

« L'évaluation quantitative internationale des systèmes éducatifs : apports, limites et renouvellement. »

**Nathalie Mons
(IREDU-CNRS/Université de Bourgogne)**

« We cannot wander at pleasure among the educational systems of the world like a child strolling through a garden and pick off a flower from one bush and some leaves from another, and then expect that if we stick what we have gathered into the soil at home, we shall have a living plant. » C'est par cette métaphore que Sadler (1900) plante l'opposition qui existe traditionnellement entre les deux écoles de l'éducation comparée, le particularisme et l'universalisme.

Pour les deux écoles, la comparaison a des sens et des objectifs différents.

D'un côté, le particularisme (Sadler (1930), Mallison (1975)...) en se focalisant sur l'étude des particularités de chacun des systèmes éducatifs, utilise la comparaison pour saisir le caractère original de chaque système éducatif. De l'autre, le positivisme (Jullien (1817), Noah et Eckstein (1969), Farrell (1979), Le Thanh Khoi (1986)...) cherche à déterminer des relations invariantes, des lois générales, qui transcendent les frontières et utilise donc les comparaisons dans un but de généralisation.

Ancrées dans cette seconde famille, les évaluations quantitatives dont les grandes enquêtes internationales sont emblématiques (IEA, OCDE-PISA, IEAP...) se sont imposées comme un lien entre recherche et politique éducative.

Après avoir, dans une première partie, présenté l'ancrage philosophique et le développement historique des évaluations internationales quantitatives, nous présenterons dans un second temps leurs apports et leurs limites. Une troisième partie sera consacrée aux nécessités de leur renouvellement, avec des évolutions nécessaires tant en amont lors de la conception des enquêtes, qu'en aval lors de leur lecture et de leur exploitation.

I. Ancrage philosophique et développement historique des évaluations quantitatives internationales des systèmes éducatifs

Traditionnellement, l'éducation comparée oppose deux conceptions de la comparaison : le particularisme et le positivisme

Le particularisme, en se focalisant sur l'étude des traits originaux de chacun des systèmes éducatifs, utilise la comparaison pour saisir le caractère national et unique des organisations scolaires. Dans l'étude descriptive de chacun des systèmes scolaires, l'accent est mis sur les contextes philosophiques, historiques, sociaux et économiques dans lesquels ils sont ancrés ainsi que sur les traditions (pédagogiques, administratives...) qui sont propres à l'organisation scolaire. Autant que l'analyse de l'école elle-même, le mouvement privilégie donc l'étude des « things outside the school » (Crossley, 1999). En France, la « méthode sociétale » (Maurice et al., 1982) met également l'accent sur les cohérences globales spécifiques au contexte de chaque pays et met en garde contre la difficulté d'opérer des comparaisons partielles de systèmes éducatifs.

Le positivisme a une vision radicalement différente de la comparaison. Si le particularisme met en avant l'unicité de chaque système, le mouvement universaliste, au fondement de l'éducation comparée avec les écrits de Jullien (1817), cherche à déterminer des relations invariantes, au-delà du temps et de l'espace, des lois générales qui transcendent les frontières et utilise donc les comparaisons dans un but de généralisation. D'après Noah et Eckstein (1969), « l'éducation comparée est une tentative d'utiliser des données internationales pour

tester des propositions sur les relations qui peuvent exister entre éducation et société ainsi qu'entre pratiques pédagogiques et résultats scolaires.»

L'ensemble des positivistes se retrouvent sur un certain nombre de propositions (Farrell, (1979), Le Thanh Khoi (1986)...): la comparaison entre les systèmes éducatifs doit être l'ultime forme d'analyse de l'école, seules des études empiriques sont scientifiques, et seules des affirmations scientifiques ont de l'intérêt. Les positivistes affirment ainsi passer grâce à leurs méthodes comparatistes de la description à l'analyse des systèmes, d'un contenu factuel à des relations généralisables. Aux affirmations péremptoires telles que « réduire la taille des classes améliore les apprentissages des élèves », « le niveau scolaire progresserait si l'on dépensait davantage » ou encore « les structures décentralisées sont bénéfiques en terme d'apprentissage », les positivistes préfèrent des validations statistiques opérées en collectant un ensemble de données réelles sur les productions des systèmes éducatifs, les caractéristiques des élèves et les conditions d'enseignement.

Les enquêtes internationales quantitatives s'inscrivent dans le mouvement universaliste de l'éducation comparée

Les grandes enquêtes internationales quantitatives s'inscrivent dans le mouvement positiviste de l'éducation comparée. S'appuyant sur la collecte de statistiques internationales, qui s'est développée aux lendemains de la seconde guerre mondiale grâce à la création des organismes internationaux, elles visent à mettre en relation, d'un côté, les performances des systèmes éducatifs et, de l'autre, les structures qui les caractérisent, les caractéristiques des élèves qu'ils accueillent et les ressources mises à contribution (ressources financières, humaines, pédagogiques...). Objectif : déterminer, au-delà des caractéristiques propres aux élèves, les facteurs scolaires qui agissent positivement sur le développement et la qualité des systèmes éducatifs de façon à dessiner, à terme, grâce à un processus d'accumulation des connaissances, un système éducatif « idéal ». Comme nous le verrons par la suite, pour l'instant, ce système « idéal » a surtout été défini comme coût-efficace.

Les performances scolaires analysées vont prendre des formes multiples. Aux lendemains de la seconde guerre mondiale, les indicateurs utilisés sont rustiques. La production des systèmes éducatifs s'analyse alors à travers : les taux de scolarisation (on examine les couvertures scolaires atteintes par les pays), les taux de transition entre les différents niveaux scolaires (on observe quel pourcentage d'une cohorte réussit à se frayer un chemin dans les systèmes éducatifs), les taux d'abandons (on se focalise sur les capacités des systèmes à garder les élèves) ou les taux de redoublement (on s'interroge alors sur les capacités des systèmes à amener les élèves à un niveau scolaire en un certain nombre d'années).

De l'autre côté de cette équation « ressources-résultats », outre les caractéristiques des élèves, sont précisés les facteurs dont on pense qu'ils agissent sur les performances académiques : les caractéristiques des enseignants (formation, âge, sexe, années d'expérience, salaires...), la taille des classes ou des établissements, la mise à disposition de matériaux pédagogiques, les structures institutionnelles, les dépenses éducatives, le volume d'heures d'enseignement... Les facteurs explicatifs sont multiples.

A partir des années 60, l'évaluation de la qualité des apprentissages donne un nouvel essor aux enquêtes quantitatives internationales

Les années 60 marquent un tournant important dans la conception des enquêtes internationales. L'analyse des performances des systèmes éducatifs, telle qu'elle a été pratiquée jusque là, semble insuffisante. Les instruments utilisés apparaissent limités. Certes, un système éducatif peut scolariser une majorité de ses élèves jusque dans le secondaire, mais la qualité des apprentissages tout autant que le nombre d'années moyen de scolarisation devient une donnée importante. Au quantitatif, on oppose un qualitatif jusqu'alors jamais mesuré.

C'est le défi que se donne un petit groupe de spécialistes en évaluation réunis en 1958 à Hambourg dans les locaux de l'UIE (Unesco Institute of Education) : évaluer la qualité des apprentissages des élèves dans différents pays à travers un test de connaissances international unique (Husen et Postlethwaite, 1996). Alors que les Russes s'octroient la première place dans la course à la conquête de l'espace en lançant Spoutnik, les

Américains, ulcérés, en proie au doute, décident de financer ce projet pilote capable de leur dire si réellement leur système éducatif, perçu comme un élément capital de la nouvelle compétition technologique, est vraiment le fautif. Sont-ils relégués au second rang parce que le niveau des élèves américains en sciences et en mathématiques est devenu trop faible ?

Mais établir un palmarès international des systèmes éducatifs n'est pas le principal objectif de ce petit groupe de spécialistes en évaluation. Les futurs fondateurs de l'IEA (International Association for the Evaluation of Educational Achievement) visent, surtout, grâce à ce « laboratoire mondial géant » qui offre une grande variété de situations éducatives (Husen, 1987) à se doter de moyens jusqu'alors jamais mis en œuvre pour analyser les facteurs explicatifs des performances scolaires. Aux côtés des tests pratiqués auprès des élèves, des données sont recueillies sur les conditions d'apprentissage et les caractéristiques des élèves. Une attention tout particulière est portée aux programmes scolaires (leurs contenus et la réalité de leur mise en œuvre).

Vont ainsi naître toutes les grandes enquêtes de l'IEA. Plus d'une quinzaine d'études, couvrant principalement les champs disciplinaires des mathématiques, des sciences et la maîtrise de la langue maternelle, ont vu participer, en moyenne, une vingtaine de pays (cf annexe n°1, sur la description des principales évaluations quantitatives internationales).

Dans la lignée de l'IEA, les grandes études internationales se développent, à partir des années 80. Les deux enquêtes de l'IAEP (International Assessment of Educational Progress) sont lancées, dans les disciplines scientifiques, par un service américain avec des moyens nettement inférieurs. L'OCDE, qui envisage dès les années 80 une enquête concurrente de celles de l'IEA, ne lancera finalement le programme PISA (Programme for International Student Achievement) qu'en 2000. Il couvre une quarantaine de pays et vise à tester, tous les trois ans, les compétences des élèves de 15 ans en mathématiques, en langue maternelle et en sciences.

Les pays développés sont historiquement sur-représentés dans les enquêtes d'évaluation. Il faut attendre les années 90 pour que les pays en voie de développement conçoivent leurs propres études d'évaluation, souvent au niveau régional. Leur priorité est davantage orientée vers le primaire. Ainsi, à partir de 1994, dix-huit pays d'Amérique Latine se regroupent dans le projet LLECE (Laboratorio Latinoamericano de Evaluacion de la Calidad de la Educacion) pour tester leurs élèves en 3^{ème} et 4^{ème} année d'étude. Dans les pays sub-sahariens francophones, le programme PASEC (le Programme d'Analyse des Systèmes Educatifs) permet l'évaluation des élèves en seconde et cinquième années d'études, en mathématiques et en français. Les quinze pays africains, réunis dans le réseau SACMEQ (Le Groupement d'Evaluation de la qualité de l'Enseignement en Afrique Australe) mettent en place en 1995 et 1999 deux évaluations en mathématiques et en lecture pour les enfants en sixième année d'étude.

Selon Kellaghan (1996), un ensemble de facteurs expliquent ce développement accéléré des évaluations quantitatives internationales, à partir des années 80. Alors que la plupart des pays développés ont réussi une démocratisation de leur enseignement secondaire, se pose de façon plus cruciale la question de l'amélioration de la qualité de l'enseignement. Si de simples indicateurs descriptifs, tels que les taux de couverture ou d'abandons, permettaient jusqu'alors de piloter les évolutions des systèmes éducatifs, on s'interroge désormais davantage sur l'organisation scolaire permettant de maximiser les apprentissages des élèves. Les pays en voie de développement, à partir de la conférence de Jomtien, en 1990, prennent conscience de façon accrue du fait que le double combat de l'accès et de la qualité de l'éducation doit être mené simultanément. La progression des dépenses d'éducation - globales et unitaires - dans les pays industrialisés et les contraintes budgétaires fortes dans les pays en voie de développement poussent également les gestionnaires à s'interroger sur un meilleur rapport coût-efficacité. Enfin, dans la lignée de la thèse du capital humain (Schultz, 1971), le système éducatif et sa qualité sont désormais perçus comme un facteur fondamental de la compétition économique entre les pays.

II. Apports et limites des enquêtes quantitatives internationales

Des apports importants malgré des conclusions encore limitées

Les recherches quantitatives internationales, menées par les organismes spécialisés dans l'évaluation (IEA, SACMEQ, LLECE...) et les grandes organisations internationales (Banque mondiale, UNESCO, OCDE...) ont permis des avancées dans de multiples domaines.

Le benchmarking international

Ces études internationales ont, tout d'abord, conduit à un renouvellement de la conception de la qualité de l'enseignement. Alors que jusqu'alors les recherches se focalisaient sur la qualité des processus et des ressources mises à la disposition des systèmes éducatifs, pour la première fois les efforts se concentrent sur l'analyse des résultats. On s'intéresse à la qualité de ce que produit l'organisation scolaire et non, simplement, à celles de ses ressources. Ces recherches quantitatives imposent également des critères internationaux d'évaluation à l'aune desquels les pays peuvent se confronter. La plupart des études répartissent ainsi les élèves dans plusieurs niveaux de performances. Par exemple, dans PISA-OCDE, les élèves sont distribués dans cinq niveaux. Les performances inférieures au niveau 1 indiquent « la présence de déficiences graves dans la capacité des élèves à utiliser la lecture comme un outil d'acquisition de connaissances et de compétences » (OCDE, 2001). Tout autant que leur score moyen, les pays doivent donc scruter avec attention cet indicateur qui montre la capacité (ou l'incapacité) du système éducatif à transmettre aux élèves le socle minimum de compétences permettant l'intégration future dans la société.

Une connaissance large de multiples systèmes éducatifs

Ces grandes enquêtes apportent, de plus, une connaissance large de nombreux systèmes éducatifs. Au-delà des performances des élèves, sont collectées des informations sur les contenus et les conditions d'apprentissage. Alors que les études qualitatives permettent la comparaison détaillée d'une poignée de pays, les questionnaires distribués aux élèves, aux enseignants et aux chefs d'établissement ont conduit, dans certaines études (TIMSS, PIRLS, PISA, Laboratorio...) à la réalisation d'état des lieux dans une quarantaine de pays sur de nombreuses thématiques : conditions d'apprentissage en mathématiques ou en lecture, contenu des programmes dans différentes disciplines, réalité de l'autonomie des établissements, réalité de la violence dans les écoles, niveau réel de formation, âge et sexe des enseignants, conditions d'admission des élèves dans les écoles... La liste est longue des données relatives aux contextes scolaires recueillies dans le cadre des recherches quantitatives internationales.

L'influence relative des facteurs scolaires et individuels et les potentialités des politiques éducatives

Alors que les sociologies locales prennent ces éléments pour donnés, les évaluations internationales ont montré que les dispersions de résultats entre les élèves au sein des pays ainsi que la force de la relation entre performances académiques et caractéristiques socio-économiques des élèves varient fortement suivant les pays. Elles ont ainsi montré la relativité de l'influence des facteurs scolaires et non scolaires, et corollaire, les potentialités des politiques éducatives. Contrairement aux affirmations du rapport Coleman (1966), l'école peut donc faire la différence. Mais dans certains pays, et suivant les politiques éducatives menées, elle aggrave et dans d'autres elle corrige les inégalités sociales. Ainsi, le programme PISA (OCDE, 2001) a-t-il permis d'opposer, d'un côté, le modèle japonais et coréen caractérisé, en fin de secondaire inférieur, par une population scolaire homogène et une relation limitée entre les caractéristiques individuelles des élèves et les résultats scolaires et, de l'autre, le modèle germanique qui présente une grande dispersion de performances entre les élèves ainsi qu'une influence forte des données personnelles des élèves sur les résultats académiques. Les enquêtes quantitatives internationales montrent que les inégalités scolaires ne sont pas une fatalité.

La valeur heuristique de la diversité des situations internationales

Autre atout des enquêtes internationales, par rapport cette fois-ci à des comparaisons intra-systèmes : grâce à la diversité des situations éducatives qu'elles apportent (dans le cadre des études à couverture internationale plutôt que régionale), elles ont permis de mettre en évidence des relations qui n'avaient pas été décelées jusqu'alors dans des études comparatives intra-systèmes. Ainsi, sur le débat houleux de l'« impact » des dépenses éducatives sur les résultats des élèves, l'étude PISA, grâce à son panel diversifié de pays industrialisés et

émergents, a permis d'apporter de nouveaux éléments. Alors que des études intra-systèmes, principalement menées aux Etats-Unis, n'avaient toute liaison entre le niveau des dépenses en éducation et le niveau scolaire des élèves, PISA a démontré, grâce à un indicateur de dépenses cumulées, que l'investissement dans l'éducation avait bien une influence sur les apprentissages. Mais cette relation n'est pas linéaire : passé un certain niveau de dépenses, un accroissement des investissements n'est plus associé à une progression des résultats scolaires. Seule une étude internationale mixant des pays aux profils variés pouvait permettre d'aboutir à ce résultat.

Les conclusions encore partielles des enquêtes internationales quantitatives

Si globalement donc les avancées ne sont pas négligeables, les conclusions sont encore à ce jour en retrait par rapport aux attentes. En effet, sur de nombreux points les conclusions des différentes études divergent trop pour être considérées comme définitives.

On peut cependant aujourd'hui, très schématiquement, avancer quelques points sur lesquels les études concordent (Keeves (1995), Lambin (1995), UNESCO (2000), Psacharopoulos (1990), Heyneman et Loxley (1983)).

Deux grandes familles de facteurs, liés pour les uns aux caractéristiques des élèves (milieu socio-économique, taille de la famille, lieu d'habitation...) et pour les autres à l'organisation scolaire (caractéristiques de l'enseignant, taille des classes, existence de matériels pédagogiques, répartition dans l'organisation des compétences entre les différents niveaux de pouvoirs, volume horaire d'enseignement...), sont associés aux résultats académiques des élèves. Dans les pays en voie de développement, l'influence des facteurs scolaires serait supérieure à ce qu'elle est dans les pays développés, dans ces derniers ce sont surtout les caractéristiques des élèves qui entretiendraient une relation forte avec leur niveau d'apprentissage.

Parmi les facteurs liés à l'école, certains ont un rôle positif sur l'enseignement. Ainsi, augmenter le temps d'instruction d'une matière (mais au détriment de laquelle ?), accroître l'exposition des élèves aux apprentissages (la fameuse « opportunity to learn » des études IEA), distribuer des manuels scolaires (la relation est surtout existante dans les pays en voie de développement et dans les premières années du primaire), présenter des attitudes positives face à l'enseignement, distribuer des devoirs à la maison, allonger le tronc commun d'enseignement (et ne pas multiplier les formes de tracking), mettre en place une politique de promotion automatique, définir une organisation des pouvoirs mêlant centralisation de certaines compétences et décentralisation d'autres (mais quelles organisation précise ?), l'ensemble de ces politiques éducatives serait associé à des performances académiques supérieures.

De nombreuses critiques

Les conclusions limitées ou les résultats divergents des enquêtes quantitatives font, bien sûr, le jeu de leurs détracteurs. Car les critiques à l'encontre de ces recherches sont nombreuses. Elles portent à la fois sur la mesure des performances et des déterminants scolaires ainsi que sur les méthodes et l'exploitation de ces études.

Les critiques à l'encontre des mesures des performances et des déterminants scolaires

Les critiques sont multiples :

- la plupart des enquêtes ne mesureraient pas l'ensemble des dimensions d'une discipline mais seulement les caractéristiques les plus facilement internationalisables (par exemple, Cumming (1996) et son analyse des études de lecture par l'IEA). A l'opposé, quand plusieurs dimensions sont présentes, certains auteurs (Goldstein, 1995) remettent en cause les scores agrégés dans une discipline. Par exemple, PISA divise les compétences de compréhension de l'écrit en trois sous-domaines « retrouver de l'information », « interpréter le texte », « réfléchir sur le contenu du texte », puis les agrège dans une échelle combinée de compréhension de l'écrit. Certaines enquêtes ont préféré, au contraire, ne pas donner de scores agrégés qui ne paraissent pas avoir de sens. C'est le cas de l'enquête de l'IEA en lecture (1992) qui a distingué les scores en lecture narrative et en lecture documentaire.

- l'élaboration des items des enquêtes internationales ferait souvent défaut (par exemple, la technique du QCM est trop souvent utilisée et ne permet qu'une vision parcellaire des performances des élèves).
- les échantillonnages ne seraient pas toujours parfaits. Prais (2003) explique ainsi les différences de résultats, pour l'Angleterre, entre les études PISA et IEA, par le fait que l'échantillon d'élèves anglais de PISA élimine une partie des « mauvaises écoles » et des élèves les plus faibles (il est, cependant, à noter que les pays qui ne suivent pas les procédures décidées par les organisateurs des enquêtes sont mentionnés explicitement dans les résultats, ce fut le cas des Pays-Bas dans PISA, que nous n'avons donc pas repris dans notre recherche).
- Les marges d'erreur peuvent, si elles ne sont pas précisées (ce qui n'est pas le cas dans les enquêtes de l'IEA et PISA) largement entacher les résultats et en particulier les classements.
- Plus globalement, la comparabilité internationale des indicateurs de performances est attaquée. Du fait de fortes différences de contexte, la faisabilité d'évaluations internationales est remise en cause pour plusieurs raisons. La réalisation des tests dans une ou des langues dominantes (anglais le plus souvent, anglais et français dans PISA) pose des problèmes de traduction. Le background culturel des élèves des principaux pays organisateurs du test leur permettrait, de plus, une meilleure appréhension de l'évaluation. Les comportements nationaux diffèreraient également face aux évaluations : tous les élèves de tous les pays ne sont pas habitués aux enquêtes standardisées, leur style de réponse varie (les Français préfèrent ne pas répondre là où les élèves américains choisissent par hasard), les motivations sont différentes face à ce type d'enquêtes. Les taux de scolarisation fortement variables suivant les pays (surtout dans le secondaire) biaisent les comparaisons (c'est un des principaux reproches faits à PISA, qui compare les performances des élèves *scolarisés* à 15 ans). Enfin, les curriculums divergent trop suivant les pays et la méthode censée permettre la réalisation de sub-tests comparables (l'Item Response Theory) ferait défaut.
- De la même façon, la comparabilité des indicateurs relatifs aux facteurs scolaires est mise en cause. La forte diversité dans les structures scolaires (âge d'entrée, filières, nombre d'années d'enseignement) ainsi que dans les concepts (le terme d' « enseignement privé », par exemple, ne représente pas la même chose dans tous les pays) rendent difficiles l'utilisation des statistiques internationales existantes. Nous survolons ce point développé plus en détail dans ce séminaire par la communication de Jake Murdoch.
- Enfin, le coût de ces enquêtes, particulièrement lourd pour les pays en voie de développement, est pointé du doigt et confronté aux bénéfices qu'ils peuvent en tirer. (Kellaghan, 1996).

Plusieurs remarques peuvent être tirées de cet ensemble de critiques :

- Ces critiques permettent de mieux cerner les conditions qui doivent prévaloir à la réalisation d'enquêtes internationales de qualité. En particulier, une attention forte doit être portée aux procédures d'évaluation : choix de l'échantillon, indication des marges d'erreurs, indication de la fiabilité des données...
- Face aux problèmes de comparabilité internationale, on doit se poser la question essentielle résumée par Schleicher (1995) : est-ce que les erreurs qui sont introduites par les problèmes de comparabilité sont faibles par rapport aux différences que les comparaisons sont censées expliquer ? Aucune mesure n'est parfaite mais toutes les erreurs ne discréditent pas la qualité des données et leur usage dans des travaux de comparaison.
- Une partie de ces critiques peut être adressée à l'encontre de toute tentative d'évaluation externe standardisée nationale, le caractère international de ces évaluations venant exacerber ces attaques. Face à ces critiques, l'alternative serait l'absence de toute évaluation des systèmes éducatifs. Est-ce mieux ?

Les méthodes et l'exploitation politique et médiatique des enquêtes internationales quantitatives sont également mises en cause

Au-delà des mesures des résultats académiques et des facteurs scolaires, c'est la méthode même et son exploitation médiatique et politique qui sont critiquées.

- Les détracteurs des enquêtes internationales dénoncent tout d'abord leur manque de fondements théoriques et méthodologiques (par exemple le choix des pays participants, Schriewer, 1988).
- Ils soulignent aussi l'absence de prise en compte des contextes nationaux et, corollaire, l'impossibilité de duplication des politiques éducatives qui ont réussi dans certains pays. On retrouve là la traditionnelle opposition entre particularistes et positivistes. Certains auteurs admettent cependant que des études quantitatives internationales longitudinales pourraient permettre des comparaisons à contexte égal.
- Ils affirment également que, les facteurs institutionnels agissant de façon concertée et complexes sur les systèmes éducatifs, les méthodes quantitatives qui analysent l'influence des différents facteurs individuellement ne peuvent conduire à des résultats valides (Kellaghan 1996).
- Certains auteurs dénoncent aussi le caractère « impérialiste » de ces enquêtes qui visent à imposer dans le monde entier un modèle éducatif dominant unique (Mollis et Marginson, 2002). Selon ces auteurs, tous les pays ne réagissent pas de façon identique aux résultats apportés par les enquêtes. De mauvaises performances signifient le début d'une remise en cause et une analyse plus précise des facteurs défaillants, dans les pays dominants (comme les USA, avec le rapport « A nation at Risk » (1983), suite à l'enquête SISS de l'IEA). Au contraire, les plus petits pays seraient plus prompts face à de mauvais résultats à adopter les politiques présentées comme performantes par les grandes institutions internationales.
- Les palmarès simplifiés, médiatisés et souvent utilisés par les politiques, sont jugés néfastes et ne reflétant pas la réalité. En effet, du fait des marges de fiabilité des estimations chiffrées, on ne devrait pas effectuer un classement traditionnel des pays mais plutôt indiquer, pour chacun d'entre eux, les pays dont les scores sont significativement supérieurs, inférieurs ou égaux. Ainsi, sur les 31 pays testés dans PISA-2000 (OCDE, 2001), il n'est pas exact d'attribuer à la France une quatorzième position, il est plus juste de préciser que seuls 7 pays ont des scores significativement supérieurs à celui de l'Hexagone. Il est à noter que les études elles-mêmes, comme TIMSS et PISA, présentent ainsi leurs résultats, la simplification du classement traditionnel est le fait des médias grand public.
- Enfin, les grandes enquêtes internationales sont remises en cause parce que leur niveau d'analyse - l'Etat nation -, a aujourd'hui perdu de son importance, face aux forces locales et internationales (Bray et Thomas (1995), Crossley (1999)...).

III. Un nécessaire renouvellement des enquêtes quantitatives internationales

Si ces critiques nombreuses sont, pour certaines d'entre elles, justifiées, un renouvellement de la conception et de la lecture des résultats des évaluations quantitatives internationales devrait permettre une double réconciliation entre méthodes quantitatives et qualitatives, entre positivisme et particularisme.

Nous avons essayé d'emprunter ce chemin original dans une recherche qui a porté, grâce aux données PISA-OCDE, sur l'évaluation des politiques de différenciation menées actuellement dans la plupart des pays de l'OCDE : différenciation territoriale avec les réformes de décentralisation, différenciation scolaire avec les politiques de diversification des parcours scolaires, différenciation des agents en charge des activités éducatives avec les politiques d'appui à l'enseignement privé. Les quelques pistes de réflexion suivantes sont tirées de ces travaux menés dans le cadre de notre thèse.

Le renouvellement des recherches internationales quantitatives peut à la fois porter, en amont, sur leur conception et, en aval, sur leur lecture et leur exploitation.

Renouvellement en amont dans la conception des études quantitatives internationales

La construction des grandes enquêtes peut évoluer dans cinq domaines.

Premièrement, l'Etat-nation ne doit plus être aujourd'hui le niveau d'analyse exclusif des évaluations internationales

Deux mouvements contradictoires ont en effet remis en cause le pouvoir exclusif des Etats centraux en matière d'éducation. Du fait d'une tendance généralisée à la décentralisation des structures éducatives, dans les pays développés et émergents (Mons, 2004), une part de plus en plus importante des décisions est désormais prise par les autorités locales ou les écoles elles-mêmes. D'après Broadfoot (1999), l'Union Européenne compte aujourd'hui plus d'une cinquantaine de systèmes éducatifs. Certaines données nationales, comme celles sur le volume horaire d'enseignement, les contenus des programmes ou les dépenses scolaires, ne reflètent plus qu'imparfaitement la réalité. Ceci expliquerait le fait que des recherches internationales, prenant comme niveau d'analyse exclusif les Etats, ne puissent plus mettre en lumière l'influence de certains facteurs explicatifs des apprentissages (Duru-Bellat, Mons et Suchaut, 2004).

A l'opposé de la décentralisation, l'existence de forces internationales exerçant une influence significative sur les réformes éducatives, avec des thèmes de prédilection comme la décentralisation, la privatisation, l'évaluation ou le choix de l'école (Carnoy et Rhoten, 2002) ont conduit à la création de zones éducatives globales dépassant le niveau de l'Etat-nation.

De façon à ne pas se cantonner à une analyse exclusivement au niveau pays, notre recherche a donc retenu non pas le critère de l'Etat-nation mais celui de système éducatif. Nous avons été aidés en cela par le fait que les données PISA sont agrégées par grandes provinces ou Etats pour les pays fédéraux. Quand des Etats présentaient, en interne, à la fois des performances scolaires territorialement très variables et des systèmes éducatifs distincts, nous avons choisi d'intégrer non pas l'Etat en tant que tel mais les systèmes éducatifs constituant l'Etat. Ainsi, les communautés belges francophone et flamande apparaissent-elles au même niveau d'analyse que la France.

De la même façon, nous avons analysé des zones géographiques et culturelles plus larges que les Etats-nation, zones dont les comportements en matière de décentralisation de l'éducation se sont avérés relativement homogènes.

Deuxième évolution nécessaire : la prise en compte des contextes historiques, politiques, économiques et sociaux dans lesquels se situent les systèmes éducatifs

Alors que la majorité des études quantitatives se limite à l'analyse statistique d'indicateurs chiffrés, nous avons souhaité appréhender la complexité de la réalité en définissant des indicateurs qualitatifs plus larges. Au-delà du phénomène étudié, ces indicateurs ont permis d'introduire des informations sur les contextes et les philosophies des politiques éducatives.

Ainsi pour l'enseignement privé, au-delà des évaluations traditionnelles quantitatives (effectifs scolarisés dans le privé, fonds publics destinés aux établissements privés...), les indicateurs développés ont cherché à rendre compte de la philosophie qui sous-tendait les politiques de développement du privé dans chaque pays (philosophie de compétition avec le public, de complémentarité...).

De la même façon, alors que les indicateurs de l'OCDE sur la répartition des pouvoirs pour la conception des programmes décrivent principalement les intervenants (gouvernement central, autorités locales ou écoles), nous

avons souhaité documenter la philosophie qui sous-tend la création des contenus scolaires. Ainsi, avons-nous distingué, dans notre base de données a) les pays dans lesquels la conception des programmes est unitaire - , un seul acteur est en charge de cette compétence que ce soit un gouvernement central (France ou Luxembourg) ou une autorité locale (comme les Länder allemands) – b) les Etats dans lesquels les programmes fonctionnent grâce à l'assignation d'objectifs aux écoles (la Grande-Bretagne par exemple) et c) les pays dans lesquels les programmes voient cohabiter une partie nationale et une partie décidée par les autorités locales (comme l'Espagne). Il est à noter que ce sont souvent ces indicateurs synthétisant plusieurs dimensions qui ont révélé les associations les plus fortes avec les performances scolaires.

Enfin, dans notre recherche, les contextes économiques, sociaux et éducatifs des pays ont été pris en compte grâce à l'introduction dans les modèles de variables permettant de refléter les différences de niveaux de développement. Cela nous a permis de comparer des pays présentant des contextes globaux disparates (depuis les pays développés de l'OCDE jusqu'aux pays émergents d'Amérique Latine et d'Asie), sans que ces différences de développement ne biaisent complètement nos résultats.

Ainsi, nous avons pu montrer que certains facteurs agissaient, quel que soit le niveau de vie du pays (comme les dépenses éducatives...) ou, au contraire, que l'influence de certains facteurs mis en avant dans d'autres études n'était que le résultat de différences de développement économique entre les pays de l'échantillon. Par exemple, Wossmann (2000) avait montré grâce aux données TIMSS que l'existence d'un examen externe national était associée à des résultats académiques supérieurs. En fait, cette association positive semble s'expliquer par le fait que les examens nationaux externes sont caractéristiques des pays les plus riches, pays présentant les résultats académiques les plus élevés. Une fois le PIB par habitant pris comme variable de contrôle, cette centralisation de l'évaluation s'est avérée neutre.

De telles ambitions – prendre en compte plusieurs niveaux d'analyse territoriale, situer le plus possible les phénomènes dans leur contexte – nécessitent un renouvellement dans la conception des indicateurs internationaux, c'est là une troisième évolution souhaitable.

Un renouvellement dans la conception des indicateurs internationaux

Se limiter aux bases de données internationales semble, en effet, aujourd'hui impossible pour atteindre ces objectifs. Outre les problèmes soulignés précédemment (validité, comparabilité...), des indicateurs créés sur mesure pour chacune des recherches permettent de conduire efficacement l'analyse souhaitée.

Dans notre cas, des indicateurs internationaux existaient sur l'ensemble des thèmes traités, cependant la création d'indicateurs propres a été nécessaire. Ainsi, si l'OCDE produit une batterie large d'indicateurs sur la répartition des pouvoirs en éducation, il s'est avéré nécessaire de produire une base de données originale d'une trentaine d'indicateurs sur ce sujet pour une quarantaine de pays, avec des nomenclatures plus synthétiques que celles de l'OCDE et donc plus adaptées aux analyses impliquant des modèles multivariés.

De la même façon, dans notre analyse des politiques de diversification des parcours scolaires, nous souhaitons évaluer l'association entre la longueur du tronc commun et les performances des systèmes. Là aussi, l'OCDE proposait un indicateur, renseigné certainement par les pays participants eux-mêmes. Or, comme en France, le développement de filières dans le premier cycle du secondaire est l'objet de tabou chez nos voisins. Certains pays ont donc déclaré qu'il n'existait pas de filières chez eux dans ce cycle quand il n'y avait pas une séparation stricte entre plusieurs voies scolaires (comme c'est le cas en Allemagne). La République Tchèque et la Hongrie ont ainsi déclaré que, dans leur système éducatif, le premier cycle du secondaire était unifié. Or, dans ces deux pays, des filières gymnasium sélectives ont été recrées depuis la chute du communisme et accueillent actuellement près de 20% des élèves les plus performants à leur entrée en « collège ». Ces pays, qui tendent à « refamiliariser » leur premier cycle du secondaire, sont donc dans des situations intermédiaires, entre l'Allemagne et la France. Pour les classer, nous avons, pour notre part, retenu une définition du tronc retenu plus rigoureuse que celle de l'OCDE : au-delà de 10% des élèves empruntant une voie spéciale nous avons considéré qu'il n'existait plus de tronc commun. Alors que l'indicateur de l'OCDE ne s'est pas révélé être associé aux

performances des élèves, notre indicateur l'a été : plus le tronc commun est long, plus les performances moyennes des élèves progressent.

Il n'y a donc pas de bons ou de mauvais indicateurs (Schleicher, 1995), mais seulement des indicateurs adaptés ou non à la recherche menée. Les bases de données internationales demeurent des références essentielles mais doivent être complétées par des indicateurs propres à forte valeur ajoutée.

De la même façon, des indicateurs sur-mesure se révèlent nécessaires pour pouvoir mener des enquêtes internationales efficaces incluant des pays de niveaux de vie très disparates, conjuguant ainsi Nord et Sud. Or, pour l'instant les bases de données internationales se focalisent le plus souvent exclusivement sur des données Nord ou Sud. L'OCDE par définition s'intéresse aux pays développés tandis que les grandes organisations internationales (banques de développement...), à l'exception notoire du PNUD et de l'UNESCO, développent des indicateurs centrés sur le Sud. Aujourd'hui ces deux mondes ne communiquent pas statistiquement parlant, seule la création d'indicateurs sur-mesure permet de faire des analyses Nord-Sud.

Autre raison expliquant le fait que des indicateurs sur-mesure doivent être développés : les indicateurs traditionnels tendent exclusivement à rendre compte du caractère formel des systèmes éducatifs et non de leur réalité. L'exemple du Japon est à ce titre significatif dans plusieurs domaines. Par exemple, la consultation des statistiques sur l'enseignement privé fait apparaître une très faible présence d'effectifs scolarisés dans le privé. Ce qui est le cas dans le système formel. C'est sans compter cependant sur les cours du soir privés, qui constituent une composante essentielle du modèle éducatif japonais. Recourir exclusivement aux statistiques existantes ne permet pas d'appréhender la totalité de la réalité des systèmes éducatifs.

Dernière raison expliquant la nécessité de création d'indicateurs qualitatifs ad hoc : l'intégration dans l'analyse d'indicateurs globaux représentatifs de typologies larges (modèles globaux d'organisation des pouvoirs, toutes compétences confondues, par exemple) permet de ne plus seulement tester l'effet d'un facteur pris individuellement mais d'évaluer l'influence de la conjonction de facteurs agissant de concert. On ne s'attache plus alors à une variable considérée isolément, on teste des formes globales d'organisations scolaires. Pour l'instant, ces tentatives ont cependant donné peu de résultats.

Dans tous les cas, ces nouveaux indicateurs nécessitent une vision multi-disciplinaire de chaque phénomène. Des incursions historiques, économiques, sociologiques et politiques ont été nécessaires pour appréhender plus globalement la complexité des systèmes éducatifs.

L'alliance nécessaire entre des observateurs internes et externes des systèmes éducatifs

Enfin, quatrième évolution nécessaire dans la conception amont des enquêtes internationales : ces évaluations nécessitent l'adoption d'une vision insider/outsider (Schriewer (1988), Crossley (1999)). Alors que l'observateur interne peut apporter une vision qualitative approfondie de son système éducatif, le point de vue extérieur est nécessaire pour homogénéiser et contrôler les informations. Si les enquêtes comme celles de l'IEA ont déjà adopté ce type de fonctionnement, la collecte aujourd'hui des statistiques par les grandes organisations internationales, comme l'UNESCO ou l'OCDE, ne suit pas toujours cette logique.

Dans notre recherche, pour le recueil des données, si nous avons eu recours à des interviews d'acteurs locaux ou des descriptifs locaux des situations nationales, nous nous sommes toujours fondés sur des paramètres factuels précis de façon à éviter les jugements internes sur les systèmes. Ainsi, nous n'avons pas demandé aux observateurs locaux si les programmes étaient centralisés ou pas, mais quels étaient les intervenants et de quelle manière précisément ils intervenaient dans la création des programmes (contenus, volumes horaires, pourcentage de marge de manœuvre laissée au local, existence d'options locales ou non...).

L'ensemble de ces évolutions nécessaires montre que, dans la conception des recherches internationales, des ponts doivent être jetés entre analyse quantitative et étude qualitative, la seconde n'étant pas perçue comme un simple ingrédient de la première, mais comme une complémentarité nécessaire. Comme le souligne Epstein

(1988), il n'existe pas de lien entre les positions épistémologiques (particularisme vs positivisme) et les méthodes utilisées (quantitatif vs qualitatif).

Au-delà de la conception amont des recherches, un renouvellement dans leur lecture et leur exploitation devrait également permettre de faire converger les approches particularistes et universalistes.

Renouvellement de la lecture et de l'exploitation des études quantitatives internationales.

Pour une lecture des études quantitatives plus précautionneuse

Si certains auteurs, normatifs, ont pu parler, dans leur exposition des résultats d'études quantitatives, de « facteurs explicatifs », ou d'« impact » de tels facteurs sur les performances des systèmes éducatifs, il nous paraît, cependant, essentiel de réaffirmer que les études quantitatives ne permettent pas de mettre en lumière des effets de causalité mais seulement des corrélations, des associations entre plusieurs variables.

De même, seul un ensemble de recherches quantitatives pointant vers le même résultat, à partir de méthodes d'analyse et de données différentes, peut permettre de conclure au moins temporairement sur un sujet. Nous l'avons vu, ces cas sont rares. Outre l'analyse quantitative, notre recherche a donc, de façon assez traditionnelle, fait la synthèse des études portant sur les mêmes sujets pour des pays comparables (effets des politiques de décentralisation...). Nos résultats sont replacés dans le contexte des recherches précédentes.

Seconde évolution nécessaire : une conception multiple des performances des systèmes scolaires

La plupart des études quantitatives se sont focalisées sur l'analyse de l'efficacité des systèmes (on étudie alors le score académique moyen du pays). La richesse des données collectées permet cependant d'appréhender les systèmes éducatifs de plusieurs points de vue. L'analyse peut porter sur d'autres valeurs que l'efficacité, l'égalité par exemple, déclinée sous plusieurs versions : égalités territoriales, sociales, de genre, ethniques. Notre recherche a ainsi consisté à déterminer quelles répartitions des pouvoirs en éducation étaient davantage associées aux systèmes les plus égalitaires –globalement ou socialement parlant. Plus les conceptions des performances sont multiples, moins les résultats des études se laissent enfermer dans un palmarès forcément schématique et réducteur. Une pluralité d'indicateurs doit appuyer chacun des concepts développés. Des visions multiples de la réalité permettent de l'appréhender plus justement. Les études quantitatives internationales n'enferment, donc, pas le chercheur dans la quête d'un système éducatif « idéal » coût-efficace, elles peuvent s'adapter à l'analyse de la maximisation de n'importe quelles valeurs (efficacité mais aussi équité...). Encore faut-il définir les valeurs dont le système éducatif doit être porteur.

Troisième évolution nécessaire : une approche relativiste des lois sociologiques

On reproche souvent à l'école universaliste, qui s'appuie aujourd'hui, entre autres, sur les études quantitatives internationales, de rechercher des lois invariantes, en dehors de tout contexte temporel et géographique. Certains auteurs comme Holmes (1981) propose une vision moins déterministe des conclusions des études quantitatives. Aux lois intemporelles, infaillibles (Le Thank Hoi, 1984), ils préfèrent parler de généralisations contextualisées dans lesquelles les comportements sont gouvernés par des conditions spatio-temporelles socio-psychologiques, linguistiques ou économiques spécifiques.

Ainsi, avons-nous mis l'accent dans notre recherche sur le fait que les résultats trouvés n'étaient pas des lois universelles mais s'avéraient relatifs à l'échantillon analysé. Les deux vagues d'évaluation de PISA - avec 1) « PISA-2000 » intégrant quasi-exclusivement des pays industrialisés et 2) « PISA+ » élargissant, en 2002, l'échantillon à des pays émergents (Indonésie, Pérou, Argentine, Roumanie, Albanie...)- ont permis de montrer que les résultats évoluaient en fonction des pays testés.

De même, des politiques éducatives en apparence identiques - le développement du choix de l'école par exemple en Suède et en Grande-Bretagne - ont abouti à des résultats divergents, dans des contextes différents : en Angleterre, les familles, surtout en milieu urbain se sont emparées de ce droit tandis qu'en Suède les élèves ont continué à être très majoritairement scolarisés dans les écoles de leur secteur. Des politiques identiques n'ont pas les mêmes conséquences dans des contextes différents. Toute exportation de politiques jugées performantes à l'étranger doit donc se faire après analyse du contexte initial de ces réformes. L'analyse quantitative reboucle donc sur une étude qualitative.

En conclusion :

- Des évaluations quantitatives internationales et le benchmarking qu'elles permettent sont nécessaires dans une perspective d'amélioration de la qualité de l'enseignement. Toute tentative d'évaluation peut prêter le flanc à la critique mais leur absence est préjudiciable.
- Un ensemble de conditions doit être réuni pour permettre des études de qualité, ceci d'autant plus que du fait de leur caractère international ces recherches acquièrent souvent une portée politique supérieure à celles des évaluations nationales.
- Analyse quantitative et qualitative doivent se compléter à tous les stades de la recherche : travail en amont sur les contextes, travail en aval pour valider la pertinence des résultats. Ce travail sera aidé par la création d'indicateurs ad hoc.
- L'analyse des performances doit être multiple : l'efficacité mais aussi d'autres valeurs comme l'équité doivent être prises en compte. Les indicateurs statistiques présentés doivent être nombreux car chacun ne révèle qu'une partie de la réalité de la situation.
- La prudence est de mise dans la lecture et l'exploitation politique des recherches quantitatives internationales. Les résultats, qui ne révèlent que des associations entre variables, ne sont pas des « recettes » facilement transposables d'un pays à l'autre.
- Malgré les critiques exacerbées et virulentes dont elles font l'objet, les recherches quantitatives internationales restent peu nombreuses, en particulier du fait de leur coût. Le champ de l'éducation comparée reste encore largement dominé par les études qualitatives portant sur un nombre très faible de pays. Seule une masse critique de recherches quantitatives internationales permettra de corroborer les premiers résultats et d'avancer des pistes de réflexion plus solides.

ANNEXE 1

Principales enquêtes internationales sur l'évaluation des connaissances/compétences des élèves dans le primaire et secondaire (1960-2003)

| Date | Nom de l'enquête | Organisme en charge de l'enquête | Zone géographique couverte | Age des élèves évalués | Disciplines (nombre de pays participants) |
|---------|--|----------------------------------|----------------------------|------------------------|---|
| 1960 | Projet Pilote | IEA | International | 13 | Mathématiques, sciences, compréhension de l'écrit, géographie (12 pays) |
| 1964 | FIMS (First International Mathematics Study) | IEA | International | 13, fin du secondaire | Mathématiques (12 pays) |
| 1970-71 | Six-subject | IEA | International | 10, 14, fin | Science (19 pays), |

| | | | | | |
|------------|---|-------------|------------------------------------|--|---|
| | study | | | du secondair e | compréhension de l'écrit (15 pays), littérature (10 pays), langue étrangère (Anglais et Français, 18 pays), éducation civique (10 pays) |
| 1980-82 | SIMS (Second International Mathematics study) | IEA | International | 13, fin du secondair e | Mathématiques (20 pays) |
| 1982-83 | Classroom environment | IEA | International | 9, 15 | Mathématiques, sciences, histoire (10 pays) |
| 1984 | SISS (Second International Science Study) | IEA | International | 10, 14, fin de secondair e | Science (24 pays) |
| 1985 | Written composition | IEA | International | 10, 14-16, fin du secondair e | Composition écrite (10) |
| 1988 | IEAP International assessment | IEAP | International | 13 | Mathématiques (6 pays) |
| 1989, 1992 | Computer in education | IEA | International | 10, 13 | Maîtrise des technologies de l'information (23) |
| 1986-2002 | Pre-primary project | IEA | International | Mêmes enfants suivis de 4 à 7 ans | Expérience pré-scolaire(15 pays) |
| 1990-91 | IEAP International assessment | IEAP | International | 9, 13 | Mathématiques e sciences (20 pays) |
| 1991 | Reading literacy | IEA | International | 9, 14 | Lecture (31 pays) |
| 1994-99 | TIMSS (Third Internatinal Mathematics and Science Study) | IEA | International | 9, 13, fin du secondair e | Mathématiques, sciences (40) |
| 1995-98 | PASEC | CONFENEM | Afrique sub-saharienne francophone | 8, 11 ans | Mathématiques et français (9 pays) |
| 1997 | Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados | Laboratorio | Amérique Latine | 6-7 ans | Mathématiques, langue maternelle (13 pays) |
| 1998-99 | SACMEQ I | SACMEQ | Afrique Australe | 12 ans | Lecture/écriture (7 pays) |
| 1997-2000 | TIMSS-R (Third Internatinal Mathematics and Science Study- Repeat) | IEA | International | 9, 13, fin du secondair e | Mathématiques et sciences (38 pays) |
| 1994- | Civic Education | IEA | International | 14, 16-18 | Education civique (32) |

| | | | | | |
|-----------|------------|--------|----------------------------|------------|--|
| 2002 | Study | | | | |
| 1999-2002 | SACMEQ II | SACMEQ | Afrique de l'Est et du Sud | 12 ans | Lecture/écriture, mathématiques (15 pays) |
| | PIRLS | IEA | International | 9 ans | Lecture (36 pays) |
| 2000-2003 | PISA-PISA+ | OCDE | International (dont OCDE) | 15 ans | Langue maternelle, mathématiques (44 pays) |
| 2000-2004 | TIMSS-2003 | IEA | International | 10, 14 ans | Mathématiques, sciences (50 pays) |

Sources : Goldstein (1995), complété par l'auteur.

Enquêtes à portée régionale :

Références

BRAY M., THOMAS M. (1995). – Levels of Comparison in Educational Studies : Different Insight from Different Literatures and the Value of Multilevel Analyses. **Harvard Educational Review**, 65, 3, p. 472-490

BROADFOOT P. (1999). – Stones from Other Hills may Serve to Polish the Jade of this One : towards a neo-comparative "learnology" of education. **Compare**, 29, 3, p. 217-231

CARNOY M., RHOTEN D. (2002). – What Does Globalization Mean for Educational Change ? A Comparative Approach. **Comparative Education Review**, 46, 1, p. 1-9

COLEMAN J.S. *et al* (1966). – **Equality of Educational Opportunity** (2 vols). Washington D.C. : U.S. Government Printing Office

CROSSLEY M. (1999). – Reconceptualising Comparative and International Education. **Compare**, 29, 3, p. 249-267

CUMMING J. (1996). – The IEA Studies of Reading and Writing Literacy : a 1996 perspective. **Assessment in Education**, 3, 2, p. 161-178

DURU-BELLAT M., MONS N., SUCHAUT B. (2004). - **Caractéristiques des systèmes éducatifs et compétences des jeunes de 15 ans : l'éclairage des comparaisons entre pays**. Rapport pour le M.E.N., DIJON : IREDU-CNRS

EPSTEIN E. (1988). – The Problematic Meaning of "Comparison" in Comparative Education. In J. SCHRIEWER, B. Holmes (eds.) **Theories and Methods in Comparative Education**. Frankfurt : Verlag Peter Lang

FARRELL J. (1979). – The Necessity of Comparisons and the Study of Education : The Saliency of Science and the Problem of Comparability. **Comparative Education Review**, 23, 1, p. 3-16

GOLSTEIN H. (1995). – **Interpreting international comparisons of student achievement**. Paris : UNESCO Publishing

HEYNEMAN S.P., LOXLEY W.A. (1983). – The effect of primary school quality on academic achievement across twenty-nine high- and low-income countries. **American Journal of Sociology**, 88, 6, p.1162-94

HUSEN T. (1987). – Policy Impact of IEA Research. **Comparative Education Review**, 31, 1, p. 29-46

HUSEN T., POSTLETHWAITE N. (1996). – A Brief History of the International Association for the Evaluation of Educational Achievement (IEA). **Assessment in Education**, 3, 2, p. 129-141

JULLIEN M.A., (1817). – **Esquisse et Vues préliminaires d'un Ouvrage sur l'Education Comparée**, cité dans H. VAN DAELE (1993). – L'Education Comparée, collection Que Sais-je ? Paris : PUF

KEEVES J. (1995). – **Selected Key Findings from 35 Years of IEA Research**, The Hague: IEA

KELLAGHAN T. (1996). – IEA Studies and Educational Policy. **Assessment in Education**, 3, 2, p. 143-160

LE THANH KHOI (1986). – Toward a General Theory of Education. **Comparative Education Review**, 30, 1, p. 12-29

LAMBIN R. (1995). – What can planners expect from international quantitative studies ?, in : W. Bos et R. H. Lehmann (Eds) **Reflections on Educational Achievement. Papers in honor of T. Neville Postlethwaith**, p. 169-182, New York : Waxman

MARGINSON S., MOLLIS M. (2001). – « The door opens and the tiger leaps » : Theories and Reflexivities of Comparative Education for a Global Millennium. **Comparative Education Review**, 45, 4, p. 581-615

MALLISON V. (1975). – **An Introduction to the Study of Comparative Education**. London : Heinemann

MAURICE M., SELLIER F., SILVESTRE J.J. (1982). – **Politique d'éducation et organisation industrielle en France et en Allemagne**. Paris : PUF

MONS N. (2004). – Politiques de décentralisation en éducation : diversité internationale, légitimations théoriques et justifications empiriques. **Revue Française de Pédagogie**, 146, p. 41-52

NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION (1983). – **A Nation at Risk : the imperative for educational reform**. Washington D.C. : US Government Printing Office.

NOAH H. J., ECKSTEIN M. A (1969). – **Toward a Science of Comparative Education**. New York : Macmillan

OCDE (2001). – **Connaissances et compétences : des atouts pour la vie. Premiers résultats de PISA 2000**. Paris : OCDE

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

PSACHAROPOULOS G. (1990). - Comparative Education : From Theory to Practice, or Are You A:neo.* Or B:*.ist ? **Comparative Education Review**, 34, 3, p. 369-381

SADLER M. (1900). – How far can we learn anything of practical value from the study of foreign systems of education, in J. H. HIGGINSON (Eds, 1979) **Selections from Michael Sadler**. Liverpool : International Publishers Ltd

SCHLEICHER A. (1995). – Comparability Issues in International Education Comparisons, in : W. Bos et R. H. Lehmann (Eds) **Reflections on Educational Achievement. Papers in honor of T. Neville Postlethwaith**, p. 216-229, New York : Waxman

SCHRIEWER J. (1988). – The Method of Comparison and The Need for Externalization : Methodological criteria and sociological Concepts, in : J. SCHRIEWER, B. HOLMES (eds.) **Theories and Methods in Comparative Education**. Frankfurt : Verlag Peter Lang

UNESCO (2000). – **Education pour tous. Situation et tendances 2000. L'évaluation des acquis scolaires.**
Paris : UNESCO

WÖSSMANN L. (2000). - **Schooling Resources, Educational Institutions, and Student Performance : The International Evidence**, Kiel Working Paper n°983. Germany : KIWE

FONTE:

http://www.prisme-asso.org/IMG/pdf/L_evaluation_quantitative_internationale_des_systemes_educatifs.pdf