

Quelques réflexions méthodologiques à propos des enquêtes internationales dans le domaine de l'éducation¹

Christian MONSEUR et Marc DEMEUSE *

INTRODUCTION

A en croire, par exemple, les articles de presse qui paraissent dans les quotidiens belges francophones³ depuis un peu moins d'une décennie et qui relatent les résultats des enquêtes internationales, les élèves de la Communauté française de Belgique détiennent la palme en matière de contre-performances académiques. Si la publication du rapport international (Beaton *et al*, 1996) de la *Troisième étude internationale en mathématiques et en sciences* (TIMSS) conduite sous l'égide de l'IEA a produit des titres plutôt alarmistes⁴, le rapport international de l'enquête PISA menée par l'OCDE (2001) a déchaîné l'imagination de la presse, y compris en matière de jeux de mots, de calembours et d'autodérision⁵.

Ces commentaires et interprétations ont le mérite de poser deux questions fondamentales qui dépassent largement le contexte belge. La première concerne

¹ Les auteurs tiennent à remercier Marcel Crahay, Dominique Lafontaine, Aletta Grisay, Denis Meuret, Sophie Morlaix et Jean-Jacques Paul, ainsi que deux lecteurs anonymes, pour leurs remarques et commentaires à propos de cet article.

* *Assistant de Recherches (Service de Pédagogie théorique et expérimentale, Université de Liège) et Chercheur associé à l'Australian Council of Educational Research. Email : cmonseur@ulg.ac.be ; Chargé de Recherches (Service de Pédagogie théorique et expérimentale, Université de Liège) et Chercheur associé à l'Institut de Recherche sur l'Éducation (IREDU, Université de Bourgogne). Email : marc.demeuse@ulg.ac.be*

³ On peut trouver sur le site de l'OCDE (<http://www.pisa.oecd.org/News/cntry.htm>), une revue de presse internationale, réalisée lors de la sortie des résultats de PISA 2000.

⁴ "Ils sont nettement plus forts que nous, de loin (La Wallonie, 22/11/1996), "Moyens en math, nos élèves sont quasi les plus mauvais en sciences" (Le Soir, 10/10/1997), "L'enseignement des sciences est-il dans le 36^{ème} dessous ?" (Le Soir, 10/10/1997).

⁵ "Pisa, la tour de la honte", "Le Bulletin francophone est déplorable", "Rapport affligeant pour les élèves francophones", "Ces cancrs francophones", "Mauvais bulletin pour nos élèves : l'OCDE classe les ados francophones en queue de peloton", "Nos élèves au fond de la classe : peut mieux faire !", "Cela va de mal en pis(a)...", "On a enfin trouvé un domaine où la Wallonie n'est pas à la traîne : la fabrication de cancrs".

la nature des enquêtes internationales. S'agit-il d'une recherche en sciences de l'éducation ou d'un simple concours, au même titre que les *Olympiades de Mathématiques* ? Cette question suscite également le débat parmi les chercheurs. La mise en œuvre de PISA 2000 par l'OCDE, dont la vocation première, à la différence de l'IEA, n'est pas la recherche scientifique, ne peut qu'attiser les différences de sensibilité. Les critiques classiquement évoquées, tant par le "grand public" que par certains spécialistes de l'éducation, concernent la qualité des mesures opérées, la présence de biais culturels, le manque de standardisation des conditions de collecte des données ou, au contraire, l'absurdité d'une telle standardisation, la validité et la représentativité des échantillons testés, et la réduction d'un processus aussi complexe que l'éducation à quelques indicateurs numériques. Ces réactions, qui semblent, en première analyse, se placer dans le champ méthodologique, dissimulent généralement une opposition de principe à toute évaluation internationale quantitative à visée comparative.

La seconde question concerne les limites qu'il convient d'assigner à l'interprétation des indicateurs produits. Le classement de tel pays dans telle discipline signifie-t-il que l'enseignement de celle-ci au sein des écoles primaires et secondaires est de mauvaise qualité ? L'ampleur des différences de rendement entre filles et garçons, entre natifs et immigrés, entre établissements scolaires est-elle nécessairement le reflet d'un manque d'équité des systèmes éducatifs ? L'évaluation de l'efficacité d'un système éducatif ou de son équité doit prendre en considération les finalités qui lui sont assignées et les moyens mis en œuvre pour les rencontrer. Dans ce contexte, un indicateur de performance des élèves par rapport à un ensemble de savoirs et de savoir-faire ne peut à lui seul témoigner de l'efficacité de l'institution scolaire.

Or, il est malheureusement fréquent d'attribuer à un ensemble de chiffres une signification qui dépasse largement les limites imposées par le contexte méthodologique dans lequel il a été produit. Ainsi, a-t-on pu lire, à propos des résultats de PISA 2000 : "*Rappelons d'abord que l'école n'est pas toute seule. L'étroite corrélation entre origine sociale et résultats scolaires est bien connue. Or, au cours des deux dernières décennies, la Wallonie a continué de s'appauvrir par rapport à la Flandre. Il ne faut donc pas s'étonner si les résultats scolaires moyens y sont moins bons*"⁶. Une relation observée au niveau des élèves se voit ici immédiatement transposée au niveau des systèmes éducatifs, sans aucune vérification empirique. Sans remettre en question la relation entre la richesse d'un pays et les performances académiques de ses élèves (Duru-Bellat, Mons et Suchaut, 2003), ce passage traduit avec une certaine acuité l'absence d'un cadrage méthodologique des indicateurs publiés. Cette tendance, qui renforce la conviction des opposants à toute démarche quantitative, ne se retrouve pas uniquement chez les journalistes : elle peut être présente également parmi les chercheurs.

⁶ Le Soir, 06/12/2001.

Cet article a pour objectif de montrer, à travers trois exemples issus de l'analyse des résultats du *Programme International pour le Suivi des Acquis des Elèves* (PISA, 2000) de l'OCDE, l'impact de certains choix méthodologiques sur les résultats. Trois indicateurs de disparité seront discutés : les différences de résultats entre filles et garçons, entre natifs et immigrés et entre établissements scolaires. Plus précisément, on traitera des implications :

1. de la forme des réponses (ouvertes vs. fermées) sur les taux de réussite des filles et des garçons,
2. de la manière de définir le concept "d'immigré" sur les résultats des élèves nés ou non dans le pays où ils sont testés et instruits,
3. de la définition de la population cible (une année d'étude particulière vs. les élèves d'un âge donné) sur les différences observées entre établissements d'enseignement.

Il ne s'agit nullement ici de développer de nouveaux arguments à l'encontre des études internationales, mais bien de rappeler l'importance et la nécessité d'un cadrage méthodologique lors de l'interprétation d'indicateurs. Les auteurs sont en effet persuadés de l'intérêt des comparaisons internationales telles que celles menées actuellement par l'OCDE et l'IEA. Ils y participent et en utilisent les résultats (par exemple, Demeuse, Crahay, Monseur, 2001 ; Groupe Européen de Recherche sur l'Equité des Systèmes Educatifs, 2003 ; Monseur, 1997 ; Monseur et Demeuse, 1998). Ils se veulent cependant attentifs à ne pas faire dire aux chiffres ce que manifestement ceux-ci ne peuvent pas dire et soulignent dans cette perspective la nécessité de bien maîtriser les outils statistiques et méthodologiques indispensables à ce genre d'approche.

DISPARITÉS ENTRE FILLES ET GARÇONS

Les rapports internationaux accordent classiquement une certaine importance aux disparités entre filles et garçons. L'interprétation de ces indicateurs nécessite une grande prudence car ils sont à la fois affectés par la nature de la population cible, le cadre de référence de l'évaluation, la nature des questions posées...et les différentes interactions qui pourraient exister entre ces éléments.

Les études de l'IEA ont pour habitude de se référer au grade modal fréquenté par les élèves d'un âge donné. A l'inverse, l'étude PISA 2000 s'intéresse à une population d'âge : les élèves de 15 ans encore scolarisés, indépendamment de l'année scolaire qu'ils fréquentent.

Dans les pays européens qui ont opté pour la promotion automatique⁷, le choix de l'une ou de l'autre définition de la population cible ne génère, en principe, aucune différence. Plus de 95% des élèves d'un âge donné se retrouvent au sein d'un seul et même niveau scolaire, appelé "grade"⁸. Par contre, au sein des pays qui pratiquent le redoublement, les élèves d'un âge donné se distribuent sur deux, voire trois et même parfois plus de trois grades, comme en Communauté française de Belgique (tableau 1). Par ailleurs, l'importance des retards peut être influencée par le sexe, les garçons étant, en moyenne, plus sujets au redoublement. Le tableau 1 permet de mettre cet effet en évidence en Communauté française de Belgique : si 62,38% des filles fréquentent le grade attendu (grade 10), seuls 49,95% des garçons sont scolarisés à ce niveau.

Tableau 1 : Distribution des élèves de 15 ans en fonction du grade fréquenté et du sexe en Communauté française de Belgique (en % de la classe d'âge pour chacun des deux sexes et pour l'ensemble de la population)

Grade	7	8	9	10	11	12	Total
Filles	0,60	7,35	28,40	62,38	1,27	0,00	100,00
Garçons	0,17	9,43	39,43	49,95	0,89	0,13	100,00
Total	0,38	8,39	33,93	56,14	1,08	0,07	100,00

Dans le cadre des études IEA où la population cible est définie en terme d'année scolaire, il s'ensuit que les garçons sont en moyenne plus âgés que les filles. L'absence d'indépendance statistique entre l'année scolaire fréquentée et le sexe (PISA 2000) ou la différence d'âge moyen entre filles et garçons au sein d'un grade donné (études IEA) incite à se pencher sur l'influence du choix de la population cible sur les disparités entre sexes.

Un second facteur souvent négligé dans l'explication des différences entre sexes réside dans le cadre de l'évaluation (OCDE, 1999, 2000 pour PISA 2000). Dans les résultats obtenus lors d'études internationales, on observe des variations de l'amplitude des disparités liées au sexe selon les domaines et sous-domaines d'investigation. Ainsi, dans un travail de synthèse de la littérature scientifique relative à l'enseignement des mathématiques, Brusselmans et Henry (1994) indiquent que "*Girls seem to outperform boys on computational tasks and do less well on problem solving. Moreover boys outperform girls on spatial tasks*". Dans le domaine des sciences, certaines tendances se dégagent également : "*In two content*

⁷ C'est-à-dire le passage d'un niveau scolaire au niveau scolaire supérieur sans possibilité de répéter une année en raison d'insuffisances graves (Crahay, 1996, pour une analyse approfondie du redoublement et de l'échec scolaire en Europe).

⁸ Le terme "grade" est utilisé pour éviter la confusion induite, en français, par le terme "année scolaire", aussi bien utilisé pour décrire une année de scolarité, comprise entre le jour de la rentrée des classes et le dernier jour de classe, et un niveau donné du cursus scolaire. Dans le système éducatif belge, le grade 10 correspond à la 4^e année de l'enseignement secondaire alors que le grade 12 correspond à l'année terminale de cet enseignement. Les élèves de 15 ans, sans retard scolaire, devraient y fréquenter le grade 10.

areas, life science and scientific inquiry and the nature of science, there were no statistically significant gender differences, either across all countries and within each country. However, boys outperformed girls on average internationally ; in each of the other content areas. The gender difference was greatest in physics, in which boys scored higher than girls by 21 scale-score points on average internationally in 12 countries boys performed significantly higher than girls. The next largest gender difference was in earth science, in which boys outperformed girls by 17 scale-score points on average internationally ; in six countries boys performed significantly higher than girls." (Martin *et al.*, 2000). Selon l'importance relative accordée aux différentes sous disciplines ou aux différents contenus au sein d'une épreuve internationale, les disparités liées aux sexes peuvent varier (pour une synthèse des différences garçons-filles dans les enquêtes internationales, voir Lafontaine et Blondin, à paraître). Il importe donc de bien garder à l'esprit, lorsqu'on compare une même grandeur dans deux évaluations différentes, la définition précise de ce qui, à chaque fois, est évalué réellement par rapport au curriculum.

Toutes les épreuves internationales ne sont cependant pas définies par rapport à un curriculum d'enseignement. Le concept de "littératie", tel qu'il est utilisé dans l'enquête PISA constitue un bon exemple d'une évaluation internationale ne portant pas sur un curriculum d'enseignement (Lafontaine, 2001). Selon le cadre théorique de PISA 2000 (OCDE, 2000) "*Le lecteur réagit de diverses manières à un texte donné, lorsqu'il cherche à exploiter et à comprendre ce qu'il lit. Ce processus dynamique intègre de nombreuses dimensions, dont trois ont été utilisées pour structurer l'évaluation PISA : (i) Processus - tâches de lecture, (ii) Contenu - types de texte, (iii) Contexte - but du texte*". Comme on peut le constater, les problèmes posés par les évaluations relatives à un curriculum d'enseignement se retrouvent ici également puisque, par exemple, cinq types de tâches de lecture ont été préalablement définis : (i) comprendre globalement un texte, (ii) trouver l'information, (iii) développer une interprétation, (iv) réfléchir sur le contenu du texte et (v) réfléchir sur la forme et la structure d'un texte. Les textes sélectionnés ont été classés en deux grandes catégories : (i) les textes continus, incluant les textes descriptifs, narratifs, informatifs, argumentatifs et injonctifs, et (ii) les textes non continus, incluant les formulaires, les annonces et réclames, les graphiques, les diagrammes, tableaux et matrices et enfin les cartes. La troisième dimension, le but du texte, comporte les finalités suivantes : (i) la lecture à usage privé, (ii) la lecture à usage public, (iii) la lecture à des fins professionnelles, (iv) la lecture à des fins éducatives. Afin d'évaluer la compétence des populations scolaires en littératie, différents types de questions ont été développés : (i) les questions à choix multiple, (ii) les questions complexes à choix multiple, (iii) les questions à réponse construite fermée, (iv) les questions à réponse courte et (v) les questions à réponse construite ouverte.

Différentes échelles de performance ont été construites à partir des réponses des élèves : (i) une échelle combinée de littératie, (ii) trois échelles de compétences (retrouver des informations, interpréter, réfléchir et évaluer) et (iii)

deux échelles relatives au type de textes (une échelle pour les textes continus et une échelle pour les textes non continus).

Cette décomposition de la compétence en littératie en fonction des processus ou en fonction des textes permet de compléter, nuancer, affiner la description de la performance des populations scolaires, offrant ainsi plus de pistes de réflexion pour les décideurs politiques.

Les différences de rendement entre filles et garçons ont par ailleurs été calculées pour ces différentes échelles et publiées dans divers rapports internationaux (OCDE, 2001, Kirsch *et al.*, 2003) et de nombreux documents nationaux⁹. Le tableau 2 résume la moyenne OCDE¹⁰ des différences filles-garçons pour les échelles disponibles.

Tableau 2 : Différence brute entre filles-garçons, écart-type et différence standardisée (une différence positive indique un avantage pour les filles)

Type d'échelle	Différence brute	Ecart type	Différence standardisée ¹¹
Echelle combinée	32	100	0,32
Retrouver l'information	24	111	0,22
Interpréter	29	100	0,29
Réfléchir et évaluer	45	106	0,42
Textes continus	39	101	0,39
Textes non continus	17	109	0,16

Selon les compétences (ii) ou selon le type de textes (iii) envisagé, les différences entre les filles et les garçons varient plus ou moins du simple au double, lorsqu'on considère les différences standardisées. Alors que l'écart n'est que de 0,22 pour les tâches qui nécessitent la recherche d'informations contenues dans le texte (retrouver l'information), il s'élève à 0,42 écart type lorsqu'il s'agit de réfléchir et d'évaluer tant le contenu que la structure d'un texte. De même, l'écart filles-garçons diffère selon le type de textes : il s'élève à 0,17 pour les textes non continus et à 0,39 pour les textes continus, à l'avantage des filles. Ce dernier résultat est par ailleurs confirmé par l'étude relative

⁹ La liste des rapports nationaux, relatifs à PISA 2000, et les textes au format PDF, sont disponibles sur le site de l'OCDE (www.pisa.oecd.org/pisa/cntry.htm).

¹⁰ La moyenne OCDE est basée sur l'ensemble des pays membres de l'OCDE, à l'exception des Pays-Bas qui ne satisfaisaient pas aux conditions d'échantillonnage, et attribuée à chaque pays la même pondération. Cette option peut naturellement, elle aussi, être discutée. On pourrait, en effet, calculer la moyenne OCDE en pondérant en fonction de la population de chaque pays plutôt qu'en leur accordant un poids identique.

¹¹ Cette colonne exprime la différence brute entre filles et garçons par rapport à l'importance de l'écart type de la distribution des scores sur l'échelle considérée de manière à faciliter la lecture. Ainsi, la différence entre filles et garçons, sur l'échelle combinée, vaut 32 points ou 0,32 après avoir pris en compte l'écart type (100 points) de la distribution des notes de tous les élèves, filles et garçons, sur cette échelle.

également à la littératie et menée sous les auspices de l'IEA en 1990 (Lafontaine, 1996).

Ces variations dans les écarts observés traduisent sans aucun doute une certaine réalité. Cependant, avant d'opérer une interprétation de type cognitif ou comportemental – les filles lisent davantage des livres de littérature et par conséquent développent des processus cognitifs plus complexes que les garçons – il convient de s'interroger sur la nature même du matériel de test.

PISA 2000 comporte quelque 130 questions de littératie, réparties comme suit entre processus et type de questions (regroupées en deux grandes catégories) : (i) questions à choix multiple, incluant les questions simples et les questions complexes à choix multiple, (ii) les questions qui nécessitent une production de l'élève, incluant les questions à réponse construite fermée, les questions à réponse courte et les questions à réponse construite ouverte.

Tableau 3 : Nombre de questions de littératie en fonction des compétences évaluées et du type de questions

	<i>Questions fermées à choix multiple</i>	<i>Questions ouvertes à réponse construite</i>	<i>Total</i>
Retrouver l'information	12	24	36
Interpréter	43	21	64
Réfléchir et évaluer	5	24	29
Total	60	69	129

La statistique X^2 rejette, avec une probabilité inférieure à 0,001, l'hypothèse d'indépendance. En d'autres termes, selon les compétences évaluées, il y a plus de questions fermées que de questions ouvertes ou inversement.

Or, Lafontaine et Monseur (à paraître) ont démontré, au départ de l'analyse des réponses des élèves dans la base de données PISA 2000, que les différences entre filles et garçons sont liées aux types de questions : la différence standardisée pour les questions à choix multiple s'élève à 0,19 (en faveur des filles) et la différence standardisée pour les questions ouvertes est de 0,29 (en faveur des filles).

Il s'impose donc de déterminer la part d'influence respective des types de tâches et de la nature des questions dans l'ampleur des disparités liées au sexe. A cette fin, Lafontaine et Monseur ont calculé les scores des élèves en littératie selon six modèles unidimensionnels définis par le croisement du type de compétences et de la nature des questions. Ces six modèles sont : (i) questions à choix multiple, retrouver des informations, (ii) questions ouvertes, retrouver des informations, (iii) questions à choix multiple, interpréter, (iv) questions ouvertes, interpréter, (v) questions à choix multiple, réfléchir et évaluer et (vi) questions ouvertes, réfléchir et évaluer.

Les disparités liées au sexe pour ces six dimensions sont présentées dans le tableau 4. Les valeurs médianes résument les tendances observées dans la majorité des pays : les disparités liées au sexe sont les plus faibles pour les tâches de recherche d'informations évaluées par l'intermédiaire des questions à choix multiple et les disparités les plus importantes s'observent pour les tâches de réflexion et d'évaluation mesurées à l'aide de questions ouvertes. Quel que soit le processus ou type de tâches, le recours aux questions ouvertes génère de plus fortes disparités, en faveur des filles.

Tableau 4 : Différences liées au sexe en fonction des types de tâches et de la nature des questions (les valeurs positives indiquent que les différences sont en faveur des filles)

	<i>Retrouver des informations</i>		<i>Interpréter</i>		<i>Réfléchir et évaluer</i>	
	QCM	QO	QCM	QO	QCM	QO
Australie	35,06	29,66	36,81	37,75	56,26	42,43
Autriche	2,76	22,27	22,36	25,46	28,19	42,44
Belgique	18,26	28,18	29,30	35,42	34,02	46,89
Brésil	-0,48	16,60	14,46	15,65	16,91	28,00
Canada	23,69	28,20	29,32	34,73	33,33	47,39
Suisse	16,39	27,14	26,01	28,43	26,67	47,24
Tchéquie	18,06	34,64	29,74	40,08	37,73	59,60
Allemagne	12,26	36,78	34,24	38,23	36,64	56,83
Danemark	6,54	19,23	19,48	30,54	23,65	46,28
Espagne	12,86	20,06	21,98	25,21	25,09	40,40
Finlande	36,67	50,48	48,97	57,60	50,56	65,37
France	16,56	29,36	26,87	30,78	30,27	39,51
GBR	16,59	24,14	22,73	27,55	31,87	33,40
Grèce	25,10	31,15	31,69	38,72	33,25	57,05
Hongrie	21,00	24,22	26,95	30,56	31,48	46,48
Irlande	17,71	22,80	25,52	32,89	36,94	36,98
Islande	21,47	37,93	32,15	53,48	32,43	57,42
Italie	21,88	34,48	33,22	45,30	33,75	49,10
Japon	22,60	30,31	26,57	25,24	34,82	43,58
Corée	3,01	9,21	9,88	11,47	14,70	29,23
Luxembourg	10,00	28,76	28,11	35,56	29,37	48,06
Lettonie	29,17	52,82	49,40	55,49	35,91	74,90
Mexique	8,64	16,75	14,18	22,33	10,91	37,45
Pays-Bas	24,83	18,80	26,94	39,13	21,64	36,01
Norvège	21,56	36,53	36,71	50,88	47,65	58,32
Nouvelle Zélande	41,36	40,52	41,76	52,33	45,17	57,44
Pologne	19,98	30,13	31,10	39,95	34,27	55,92
Portugal	12,69	20,34	26,16	25,42	19,53	40,05
Russie	28,70	36,21	32,16	37,65	32,35	47,69

Suède	29,68	30,72	35,69	39,02	44,46	50,70
Etats-Unis	22,37	31,39	26,58	33,45	27,29	39,97
Médian	19,98	29,36	28,11	35,42	32,43	46,89

Comme le montre le tableau 4, selon l'importance relative accordée aux différents types de compétence et à la nature des questions utilisées, les disparités peuvent varier du simple à plus du double. Pour une même compétence, comparativement aux questions fermées, les questions ouvertes sont toujours associées à des disparités plus importantes. Elles sont toutes en faveur des filles, sauf dans un cas (questions à choix multiple, retrouver de l'information), au Brésil.

Pour estimer la contribution relative des différents facteurs, ces différences ont été soumises à une analyse de la variance. Le tableau 5 présente les pourcentages des sommes des carrés des différents facteurs (pays, type de tâches et nature des questions et interactions entre ces trois facteurs).

Tableau 5 : Décomposition de la variance totale des différences de réussite entre les filles et les garçons en fonction de l'appartenance nationale, de la nature des questions et des compétences évaluées (et des différentes interactions entre ces variables)

<i>Source</i>	<i>Pourcentage de variance expliquée</i>	<i>Pourcentage de variance expliquée sur différences standardisées</i>
Pays	46,93	41,92
Nature des questions	16,21	17,60
Type de tâche	24,08	25,39
Pays * nature des questions	4,41	2,15
Pays * type de tâche	3,06	3,35
Nature des questions * type de tâches	1,87	6,65
Résidu (interaction de 3ème ordre)	3,44	2,92

Le pays reste le déterminant premier dans l'explication des différences entre filles et garçons. Selon que l'on envisage les différences absolues ou les différences standardisées¹², les pourcentages de sommes de carrés expliqués s'élèvent respectivement à 47% et 42% de la variance totale. La nature des questions explique environ 17% et le type de tâches plus ou moins 25% de la variance observée entre les scores des filles et des garçons à travers tous les pays et les différentes formes de question. Autrement dit, si on considère que la variable "pays" recouvre les particularités non précisées de chaque pays, les variables "nature des questions" et "type de tâche" prennent respectivement en compte 16,21% et 24,08% de la variation des scores observés entre les filles et les garçons. L'explication des disparités entre les scores des filles et des garçons

¹² Différences standardisées : différences brutes divisées par l'écart type.

doit aussi prendre en compte les interactions qui existent, d'une part, entre l'appartenance à un pays particulier et, d'autre part, la forme des questions (nature des questions et type de tâche). Des interactions existent également entre la nature des questions et le type de tâche. Une interaction complexe entre le pays d'appartenance, la nature des questions et le type de tâche est encore observée dans le modèle. La décomposition de la variance permet donc de mettre en évidence l'importance respective de trois facteurs : le pays d'appartenance (variable qui recouvre de manière globale toutes les spécificités nationales), mais aussi la nature des questions et le type de tâche, deux variables dont les modalités sont parfaitement décrites et contrôlées.

Cette analyse des différences entre filles et garçons démontre que cet indicateur de disparités doit être interprété essentiellement de manière relative et non de manière absolue. En d'autres termes, il convient d'analyser les différences entre filles et garçons au sein d'un pays comparativement aux différences observées dans les autres pays.

DISPARITÉS ENTRE "NATIFS" ET "NON NATIFS"

Parmi les indicateurs souvent mentionnés pour étudier l'équité d'un système éducatif figurent les disparités qui peuvent exister entre natifs et non natifs, nationaux et étrangers. Il existe différentes définitions de "l'étranger" et les législations nationales diffèrent à cet égard. Ainsi, dans certains pays, le fait d'être né sur le territoire national confère le statut de citoyen alors que dans d'autres pays, c'est la "loi du sang" qui est invoquée. Des possibilités plus ou moins importantes existent quant à l'acquisition de la nationalité et celles-ci sont effectivement utilisées de manière très variable, en fonction, par exemple, de la possibilité de conserver ou non sa nationalité à la naissance ou de l'intérêt que représente cette acquisition.

Outre les différences de législation, il existe aussi différents profils d'immigration. En raison des aides européennes accordées aux régions en difficulté économique, certains pays comme l'Irlande ont accueilli au cours de la dernière décennie une population hautement qualifiée. Il en résulte que ces élèves "allochtones" obtiennent en lecture un score moyen de 573 alors que les élèves "autochtones" n'obtiennent que 528 points. Dans d'autres pays, les immigrés appartiennent plutôt aux classes défavorisées et dès lors les différences de rendement peuvent varier considérablement.

Le rapport initial de l'OCDE distingue à cet égard, en fonction de leur lieu de naissance et du lieu de naissance de leurs parents, les autochtones (élèves nés dans le pays de l'évaluation dont au moins un des parents est né dans le pays), les élèves de première génération (élèves nés dans le pays de l'évaluation dont les parents sont nés à l'étranger) et les élèves allochtones (élèves nés à l'étranger

dont les parents sont nés à l'étranger). Il faut noter à ce sujet qu'aucune information relative à la nationalité n'est demandée et il est possible, au moins en théorie, d'observer des individus considérés comme "de première génération" alors que leurs parents, même s'ils sont nés à l'étranger, possèdent la nationalité du pays où ils sont testés, voire des élèves qualifiés "d'allochtones" alors qu'ils possèdent la nationalité du pays où ils sont testés.

Une autre information intéressante est par contre enregistrée : on demande aux élèves testés si la langue qu'ils parlent habituellement à la maison diffère ou non de la langue du test. A la diversité culturelle et socio-économique de l'immigration réelle s'ajoute également la fréquentation d'une école par des élèves d'un pays voisin. Ainsi, la Communauté française de Belgique accueille une population non négligeable d'élèves vivant sur le territoire luxembourgeois (qu'ils soient luxembourgeois ou non) qui souhaite éviter le plurilinguisme du système éducatif et une population d'élèves français dont les parents souhaitent, par exemple, échapper à la carte scolaire. Ces élèves, dans les analyses PISA 2000, sont considérés, en principe, comme "allochtones", au même titre que les enfants de fonctionnaires européens non belges ou les enfants de parents en séjour illégal sur le territoire ou demandeurs d'asile.

Une nouvelle analyse des données PISA 2000, en croisant la variable "pays de naissance" et "langue parlée à la maison" a été réalisée par les auteurs. Les catégories élèves "de première génération" et "élèves allochtones" ont été confondues pour des raisons de taille d'échantillon. Ce croisement aboutit à quatre catégories d'élèves : (i) "élèves autochtones qui parlent la langue du test à la maison", (ii) "élèves autochtones qui ne parlent pas la langue du test à la maison", (iii) "autres élèves qui parlent la langue du test à la maison" et (iv) "autres élèves qui ne parlent pas la langue du test à la maison". Le tableau 6 présente, pour chaque modalité, le pourcentage d'élèves concerné et la moyenne du groupe en lecture.

Tableau 6 : Proportion du groupe des élèves autochtones ou non selon qu'ils utilisent ou non la langue du test à la maison (ligne supérieure) et moyenne en lecture de ces différents groupes (ligne inférieure)

	<i>Elèves autochtones</i>		<i>Autres élèves</i>		<i>Elèves autochtones</i>	<i>Autres élèves</i>
	<i>Même langue</i>	<i>Autre langue</i>	<i>Même langue</i>	<i>Autre langue</i>		
Australie	73,15	3,85	9,70	13,29		
	534,73	494,51	536,79	507,15	531.55	519.32
Belgique	71,80	17,66	5,29	5,25		
	521,66	529,18	440,57	403,61	519.54	416.56
Canada	77,70	2,24	10,82	9,24		
	540,09	493,18	544,97	508,70	536.83	525.73
Suisse	75,33	4,51	5,80	14,36		
	515,63	480,68	476,38	407,92	512.65	426.99
France	88,21	1,04	6,75	4,00		
	512,88	471,39	481,52	441,36	510.82	463.67
Royaume-Uni	90,38	0,59	5,53	4,51		
	528,90	434,93	516,68	476,50	526.89	494.93
Italie	81,75	17,35	0,33	0,58		
	500,44	448,84	471,81	446,40	488.13	449.53
Luxembourg	63,07	2,98	8,21	25,73		
	475,58	409,26	407,10	378,21	469.88	385.10
Pays-Bas	75,78	13,03	5,94	5,25		
	545,78	524,38	481,65	458,68	541.09	463.80
Nouvelle-Zélande	79,76	1,38	10,00	8,86		
	541,38	454,74	549,75	470,24	534.83	506.77
Etats-Unis d'Amérique	84,23	2,25	4,98	8,54		
	514,05	412,73	519,72	445,79	509.57	472.02

En Australie, par exemple, quelque 40% des élèves considérés comme immigrés (9,70% de la population totale) parlent anglais à la maison. Ils obtiennent un score de 536,79, soit des résultats comparables aux élèves autochtones qui parlent l'anglais à la maison (73,15% de la population totale et un score moyen de 534,73) et des résultats supérieurs à ceux des élèves autochtones qui ne le parlent pas à la maison (3,85% de la population totale et un score moyen de 494,51). Dans ce contexte, il n'est guère étonnant que la différence de rendement entre élèves allochtones et élèves autochtones ne soit que de 20 points lorsqu'on ne tient pas compte de la langue parlée à la maison. Il en est de même dans les autres pays anglophones considérés (Nouvelle-Zélande, Royaume Uni et Etats-Unis d'Amérique). La situation de la Suisse et du Luxembourg, dans ce domaine, est totalement opposée.

Malheureusement, cette complexité n'est pas nécessairement prise en compte. Ainsi, dans le rapport initial de l'OCDE, on peut lire : "*Concern about such differences is*

especially justified in those countries with significant performance gaps and comparatively large percentages of first-generation students, including Belgium, France, Luxembourg and Switzerland. (...) Australia and Canada provide interesting exceptions to the general pattern. In Australia, neither a student's place of birth nor the birthplace of his or her parents seems, on average, to influence performance in reading, mathematical and science literacy. Australia achieves these results in the face of high proportions of non-native students and first-generation students. Similar patterns of success are also found in Canada as the three groups of Canadian students score statistically above the OECD average in two out of the three domains" (OCDE, 2001).

Cette analyse rapide – on aurait également pu tenir compte du statut socio-économique des migrants ou de la formation des parents - permet de mettre en évidence la complexité de la prise en compte de l'immigration comme facteur d'explication d'une partie des résultats ou, de façon plus globale, le danger de surinterprétation d'analyses bivariées.

DISPARITÉS ENTRE ÉTABLISSEMENTS SCOLAIRES

Le rapport initial de l'OCDE (2001) ainsi que les rapports thématiques publiés ultérieurement¹³ accordent une place non négligeable à l'effet des variables scolaires et aux indicateurs de disparité entre établissements. De ces analyses découlent des recommandations aux décideurs politiques, notamment en terme d'autonomie accordée aux établissements, pour ne citer qu'un exemple. Mais l'étude PISA est-elle réellement en mesure de dégager l'effet des variables scolaires ? Différents éléments invitent à la prudence dans l'interprétation des résultats.

En premier lieu, ces différentes analyses, qu'il s'agisse du coefficient de corrélation intraclasse¹⁴, d'analyses multi niveaux ou de l'effet plus global des variables scolaires, nécessitent l'adoption d'une définition précise et standardisée du concept d'école. Sans cette définition, les effets entre pays ne peuvent être comparés et recouvrent des réalités potentiellement différentes.

De manière schématique, l'école peut être considérée soit comme une entité administrative, soit comme un lieu géographique circonscrit, composé d'un seul

¹³ La liste des rapports et le texte de ceux-ci sont disponibles, au fur et à mesure de leur publication, sur le site de l'OCDE (<http://www.pisa.oecd.org/analyses/intro.htm>).

¹⁴ Coefficient de corrélation intra-classe : ce coefficient indique la part de la variance totale d'un score expliquée par la seule appartenance de chacun des individus à l'une ou l'autre des unités, par exemple, les écoles, qui sont prises en compte dans le modèle d'analyse de la variance. Ce coefficient sera d'autant plus proche de 1 que les élèves seront groupés dans des unités de moyennes distinctes et de faible dispersion. Il se rapprochera d'autant plus de 0 que les moyennes de différentes unités seront semblables et que la variance des notes à l'intérieur de chaque unité sera grande.

ou de plusieurs bâtiments interconnectés. La législation belge francophone différencie ces deux concepts sous les appellations "école" (ou entité administrative sous la responsabilité d'un directeur) et "implantation" (bâtiment ou ensemble de bâtiments). Une école peut comporter une ou plusieurs implantations alors qu'une implantation n'appartient qu'à une et une seule école. Cependant, il n'est pas rare qu'un complexe de bâtiments scolaires qui accueille un très grand nombre d'élèves puisse être scindé en deux unités administratives (donc, constitue deux écoles). Si la réalité présente déjà au sein d'un même pays une telle complexité, il est aisé d'imaginer la difficulté de définir un concept "école" comparable d'un pays à l'autre. De ce point de vue, PISA 2000 échappe à cette difficulté... en ne définissant pas le concept d'école, pourtant employé à travers l'ensemble des documents et la base de données (Adams et Wu, 2002).

En fonction de la définition employée par un pays, le coefficient de corrélation intra-classe ou, plus globalement, le pourcentage de variance des scores des élèves expliqué par l'appartenance à l'établissement scolaire, peut varier plus ou moins selon les modalités de regroupement des élèves. Malheureusement, le rapport technique de l'étude PISA 2000 reste silencieux sur la définition adoptée par chacun des pays participants.

On peut mettre en évidence ce type de problème en considérant la situation de deux pays proches : la France et la Communauté française de Belgique. Comme PISA s'intéresse à une population d'âge en fin d'obligation scolaire (au moins à temps plein), cela se traduit dans certains pays par une répartition des élèves dans différents types d'établissements scolaires. C'est notamment le cas en France : la scolarité secondaire inférieure est assurée dans les *Collèges* et la scolarité secondaire supérieure dans les *Lycées*. Ces deux types d'établissement sont généralement distincts alors qu'en Belgique, à l'opposé, les établissements d'enseignement secondaire couvrent le plus souvent l'ensemble des 6 années de scolarité au sein d'une même structure administrative.

La population des élèves de 15 ans devrait, en principe, fréquenter la première année de l'enseignement secondaire supérieur, en France, soit la seconde (première année du lycée). Par contre, les élèves qui ont redoublé une année se rencontrent dans la dernière année de l'enseignement secondaire inférieur, c'est-à-dire dans les collèges.

Il s'ensuit qu'un collège fréquenté uniquement par des élèves qui n'ont pas connu de retard scolaire ne devrait pas compter d'élèves de 15 ans et ne devrait donc pas figurer dans la liste des écoles au départ de laquelle l'échantillon français a été constitué. De même, un lycée fréquenté uniquement par des élèves "en retard", ne sera pas repris dans la population des écoles éligibles. En conséquence, l'échantillon d'écoles sélectionné, en France, ne représente ni la population de lycées, ni la population de collèges, alors qu'il représente bien la population des élèves de 15 ans. Si un coefficient de corrélation intra-classe est

estimé au départ de l'échantillon PISA pour les seuls collèges ayant participé à l'étude, il sera sous-estimé puisque les collèges fréquentés par les meilleurs élèves, ceux qui n'ont aucun retard, ne figurent pas dans la population de référence. Il en va de même pour les lycées qui ne comptent que des élèves en retard.

Pour matérialiser ces limites, les variances entre écoles et à l'intérieur de celles-ci ainsi que les coefficients de corrélation intra-classe (en littérature) ont été calculées pour la France et la Communauté française de Belgique. Comme il est impossible de créer des écoles qui proposent à la fois la scolarité secondaire inférieure et la scolarité secondaire supérieure en France, les établissements de la Communauté Française de Belgique ont été scindés, artificiellement, selon le niveau d'enseignement fréquenté par les élèves (grades inférieurs à 10 = "pseudo-collège", grade 10 et au-delà, "pseudo-lycée").

Le coefficient de corrélation intra-classe s'élève à 0,50 en France (collèges *et* lycées inclus) et à 0,57 en Communauté française de Belgique, si on garde la définition des écoles intacte. Cette faible différence entre les coefficients de corrélation intra-classe français et belge a conduit certains chercheurs à proclamer un peu rapidement l'inefficacité de la carte scolaire en France. En effet, lorsque les écoles belges sont scindées selon le niveau scolaire, de manière à reproduire la structure française des études, le coefficient de corrélation intra-classe s'élève à 0,64, soit un accroissement de 0,07. Dans cette simulation¹⁵, la faiblesse des élèves les plus en retard (Lafontaine et Demeuse, 2003) se matérialise par un accroissement des disparités entre établissements artificiellement constitués alors qu'elle est constitutive de la variance intra-classe lorsqu'on utilise les unités administratives réelles qui englobent les 6 années de scolarité secondaire en Belgique francophone¹⁶. Ce type de discussion n'a, en principe, pas lieu d'être lorsque c'est un échantillon de niveau scolaire qui est pris en compte, comme dans les études de l'IEA, mais ces dernières posent, par

¹⁵ La simulation est, par ailleurs, certainement optimiste quant à la situation belge francophone puisque les "pseudo-collèges" et les "pseudo-lycées" ne sont pas sélectionnés de manière indépendante, comme les collèges et les lycées français.

¹⁶ L'exemple qui vient d'être traité à travers une simulation n'est pas gratuit. Si la Communauté française de Belgique a adopté comme unité d'échantillonnage l'implantation scolaire (ensemble de bâtiments contigus et appartenant à une école) lors de PISA 2000, c'est l'école (unité administrative pouvant comporter plusieurs implantations distinctes) qui a été choisie lors de PISA 2003. On peut s'attendre à observer, de ce fait, une part de variance imputable à l'école plus faible en 2003 qu'en 2000, ce qui est plutôt problématique dans la perspective du programme de permettre le suivi des acquis des élèves à travers les différentes prises d'information.

contre, d'autres problèmes¹⁷. Il faut s'intéresser, notamment, aux différences d'âge des élèves au sein des écoles puisque les élèves d'un même niveau peuvent cumuler plus ou moins de retard, selon l'école qu'ils fréquentent.

Ce troisième exemple avait pour objectif de s'interroger sur la validité des indicateurs dits "d'effet établissement" dans le cadre de la recherche PISA. Comme nous l'avons montré, sans l'adoption d'une définition internationale du concept école, la pertinence relative ou comparative de ces indicateurs se voit partiellement altérée.

CONCLUSION

Les études internationales, comme celles menées sous l'égide de l'IEA ou de l'OCDE, apportent sans conteste une description essentielle des systèmes éducatifs en permettant "de prendre de la hauteur". Leur utilité et leur impact, tant au niveau de la connaissance des systèmes que dans les tentatives qui sont faites de les piloter, ne sont plus à démontrer. Mais il serait erroné de penser que ces études peuvent répondre à n'importe quelle question des décideurs politiques ou des scientifiques, quelle qu'en soit la formulation précise. De plus, ces études se déroulent dans un cadre méthodologique donné dont il faut tenir compte lors de l'interprétation des résultats. Les trois exemples, réanalysés dans cet article, démontrent la complexité des phénomènes et l'importance de ne pas s'engager dans des explications pédagogiques trop simplistes.

En premier lieu, il importe de valoriser l'approche comparative ou relative plutôt que l'approche absolue. Les analyses menées sur les indicateurs de disparité entre filles et garçons ont démontré que l'ampleur de l'effet est aussi fonction de paramètres méthodologiques comme le choix de la population cible, la nature des questions et le type de tâches demandées aux élèves. Cela n'enlève rien à l'intérêt du constat et ne saurait invalider le test sous un quelconque prétexte de "biais". Au contraire, des analyses plus poussées permettent de retirer des informations intéressantes quant à un éventuel fonctionnement différencié des élèves des deux sexes.

En second lieu, il convient de se prémunir contre les interprétations hâtives basées sur des analyses bivariées. La réalité éducative ne peut se réduire à une causalité simple entre une variable de rendement et un prédicteur. De telles analyses peuvent conduire à classer à tort des pays comme équitables et d'autres comme inéquitables. L'étude du concept "d'immigré" en constitue un exemple.

¹⁷ Par exemple, on mélange dans les classes qui connaissent le redoublement des élèves à l'heure et en retard qui appartiennent à des cohortes différentes et on confond généralement le niveau "classe" et le niveau "école" puisque c'est souvent une seule classe qui est prélevée dans chaque école (Demeuse, Crahay et Monseur, 2001).

Enfin, chaque enquête internationale se déroule dans un cadre méthodologique donné et précis qui limite les analyses secondaires. Outrepasser ces limites peut conduire à des aberrations ou à des conclusions non fondées. L'inefficacité de la carte scolaire en France proclamée au départ du coefficient de corrélation intra-classe constitue à cet égard un bel exemple didactique d'aveuglement idéologique et de méconnaissance méthodologique. Au-delà des trois exemples, analysés en détail, c'est très souvent la surinterprétation hâtive qui constitue, en définitive, le talon d'Achille des études internationales, bien plus que des faiblesses constitutives rédhibitoires.

BIBLIOGRAPHIE

- Adams, R. & Wu, M. (eds.) (2002) *PISA 2000 Technical Report*. Paris : Organisation for Economic Cooperation and development.
- Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzales, E.J., Smith, T.A., & Kelly, D.L. (1996) *Science Achievement in the Middle School Years : IEA's TIMSS*. Chestnut Hill, MA, Boston College.
- Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Kelly, D.L., & Smith, T.A. (1996) *Mathematics Achievement in the Middle School Year : IEA's TIMSS*. Chestnut Hill, MA: Boston College.
- Blondin, C. & Lafontaine, L. (à paraître) Différences liées au sexe : l'école en relation avec son environnement. Communication présentée lors du 16^e Colloque international de l'Admécé-Europe du 4 au 6 septembre 2003 sur le thème "Evaluation entre efficacité et équité". *Cahiers du SPE*, 15-16. Liège : Service de Pédagogie expérimentale de l'Université.
- Brusselmans-Dehairs, C., Henry, G.F. (Eds)(1994) *Gender and Mathematics. International Journal of Educational Research*, 21(4).
- Crahay, M. (1996) *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck.
- Demeuse, M., Crahay, C., Monseur, C. (2001) Efficiency and Equity. In W. Hutmacher, D. Cochrane, N. Bottani (eds) *In Pursuit vs Equity in Education*. Dordrecht : Kluwer Academic Publishers.
- Duru-Bellat, M., Mons, N., Suchaut, B. (2003) Contextes nationaux, organisation des systèmes éducatifs et inégalités entre élèves : l'éclairage de l'enquête PISA. *Politiques d'éducation et de Formation. Analyses et comparaisons internationales*, 9, 95-108.
- Groupe Européen de Recherche sur l'Équité des Systèmes Educatifs (2003) *L'équité des systèmes éducatifs européens. Un ensemble d'indicateurs*. Université de Liège : Rapport remis à la Commission européenne (Projet Socrates SO2-61OBGE) Ce rapport est disponible en français et en anglais sur le site officiel de la Commission européenne :
http://europa.eu.int/comm/education/programmes/socrates/observation/equality_fr.pdf
 (version française),
http://europa.eu.int/comm/education/programmes/socrates/observation/equality_en.pdf
 (version anglaise).
- Kirsch, I., de Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J., Monseur, C. (2003) *La lecture, moteur de changement. Performances et engagements d'un pays à l'autre. Résultats du cycle d'enquête de PISA 2000*. Paris : OCDE.

- Lafontaine, D. (1996) *Performances en lecture et contexte éducatif : enquête internationale menée auprès d'élèves de 9 et 14 ans*. Paris ; Bruxelles : De Boeck Université.
- Lafontaine, D. (2001) Quoi de neuf en littérature ? : Regard sur trente ans d'évaluation de la lecture. *Cahiers du Service de Pédagogie expérimentale*, (7-8), 71-95.
- Lafontaine, D. & Blondin, C. (à paraître) *De 1964 à 2001 : regard sur les résultats des élèves en communauté française*. Bruxelles : De Boeck.
- Lafontaine, D. & Demeuse, M. (2002) Le bon (critique), la brute (médiatique) et les truands (anglo-saxons). (Controverse. L'enquête OCDE sur les acquis des élèves en débat) Le bon (critique), la brute (médiatique) et les truands (anglo-saxons). *Revue nouvelle*, 115(3-4), 100-108.
- Lafontaine, D. & Demeuse, M. (2003) Le parcours scolaire fait toute la différence. *Wallonie*, 76, 82-91.
- Lafontaine, D. & Monseur, C. (à paraître) Gender differences : effect of the framework and of the test composition.
- Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., O'Connor, K.M. (2000) *TIMSS 1999 International Science Report : Findings from the IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. International Study Center, Lynch School of Education, Boston College.
- Monseur, C. (1997) Résultats à une enquête internationale en mathématiques. *Mathématique et pédagogie*, 114, 5-12.
- Monseur, C. & Demeuse, M. (1998) Apports des études internationales à la réflexion sur la qualité des systèmes d'enseignements nationaux. Une analyse de l'éducation scientifique en Communauté française de Belgique. *Bulletin de la Société Royale de Sciences de Liège*, 67(5), 261-280.
- OCDE (1999) *Mesurer les connaissances et les compétences des élèves : Un nouveau cadre d'évaluation*. Paris : OCDE.
- OCDE (2000) *Mesurer les connaissances et les compétences des élèves : Lecture, mathématique et sciences : l'évaluation PISA 2000*. Paris : OCDE.
- OCDE (2001) *Connaissances et compétences : des atouts pour la vie – Premiers résultats de PISA 2000*. Paris : OCDE.