, 2010).

Les comparaisons internationales éclairent donc les racines de cette spécificité française que sont des inégalités sociales particulièrement fortes. De plus, même si elles montrent aussi qu'il n'existe pas un « one best way » -une solution unique pour s'améliorer-, elles peuvent suggérer des pistes politiques. A cet égard, les évolutions constatées entre les dernières enquêtes PISA (2009-2012 notamment) constituent des expériences en vraie grandeur. Ainsi, qu'a-t-on fait dans les pays où les scores moyens ont augmenté sur la période ? Revenons à l'Allemagne, où ont été développés des standards et des évaluations nationales, et où l'on a accru le nombre d'écoles offrant désormais une journée complète de cours. En Pologne, autre pays dont les progressions sont frappantes, on a créé un cycle secondaire inférieur unifié (à l'instar de notre collège), à la place des anciennes filières. Au Portugal, on peut évoquer pêle-mêle la forte augmentation des ressources pour les publics défavorisés, la réduction du retard scolaire, le pilotage plus strict du système sur la base d'évaluations nationales. *A contrario*, dans un pays comme la Suède dont les élèves avaient, dans les premières enquêtes, des performances élevées et égalitaires, et qui a vu ses performances se dégrader sensiblement, a été conduite une politique de décentralisation des établissements et de libéralisation, avec un système de chèques scolaires permettant aux élèves de choisir leur école ; il s'en est suivi une ségrégation accrue des établissements, dont les élèves les plus défavorisés ont payé le prix. Il semblerait donc que l'on puisse infléchir l'ampleur des inégalités, dans un sens comme dans un autre, par des politiques éducatives.

Mais bien sûr, d'autres facteurs peuvent être en cause. Alors que de manière régulière, les élèves des pays asiatiques obtiennent des performances remarquables, notamment en mathématiques ou en sciences, il faut invoquer, ce qui est rarement fait, les particularités de langues comme le japonais, le coréen ou le chinois, dont le système de désignation des nombres rend plus aisée la maîtrise, très jeune, du système de numération (Hee Yean *et al.*, 2008). On peut ajouter que dans ces pays, l'implication des mères auprès des tout jeunes enfants est très marquée et vise très tôt des acquisitions qui dans d'autres pays se feront à l'école élémentaire, la pression des parents se poursuivant toute la scolarité durant, redoublée par des cours de soutien privés quasi généralisés. Qui plus est, il faudrait pouvoir évaluer, ce qui est quasiment impossible, l'impact probable d'un arrière-plan culturel et religieux –en l'occurrence le confucianisme- qui valorise le savoir et le respect des maîtres.

Les comparaisons internationales peuvent aussi ébranler ce qui est souvent considéré comme évident par les chercheurs eux-mêmes. On a ainsi pu établir qu'il n'y avait pas de relation stricte entre les inégalités sociales qui existent dans un pays (qui se traduisent par des inégalités de revenus et de ressources socio-culturelles entre les familles), et les inégalités sociales de performance scolaire (Dubet *et al.*, 2010, *op.cit.*). Tous les cas de figure s'avèrent possibles. Certes, certains pays sont caractérisés, comme on s'y attendrait spontanément, à la fois par de fortes inégalités de revenus entre adultes (et donc entre familles) et par de fortes inégalités scolaires (c'est par exemple le cas en Nouvelle Zélande), alors que d'autres sont dans la situation inverse en combinant de faibles inégalités scolaires et sociales (ce sont essentiellement les pays scandinaves). Mais, moins attendu *a priori*, certains pays combinent de fortes inégalités socio-économiques (mesurées par des indices de répartition de revenus) et de relativement faibles inégalités scolaires (l'Italie, le Japon...), alors que, à l'inverse certains pays associent des inégalités socio-économiques relativement modérées et de relativement fortes inégalités scolaires (l'Allemagne jusqu'à une date récente, la France..).

Ces constats suggèrent que l'école peut accentuer ou, au contraire, atténuer l'ampleur des effets des inégalités sociales sur les inégalités scolaires, concrètement de permettre aux enfants a priori les plus défavorisés d'atteindre quand même un certain niveau de performance ou au contraire d'avantager sans s'en rendre compte les plus favorisés. En arrière-plan de cette diversité entre pays que mettent en lumière les comparaisons, et au-delà des facteurs strictement scolaires, ce sont les modalités de la stratification sociale et du fonctionnement du marché du travail qui sont également en jeu : y-a-t-il des emplois dont personne ne veut pour ses enfants, comment est régi l'accès aux emplois les plus qualifiés, existe-t-il des processus de certification continue des qualifications internes aux entreprises, à même de remédier à une moindre formation initiale, etc... ?). Quoiqu'il en soit, ces observations, qui invitent à explorer des mécanismes macrosociaux complexes sous-jacents aux inégalités sociales scolaires, qu'ils soient de nature socio-économique ou culturelles, constituent une veine des plus heuristiques pour les chercheurs.

Elles sont évidemment aussi extrêmement stimulantes d'un point de vue politique, dès lors que les pays savent se montrer sincèrement curieux de ce que font les autres, et avec quels résultats. Cependant, il faut souligner que les comparaisons entre pays ne peuvent servir de guide à l'action publique uniquement si elles permettent d'identifier des relations de nature causale, ce qui est loin d'être immédiat avec les résultats des enquêtes PISA.

4. Quelles sont les limites et les dérives possibles des enquêtes comme PISA ?

Alors que les enquêtes PISA exposent de manière récurrente à un glissement du constat à la prescription (comme en attestent les titres des ouvrages de l'OCDE qui décrivent les résultats de PISA : « *What makes a school successful ?* »), leur utilisation pose à la fois des problèmes variés qui ne sont jamais purement techniques.

Tout d'abord, comme toutes données d'enquêtes, celles-ci sont inévitablement imparfaites. Un premier point est que les résultats des enquêtes PISA sont construits sur la base d'épreuves et sur des échantillons particuliers. Concernant les échantillons d'élèves pris en compte, il faut prendre garde au fait qu'ils ne sont pas toujours comparables d'un pays à l'autre, dès lors que les élèves encore scolarisés à 15 ans dans les pays les plus pauvres comme le Brésil ou le Mexique ne constituent qu'une frange, sur-sélectionnée, de la classe d'âge (tant que le taux de scolarisation à 15 ans n'égale pas 100%), une sur-sélection susceptible de se traduire par des performances supérieures à ce qu'on observe quand tous les enfants, même les plus faibles, sont scolarisés. Quant à l'échantillon de pays intégrés dans l'enquête, et même s'il s'agit de pays parmi les moins pauvres de la planète, il reste marqué par une grande hétérogénéité (des Etats-Unis à l'Albanie ou au Qatar), ce qui fragilise le sens de toutes les valeurs moyennes. De plus, certains pays typés « tirent » parfois à eux seuls les chiffres et les corrélations : par exemple, le fait que dans les pays asiatiques, on ait à la fois des classes très nombreuses et des performances très élevées, engendre statistiquement une corrélation globale positive entre taille des classes et performance, une corrélation qui cesse d'être significative quand on retire ces pays des calculs.

Il faut aussi se souvenir que comme toute enquête sur échantillon, les estimations sont assorties d'une marge d'incertitude ; alors que les classements sont volontiers médiatisés, la position d'un pays n'est souvent pas différente significativement d'un nombre assez élevé de ses voisins. L'oublier conduit à surestimer les écarts entre pays. Les seuls écarts de performance vraiment importants se situent en fait entre les pays riches et les pays pauvres.

Rappelons qu'un autre point à ne pas négliger dans la conduite des comparaisons est le fait qu'on observe dans PISA les élèves âgés de 15 ans quel que soit leur niveau scolaire. Or, en France, où l'on compte à 15 ans un nombre (certes de moins en moins) important d'élèves en retard, et dès lors que ces élèves sont dans des filières moins exigeantes scolairement (en 3^{ème} dans leur majorité), cela crée un handicap structurel par rapport aux pays, majoritaires, où le redoublement n'existe pas ; d'où l'intérêt d'examiner spécifiquement le score des élèves n'ayant jamais redoublé, en général très bien situés dans le classement.

Concernant les épreuves, des réserves ont été émises sur les possibles « biais culturels », provenant du fait que les items sont conçus initialement en anglais et traduits ensuite, ou du format de type QCM de nombreux exercices. Néanmoins, les meilleurs élèves viennent d'aires culturelles très différentes (Finlande, Canada, Corée), et au total, quant à la construction et à la traduction des épreuves, PISA donne des assurances « techniques » assez importantes. Il reste certes des interrogations quant à savoir ce que révèlent précisément les épreuves sur les compétences des élèves : la psychométrie et la docimologie sont là pour rappeler combien il est difficile, voire impossible, d'évaluer des compétences stables sans interférence d'éléments de contexte ; l'univers de référence mobilisé par exemple n'est pas forcément identique pour tous (et conforme à ce qu'imaginent les concepteurs des tests) : va-t-on par exemple comprendre qu'il faut analyser grammaticalement une question ou penser, en rabattant le texte sur son contenu, qu'il faut donner son opinion ? Pour certains élèves, peu à l'aise avec certains termes, voilà qui peut constituer une réelle difficulté (Bautier *et al.*, 2006). L'analyse pédagogique détaillée des items peut à cet égard s'avérer aussi instructive que l'examen de scores moyens.

Mais de fait, pour en revenir aux politiques, la difficulté majeure concerne les limites des données transversales, collectées à un moment donné, quand il s'agit d'évaluer et de piloter les systèmes éducatifs. Car les liens observés entre résultats des élèves et politiques éducatives ne sont instructifs à cet égard que si l'on est raisonnablement sûr qu'on a bien affaire à des relations causales. Or d'une part les résultats prennent souvent la forme de corrélations entre deux données -une performance, une caractéristique du système-, ce qui peut évidemment masquer nombre de relations non prises en compte ; par exemple, on croirait appréhender l'impact du salaire des enseignants, alors que ce serait en fait les modalités de leur formation ou de leur recrutement qui serait en jeu. D'autre part, le caractère ponctuel des observations ne permet pas de trancher quant à l'impact spécifique de tel aspect du système éducatif, puisque toute la « fabrication » passée des diverses corrélations est ignorée et qu'à 15 ans, on enregistre nécessairement et globalement le résultat de toute la socialisation et la scolarisation antérieures. Ainsi, la meilleure performance des élèves de 15 ans d'un pays donné ne peut strictement s'expliquer par l'organisation pédagogique au niveau du premier cycle du secondaire, car ces élèves avaient déjà peut-être un meilleur niveau dès les premiers degrés de l'école élémentaire.

Notons que cette faiblesse inhérente aux données PISA n'est pas spécifique aux approches quantitatives: les approches qualitatives d'un nombre restreint de pays, qui analysent de manière fouillée des modes de fonctionnement, en les situant dans leur contexte historique et social, ne disent en rien si les quelques aspects retenus (puisque une perspective qualitative choisit de comparer quelques cas sur quelques points) ont quelque pertinence « externe », c'est-à-dire affectent effectivement les acquis ou les attitudes des élèves ; on peut par exemple pointer qu'entre le pays A et le pays B, les salaires des maîtres sont plus importants ou l'esthétique des bâtiments scolaires est plus satisfaisante, mais il reste à prouver si de manière générale, sur un nombre cette fois plus élevé de pays, ces caractéristiques semblent avoir un impact sur ce que l'on cherche à expliquer, les résultats des élèves. De plus, les comparaisons quelles qu'elles soient se fondent souvent sur les données (comparatives) disponibles, qui ne sont pas forcément très abondantes, ou encore sur ce que l'on croit a priori important de prendre en compte, ce qui n'est pas mieux !

Dans tous les cas, de multiples dérapages sont possibles, les corrélations étant abusivement interprétées selon les préférences idéologiques de chacun. Prenons l'exemple de l'interprétation des bons résultats finlandais. Si l'on est syndicaliste, on les met en relation –avec la tentation de dire qu'on les « explique »- avec le niveau élevé des rémunérations et du prestige des enseignants ; si on croit à la pédagogie, on les mettra plutôt en relation avec le suivi individualisé dont bénéficient en temps réel les enfants en échec ; si on a un *a priori* favorable à l'autonomie des établissements (comme c'est souvent le cas à l'OCDE), on soulignera que les écoles finlandaises ont un niveau d'autonomie plus élevé que les autres pays, en oubliant de préciser que ceci ne concerne que la politique d'évaluations des élèves⁶. Et ainsi de suite...

Or, il est plus que probable que les divers aspects des systèmes éducatifs jouent de concert, d'où d'ailleurs le caractère très critiquable des « emprunts » d'éléments isolés de tel ou tel système. En outre, la prise en compte du contexte historique et culturel serait nécessaire et c'est peu dire qu'elle est rarement intégrée dans l'analyse.

⁶ Pour ce qui est des autres dimensions de l'autonomie –recrutement des enseignants, budget, admission des élèves-, les écoles finlandaises sont souvent moins autonomes que la moyenne de l'OCDE.

Reprenons plus en détail l'exemple du Japon, un pays parmi les plus efficaces dans les enquêtes PISA et, plus nettement encore, les plus équitables. Un premier point à souligner concerne la valorisation spécifique dont l'éducation est l'objet dans ce pays, à la fois parce qu'apprendre a toujours joui d'une forte valorisation dans la société japonaise en raison de la tradition confucéenne, et aussi parce qu'historiquement la priorité a été donnée au système scolaire, dès le début du 20^{ème} siècle, pour conforter l'unification idéologique du pays, sa modernisation, son indépendance et sa compétitivité. S'ajoute à ces éléments de contexte l'opinion répandue au Japon selon laquelle tous les enfants sont capables d'apprendre et que les capacités innées comptent peu dans la réussite. Les élèves japonais sont beaucoup plus nombreux que les élèves américains à attribuer leurs performances scolaires à leurs efforts plutôt qu'à leurs aptitudes. Au quotidien, tant les pratiques des maîtres que les relations entre élèves sont marquées par cette conviction, et l'on croit fermement aux pouvoirs de la motivation. Mais il ne faut pas négliger le fait que les élèves fréquentent souvent deux écoles : l'école publique qui les éduque de façon assez « égalisante », et les cours privés qui les placent dans une compétition précoce. Il n'est pas exclu que l'excellence des performances des élèves japonais mesurée par PISA tienne à cette école parallèle, même si d'autres pistes culturelles ont été explorées, comme le rôle de facteurs linguistiques (Hee Yean, 2008, *art.cit*).

Les politiques doivent-ils donc courir après ce « modèle » asiatique, si peu comparable au nôtre à tant d'égards ? Examiner ce qui se passe dans un pays comme le Canada (toujours bien classé dans les enquêtes PISA, notamment moins inégalitaire alors même que la société canadienne est loin d'être égalitaire) serait certainement plus pertinent ; c'est notamment un pays où, face à des élèves d'horizon aussi varié que les nôtres, on croit bien plus que nous à la pédagogie. Au moins les pays asiatiques (et les autres « premiers de la classe » comme la Finlande) ont-ils le mérite de montrer qu'il n'y a pas d'arbitrage entre efficacité et équité : les pays dont les élèves sont les meilleurs sont aussi ceux qui réduisent le plus les inégalités. Or dans notre pays, au gré des alternances politiques, on semble balancer entre quête de l'équité et quête de l'excellence –la maternelle pour tous et l'éducation prioritaire, *versus* les Grandes Ecoles et les classes bi-langues-, comme s'il fallait choisir...

Une autre difficulté d'interprétation (qui peut entraîner des choix politiques discutables) concerne la légitimité de l'inférence des constats faits au niveau des pays à des interprétations portant sur les individus, ou vice-versa. C'est une difficulté classique de ce qu'on appelle l'analyse écologique, quand on examine divers indicateurs portant sur des « unités écologiques », qui sont le plus souvent des pays. Ces indicateurs qui entendent caractériser les pays peuvent être construits soit par agrégation de données individuelles (moyenne des attitudes des membres d'un pays), soit refléter la distribution statistique, dans le pays, de caractéristiques individuelles (indicateur d'inégalités) ; ils peuvent aussi n'avoir aucune contrepartie au niveau des individus et exprimer une caractéristique du pays en tant que tel (ici, tel ou tel trait du système éducatif).

Une difficulté de ce type d'analyse est que l'on est sans cesse tenté de passer d'un niveau à l'autre. Ainsi, pour évoquer des constats assez classiques, parce que les individus mieux formés sont en moyenne moins souvent des criminels, on en déduirait que les pays instruits doivent avoir un niveau moyen de criminalité plus faible. Ceci s'avère faux, notamment parce que la criminalité, dans un pays, dépend à l'évidence de bien d'autres facteurs que l'éducation. Ce type d'interrogation est particulièrement pertinent quand on s'interroge sur les effets de l'éducation, qui peuvent faire l'objet d'appréciations divergentes au niveau des personnes et au niveau du pays. Par exemple, la plupart des gens penseront qu'il est bon que le niveau d'éducation des parents soit associé à des aspirations plus élevées pour leurs enfants, mais au niveau collectif on pourra craindre que l'effet produit par une

élévation du niveau d'éducation des parents soit une compétition scolaire et sociale encore plus dure. Ou encore, de fortes relations entre formations et emplois peuvent être considérées comme un gage d'efficacité au niveau d'un pays, tout en étant très délétères pour certains individus (les moins diplômés, qui s'en trouvent durablement exclus du marché du travail). Ceci invite à garder sans cesse à l'esprit que les effets collectifs et les effets individuels de l'éducation doivent être soigneusement distingués.

Il est donc toujours risqué de déduire des corrélations observées à un niveau global des relations, *a fortiori* des relations causales, au niveau des individus. Cela s'explique par la multiplicité de facteurs potentiellement en cause en-deçà de ces corrélations, et c'est d'autant plus vrai que l'on travaille sur des données instantanées : il est alors impossible de dégager des séquences temporelles entre les variables et donc d'identifier précisément des évolutions causales (la variation de tel ou tel paramètre ayant eu tel ou tel effet), ce qui rend complètement inadéquate la notion d'efficacité, souvent invoquée dans les commentaires sur les résultats de ces enquêtes. Ceci doit également rendre pour le moins prudent chaque fois que l'on fait des pronostics quant aux effets des politiques ; par exemple quand, observant que les plus diplômés s'insèrent mieux, on compte sur l'élévation du niveau d'instruction pour résorber le chômage: on néglige alors les dynamiques temporelles, le fait que les relations observées à l'instant t –la meilleure insertion des diplômés- vont forcément bouger si les « quantités » bougent –si ces diplômés en viennent à constituer l'immense majorité des jeunes cherchant à s'insérer...

Conclusion

On ne peut contester l'intérêt et le grand pouvoir heuristique des enquêtes PISA, comme des enquêtes comparatives internationales de manière plus générale. Mais il faut dans le même temps admettre qu'on ne saurait « caler » la politique éducative sur ces enquêtes. Tout d'abord parce qu'un pilotage s'appuyant sur des corrélations extraites de leur contexte s'exposerait à des erreurs de préconisation ; on défendrait, par exemple, les classes nombreuses à l'asiatique ! Mais aussi parce qu'il convient de veiller à ne pas entériner sans discussion les choix éducatifs qui sous-tendent les enquêtes PISA, choix effectués de fait par des experts et/ou les instances européennes et les organisations internationales, mais qui ne font pas nécessairement l'objet d'une politique explicite et *a fortiori* de débats démocratiques. Ici comme ailleurs, privilégier certains indicateurs équivaut à privilégier certains objectifs, puisqu'on évalue les systèmes éducatifs à l'aune de ces indicateurs. Ainsi, la définition des « compétences pour la vie » incorporée dans PISA fait-elle l'objet d'un consensus, ou encore, pourquoi ne pas donner plus d'importance à des indicateurs sur le bien-être des élèves ? Pour ce qui est strictement des performances scolaires, on peut également poser la question de savoir si l'excellence d'un système se mesure à l'aune du niveau moyen des performances, et/ou à l'aune du niveau de performance des meilleurs, ou au contraire du niveau auquel on a fait accéder les plus faibles, ou encore à l'homogénéité des performances d'une classe d'âge...

De plus, il serait légitime de distinguer efficacité interne et efficacité externe des systèmes éducatif : un système performant est-il un système dont les élèves réalisent des acquis élevés ou qui se placent aisément, et s'intègrent bien dans la vie ? Rappelons à cet égard que ces élèves japonais ou coréens si performants ne semblent pas s'intégrer si facilement dans la vie adulte, si l'on en croit les taux de suicide ou de « NEET » observés dans ces

pays⁷. Et à cet égard, si l'on prend en compte cette fonction capitale de l'école qu'est l'allocation aux emplois, il faut poser la question du moment où il est légitime d'évaluer ce que « produit » l'école : à 15 ans, ou 10 ans après l'entrée dans l'âge adulte ? C'est alors peut-être plus la profession et la position sociale que l'on a obtenues que le volume des acquis scolaires (qui plus est, à 15 ans) qui importe. Les classements qui découlent des mesures d'acquis à 15 ans méritent donc d'être fortement relativisés : il semble bien que l'école inégalitaire à l'allemande se « rattrape » ensuite avec une entrée dans la vie active plus fluide et moins inégalitaire, ou à l'inverse qu'une école plus égalitaire comme dans nombre de pays de l'Europe du Sud est suivie de fait de grosses difficultés d'insertion (van de Werfhorst et Mijs, 2010).

Enfin, les critères macrosociaux censés justifier *in fine* les politiques éducatives (notamment l'élévation du niveau de formation) mériteraient tout autant d'être l'objet de débats, qu'il s'agisse –OCDE oblige– de la compétitivité ou de la croissance économique⁸, ou d'une cohésion sociale dont la définition et la mesure n'ont rien d'évident

Un risque, et non des moindres, engendré par les enquêtes PISA serait d'étouffer tous ces débats par leur simplicité et leur séduction trompeuses, et ce d'autant plus que règne une relative incertitude quant aux objectifs éducatifs intrinsèques qui valent d'être poursuivis. Oubliant les finalités, leurs propres finalités, les pays risquent de définir le progrès, en matière éducative, comme le fait d'améliorer sa place dans le palmarès, dérive potentielle qui vaut d'ailleurs pour toute évaluation... Mais de toute évidence, il ne faut pas jeter le bébé avec l'eau du bain et simplement, si l'on peut dire, se souvenir du vieil adage, comparaison n'est pas raison.

Marie Duru-Bellat (IREDU, Pôle AAFE, Esplanade Erasme, Université de Bourgogne-Franche-Comté, B.P. 26513, 21065 Dijon cedex, 0380395450; marie.duru-bellat@wanadoo.fr).

Bibliographie

Bautier, E., Crinon, J., Rayou, P., Rochex, J-Y. (2006). Performances en littératie, modes de faire et univers mobilisés par les élèves : analyses secondaires de l'enquête PISA 2000, 157, 85-101.

Duru-Bellat, M. (2015). Les compétences non académiques en question. *Formation Emploi*, n°130, 13-30.

⁷ Les « NEET » sont une nouvelle catégorie internationale regroupant les jeunes « ni en études, ni en emploi, ni en formation », qui vivent le plus souvent isolés et sans activité chez leurs parents.

⁸ L'OCDE ou « The Lisbon council », en charge de la promotion des objectifs de Lisbonne, publie régulièrement des textes soulignant que l'Europe risque d'être déclassée au cours du vingt-et-unième siècle dans la compétition économique mondiale si elle ne fait pas de l'éducation sa priorité ; cf. par exemple The Lisbon Council, « The Evidence is Clear : Education Pays » (13/03/2006).

Duru-Bellat, M. (2019). *Le mérite contre la justice*, Paris, Presses de sciences Po.

Dubet, F., Duru-Bellat, M., & Vérétoit, A. (2010) *Les sociétés et leur école. Emprise du diplôme et cohésion sociale*, Paris, Seuil.

Hee Yean C. *et al.*(2008). D'où vient la supériorité des enfants coréens en mathématiques ? *Carrefours de l'éducation*, n° 26, 185-200.

Le Donne, N. (2014). European Variations in Socioeconomic Inequalities in Students' Cognitive Achievement : The Role of Educational Policies, *European Sociological Review*, 30, n°3, 329-343.

Lafontaine, D. et Baye, A. (2012). PISA, instrument ou témoin du changement ? *Education Comparée*, vol.7,59-102.

Le Mener, M. et Meuret,D. (2017). L'accroissement de l'effet de l'origine sociale : par où est-il passé ? *Revue Française de Sociologie*, 2, 207-231.

Merle, P. (2012). *La ségrégation scolaire*, Paris, La Découverte.

Meuret, D. et Morlaix S. (2006). Origine sociale et performance scolaires, *Revue Française de Sociologie*, 1, 49-79.

Rey,O. (2011). PISA : ce que l'on en sait et ce que l'on en fait. *Dossier d'actualité*, l'IFE (institut français d'éducation, n°66.

Van de Werfhorst, H. et Mijs J. (2010). Achievement inequality and the institutional structure of educational systems. *Annual Review of Sociology*, 36, 407-428.