
ÉTUDE DE L'INFLUENCE DU FORMAT DE QUESTION ET DE LA DÉFINITION DE LA POPULATION SUR LES DISPARITÉS GARÇONS-FILLES EN LECTURE

Dominique Lafontaine
Christian Monseur

Service de pédagogie expérimentale - Université de Liège - Belgique

1. PROBLÉMATIQUE

Il est aujourd'hui devenu de pratique courante de publier des indicateurs d'efficacité et ou d'équité des systèmes scolaires en se fondant sur des études à large échelle, à caractère national ou international. La plupart des pays industrialisés voient ainsi publiés, de manière régulière, des « tableaux de bord », des « états de l'école », des « chiffres-clés » ou des « regards sur l'éducation ». L'importance de telles publications, dans une perspective de pilotage, n'est plus à démontrer. Le risque existe cependant, vu le caractère ramassé de ces publications¹, que le produit (l'indicateur) soit pris pour argent comptant et qu'il n'acquière ainsi, aux yeux des utilisateurs potentiels, une valeur absolue, alors que sa valeur n'a que le sens que lui confèrent l'enquête et les différents gestes techniques dont l'indicateur est l'aboutissement. Aussi bien dans le chef de certains experts que dans le grand public, on se trouve confronté à un risque sérieux de sur-interpréter les données. Le caractère parfois contradictoire ou divergent de certains indicateurs issus d'enquêtes différentes vient heureusement rappeler la prudence nécessaire et invite à se pencher sur l'impact que peuvent avoir certaines caractéristiques des enquêtes sur les indicateurs qui en sont tirés.

2. OBJECTIFS

Nous avons décidé d'étudier cet impact en nous intéressant :

- aux indicateurs d'équité garçons-filles,
- dans le domaine de la lecture ou littérature,

¹ C'est le principe même de ces publications de condenser des informations quantitatives (les indicateurs) sans s'attarder sur le détail des études sources qui ont servi à leur élaboration.

- dans les deux grandes enquêtes internationales les plus récentes menées dans le domaine pour une population d'élèves de 14-15 ans : IEA Reading Literacy (1991) (Elley, 1994) et PISA 2000 (Ocdé, 2001 ; Kirsch et al., 2003).

Est plus particulièrement analysé l'impact des caractéristiques suivantes:

- le choix du matériel de lecture (en particulier des types de textes) et les démarches évaluées ;
- le format des items (choix multiples ou question ouverte) ;
- la définition de la population.

3. MÉTHODOLOGIE

Différentes approches méthodologiques ont été utilisées.

- approche descriptive : les caractéristiques du cadre et des modalités d'évaluation des deux enquêtes sont présentées ainsi que les indicateurs d'équité garçons-filles issus de celles-ci ;
- dans PISA, les différences garçons-filles ont été estimées séparément pour les questions à choix multiple et pour les questions ouvertes, et standardisées. On peut estimer, sur cette base, dans quelle mesure l'indicateur d'équité garçons-filles est affecté par la présence de questions ouvertes (calcul d'un pourcentage d'accroissement de la différence) ;
- afin d'étudier l'impact de la définition de la population, on a sélectionné dans PISA les élèves à l'heure dans leur parcours scolaire. De la sorte, on peut tenter d'estimer dans quelle mesure la définition de la population - en termes d'âge dans PISA - affecte l'ampleur des disparités garçons-filles².

4. DESCRIPTION COMPARATIVE DES DEUX ÉTUDES

Pour cette description, nous nous en tiendrons aux seules caractéristiques dont l'impact sera étudié ici³. Les deux études, menées à près de 10 ans d'intervalle, la première sous les auspices de l'IEA, la seconde sous celles de l'Ocdé, portent sur la compréhension en lecture (*reading literacy*).

² Il serait bien sûr tout à fait intéressant d'opérer le même type d'analyse pour d'autres indicateurs d'équité, comme ceux en fonction de l'origine sociale ou ethnique.

³ Pour une comparaison plus complète, on peut se reporter à Lafontaine (2001) ou aux rapports des enquêtes (Elley, 1994 ; Lafontaine, 1996), Ocdé, 1999, Ocdé, 2001).

4.1. Définition de la population

	IEA Reading Literacy (1991)	PISA 2000
Définition de la population	Grade fréquenté par la majorité des élèves âgés de 14 ans	Elèves âgés de 15 ans

Les élèves de PISA sont un peu plus âgés que ceux de l'étude IEA, mais là n'est pas la différence la plus importante de notre point de vue. Pour tous les pays qui pratiquent la promotion automatique (soit environ la moitié des pays participants à PISA), le choix de l'une ou l'autre définition de la population, en termes de grade ou en termes d'âge, n'a pas d'impact : tous les élèves d'un âge donné y fréquentent en effet le même grade.

Pour les pays qui pratiquent le redoublement, en revanche, l'impact peut être important sur tous les indicateurs de disparités. En effet, dans le premier cas de figure (choix d'un grade), les élèves en retard scolaire sont dans le même grade que les élèves à l'heure, ce qui a tendance à contenir dans certaines limites les écarts de performances entre les uns et les autres. Dans le second cas de figure, les élèves (ici de 15 ans) en retard scolaire ont une ou plusieurs années de scolarité en moins que les élèves à l'heure et l'écart entre les uns et les autres a donc tendance à se creuser.

Du point de vue qui nous occupe ici, les disparités de performances entre les garçons et les filles, on peut faire l'hypothèse que le choix d'une définition de la population en termes d'âge dans PISA aura tendance à amplifier les écarts garçons-filles, car les garçons et les filles sont inégalement frappés par l'échec scolaire.

4.2. Choix des textes support pour les tâches de lecture

Les deux études ont pris le parti d'utiliser comme supports des textes continus et non continus, et à l'intérieur des textes continus, une diversité de genres de textes. Mais, comme l'illustre le tableau suivant, la répartition dans les diverses catégories diffère quelque peu dans les deux études.

Proportion d'items portant sur les types de textes suivants

	IEA Reading Literacy (1991)	PISA 2000
Textes continus		
- narratifs	33 %	13 %
- informatifs/descriptifs	29 %	35 %
- argumentatifs/injonctifs	0 %	18 %
Textes non continus	38 %	34 %
Total	100 %	100 %

Comme on peut le voir, la proportion de textes non continus et de textes informatifs est sensiblement la même dans les deux études. Il n'y a pas de textes argumentatifs dans IEARL et la proportion de textes narratifs est sensiblement moindre dans PISA. Par ailleurs, les textes non continus dans IEARL comportent moins de texte que ceux de PISA. Alors que dans PISA, la plupart des textes non continus combinent un tableau, graphique ou schéma avec des portions de texte parfois substantielles, les documents d'IEARL ne comportent qu'un élément graphique ou schématique, avec un minimum de texte.

Dans IEARL, les résultats sont présentés sur 3 échelles : textes narratifs, informatifs et non continus (documents). Dans PISA, on ne dispose que de deux échelles, l'une pour les textes continus, l'autre pour les textes non continus.

4.3. Types de démarches évaluées

Dans les deux études, l'option a été prise d'évaluer plusieurs démarches ou processus de traitement, différant notamment par le degré d'inférence exigé. Celles-ci ne sont toutefois pas définies de la même manière ou avec la même rigueur et il est donc malaisé d'opérer une comparaison stricte. Pour cette comparaison, nous avons réparti les items d'IEARL dans les catégories de démarches de PISA⁴.

Proportion d'items portant sur les démarches suivantes

	IEA Reading Literacy (1991)	PISA 2000
Retrouver ou restituer de l'information (sous forme littérale ou paraphrase)	42 %	30 %
Interpréter (inférer, trouver l'idée principale)	58 %	50 %
Réfléchir et évaluer	0 %	20 %
Total	100 %	100 %

Dans PISA, on dispose de trois échelles – une par démarche. Dans IEARL, les résultats en fonction de la démarche n'ont pas été distingués.

4.4. Format des questions

L'une des différences les plus marquantes entre les deux études réside dans la proportion de questions à choix multiples et de questions ouvertes.

⁴ Cette répartition s'est effectuée en fondant les catégories d'IEARL (localiser l'information, trouver l'idée principale...) dans celles de PISA. Nous n'avons pas opéré une re-catégorisation fine par item.

Proportion d'items par format de question

	IEA Reading Literacy (1991)	PISA 2000
QCM	75 %	45 %
Question ouverte à réponse courte	22 %	11 %
Question ouverte à réponse construite	3 %	45 %
Total	100 %	100 %

Toutes les questions d'IEARL intervenant dans le calcul du score total sont à réponse objective (les deux questions ouvertes à réponse construite constituaient une option internationale n'entrant pas dans le score total). Une seule réponse correcte est admise, que ce soit sous forme de QCM ou de réponse ouverte (qui se résume souvent à un seul chiffre ou un seul mot).

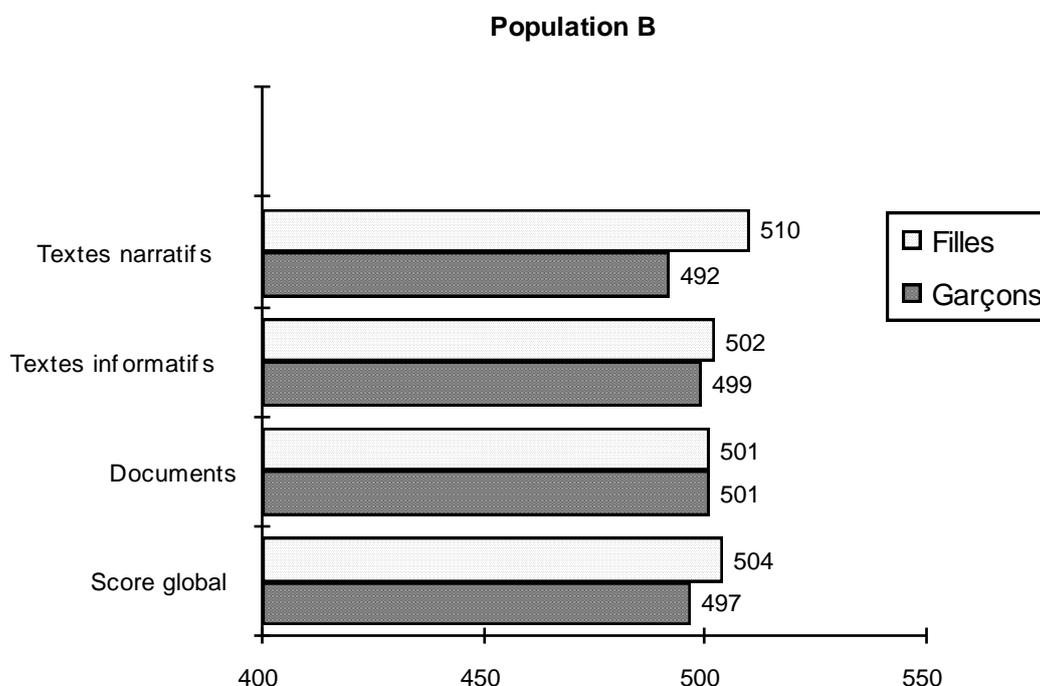
Dans ce cas, les données de la littérature permettent de formuler une hypothèse précise. D'après la synthèse dressée par Bennett (1993), « plusieurs études ont montré que comparativement aux garçons, les filles ont de meilleures performances dans les réponses construites que dans les questions à choix multiple » (Bennett, 1993, p. 20). Une étude de Mazzeo *et al.* (1991) montre que cet effet se vérifie quels que soient la discipline (Sciences, sciences sociales, sciences humaines), la longueur de la réponse (courte ou construite), le type de codage (analytique ou global) et l'origine ethnique des élèves (Américains blancs, noirs ou asiatiques).

Les rapports internationaux relatifs à PISA n'ont pas distingué les résultats en fonction du format des questions. C'est l'un des objets de la présente étude.

5. INDICATEURS D'ÉQUITÉ GARÇONS-FILLES ISSUS DES DEUX ÉTUDES

Ecarts de performances entre garçons et filles à 14 ans (Pop. B) en fonction du type de texte.

Source : I.E.A. Reading Literacy, 1991



Ecarts de performances médians standardisés entre garçons et filles sur les différentes échelles de lecture.

Source : PISA 2000 (Kirsch *et al.*, 2002)

Echelle combinée	0.32
Retrouver de l'information	0.23
Interpréter	0.28
Réfléchir et évaluer	0.41
Textes continus	0.39
Textes non continus	0.17

À l'évidence, les différences de performances entre les garçons et les filles sont plus marquées en 2000 dans PISA qu'en 1991 dans IEARL⁵. On voit dès lors

⁵ On peut aussi à titre complémentaire, comparer les résultats de IEA Reading Literacy et ceux de l'étude PIRLS (2001) pour les élèves de 10 ans. Dans IEARL, l'écart filles-garçons pour le score global est de 12 points. Il est plus important pour les textes narratifs (19 points) que pour les textes informatifs (12 points) et les documents (3 points). Dans PIRLS, l'écart filles-garçons pour le score global est de 20 points et est sensiblement plus élevé pour les tâches à orientation littéraire (21 points) que pour celles à but informatif (18 points). Toutes les questions d'IEARL sont des QCM ou des questions courtes à réponse objective. Dans PIRLS, la moitié des points de score vient de QCM, l'autre moitié de questions ouvertes à réponse construite. On constate donc aussi une augmentation de la différence garçons-filles pour les élèves de 9-10 ans entre les deux études, dans laquelle l'introduction de questions ouvertes joue sans doute un rôle.

toute l'importance que revêt la présente investigation, pour tenter de discerner ce qui, dans l'amplification manifeste des différences, peut tenir aux caractéristiques respectives des deux enquêtes et ce qui peut relever de l'évolution historique⁶.

Dans IEARL, la différence n'est significative que dans 13 pays sur 31 et elle est en moyenne de faible ampleur : 7 points de score sur l'échelle standardisée (moyenne 500, écart type 100). Dans PISA, l'écart moyen (en faveur des filles) est de 32 points ; la différence est significative dans les 32 pays participants.

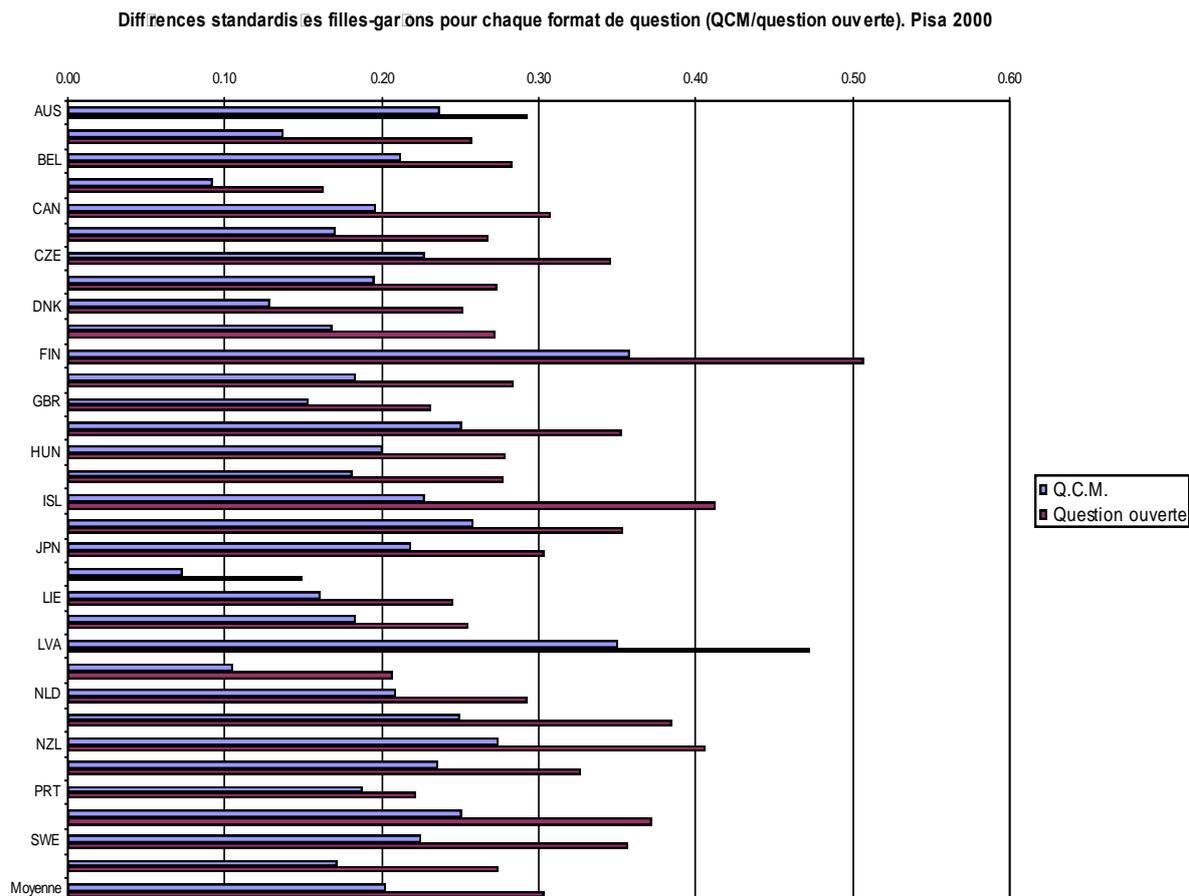
L'examen des résultats sur certaines sous-échelles permet de baliser la réflexion. Dans les deux études, on constate que l'écart garçons-filles est moindre pour les textes non continus (0 point dans IEARL ; 0.17 pour les textes non continus, 0.39 pour les textes continus dans PISA, Dans IEARL, c'est pour les textes narratifs que la différence est la plus nette avec 18 points d'écart. Dans PISA, la distinction narratif/informatif n'a pas été faite. Si la dimension type de texte avait une influence prépondérante propre, abstraction faite d'autres paramètres des tâches (format de réponse, démarche évaluée...), on aurait pu s'attendre, vu la proportion moindre de textes de fiction dans PISA, à voir les garçons sensiblement « avantagés ». Or il n'en est rien, que du contraire. L'action en sens opposé d'autres paramètres est donc sans doute plus importante. L'introduction de questions de réflexion et d'évaluation n'est probablement pas étrangère au phénomène. On voit en effet que c'est pour ce type de démarche – absent dans IEARL – que les écarts garçons-filles sont les plus marqués dans PISA (41 points, contre 23 et 28 points d'écart respectivement pour les démarches Retrouver et Interpréter).

6. ANALYSES ET RÉSULTATS

6.1. Impact du format des questions

Pour tenter d'estimer cet impact, dans PISA, on a estimé séparément les différences garçons-filles pour les questions à choix multiple et pour les questions ouvertes. Le graphique ci-dessous présente ces différences standardisées.

⁶ On ne peut, en effet exclure que l'écart entre les garçons et les filles en lecture se soit creusé au cours des dix dernières années.



Il existe dans tous les pays participants une interaction entre le sexe et le format de questions. La différence de performances entre les garçons et les filles est en effet moindre pour les questions à choix multiples et plus importante pour les questions ouvertes. Le choix de privilégier tel ou tel format de question, leur dosage respectif dans différentes études a donc un impact non négligeable sur les indicateurs d'équité garçons-filles. Selon les pays, le pourcentage d'accroissement de la différence garçons-filles lorsque l'on passe des questions à choix multiples aux questions ouvertes oscille entre 18 et 103 %. Il est en moyenne de 53,6 % pour l'ensemble des pays participants, ce qui est loin d'être négligeable.

6.2. Impact de la définition de la population

Dans la base de données PISA, on s'est attaché à quantifier le pourcentage d'élèves à l'heure dans l'échantillon des différents pays, afin de déterminer si le pays pratique ou non la promotion automatique. On a ensuite calculé la proportion de filles dans l'échantillon de PISA et la proportion de filles dans le sous-échantillon des élèves à l'heure.

Pourcentage d'élèves à l'heure dans l'échantillon initial, proportion de filles dans l'échantillon initial et dans un sous-échantillon d'élèves à l'heure (PISA 2000)

Pays	Pourcentage d'élèves à l'heure dans l'échantillon initial	Pourcentage de filles dans l'échantillon initial	Pourcentage de filles dans l'échantillon d'élèves à l'heure	Différence de pourcentage de filles entre les 2 échantillons
<i>Belgique</i>	65	48	52	+4
Danemark	89	50	50	0
<i>Espagne</i>	72	51	54	+3
Finlande	89	51	52	+1
<i>France</i>	53	51	54	+3
Islande	98	50	51	+1
<i>Italie</i>	76	49	52	+3
Japon	100	51	51	0
Corée	98	44	44	0
<i>Liechtenstein</i>	78	50	53	+3
<i>Mexique</i>	50	50	56	+6
Norvège	97	49	49	0
Nouvelle Zélande	88	50	50	0
Suède	97	49	50	+1
Moyenne	82	50	51	1.79

Légende : les pays qui pratiquent régulièrement le redoublement sont en italiques

Dans les pays qui pratiquent la promotion automatique et où le redoublement d'année constitue dès lors une mesure exceptionnelle, la proportion de filles dans l'échantillon initial et dans l'échantillon d'élèves à l'heure est identique ou quasi identique (pas plus d'un pourcent de différence). Pour ces pays, le choix d'une définition de la population en termes d'âge (comme dans PISA) ou en termes de grade (comme dans IEARL) ne devrait avoir que peu d'impact sur les différences garçons-filles.

En revanche, dans les pays où le redoublement constitue une pratique habituelle, la proportion de filles à l'heure dans leur parcours scolaire est de 3 à 6 % plus importante que celle des garçons. Le pourcentage de filles passe de 50 % à 54 % lorsqu'on ne prend en compte que les élèves à l'heure. Pour ces pays, le choix d'une définition de la population en termes d'âge devrait accentuer les indices de disparités garçons-filles, car davantage de garçons sont en retard dans leur parcours scolaire.

Pour estimer dans quelle mesure les indicateurs d'équité sont affectés par la définition de la population choisie, on a calculé, dans PISA, les écarts de performances garçons-filles pour l'échantillon original et pour le sous-échantillon d'élèves à l'heure⁷. Notre objectif était de procéder de même pour

⁷ On ne peut hélas faire cette comparaison que pour les 15 pays pour lesquels on a pu estimer, de façon fiable, compte tenu des informations disponibles, quelle proportion d'élèves se trouvent à l'heure. Il s'agit des pays pour lesquels, sur la base de la date de naissance, il est possible de déterminer quel grade devrait fréquenter

IEARL et de comparer l'ampleur de la différence entre échantillon original et sous-échantillon d'élèves à l'heure dans les deux études, l'effet du choix de la population en termes d'âge ou de grade se trouvant neutralisé dans ce dernier sous-échantillon. Malheureusement, la date de naissance des élèves ne figure pas dans la base de données IEARL et ce travail s'est donc avéré impossible.

Différences d'écarts de performances entre garçons et filles entre l'échantillon original et l'échantillon d'élèves à l'heure, selon que le pays pratique ou non la promotion automatique.

Source : PISA 2000

Pays pratiquant la promotion automatique	Écarts de performances garçons-filles	Pays pratiquant le redoublement	Écarts de performances garçons-filles
Danemark	-0.01	Belgique	-0.13
Finlande	-0.01	Espagne	-0.13
Islande	-0.01	France	-0.13
Japon	0.00	Italie	-0.05
Corée	0.00	Liechtenstein	-0.05
Norvège	-0.01	Mexique	-0.14
Nouvelle Zélande	-0.04		
Suède	-0.02		
Moyenne	-0.011	Moyenne	-0.11

Pour les pays pratiquant la promotion automatique, sans surprise, l'impact du choix de la population est minime (- 0.011). En revanche, pour les pays qui pratiquent le redoublement, l'impact est plus net : il va de -0.05 à - 0.14 et est de - 0.11 en moyenne. L'écart moyen garçons-filles, qui était, dans ce petit groupe de pays, de 0.32 au départ, passe ainsi à 0.21, soit une réduction de plus d'un tiers de l'écart (réduction de 35,5 %).

Lorsque l'on compare l'ampleur des différences filles-garçons en lecture parmi les élèves à l'heure et tous les élèves d'un groupe d'âge, on constate que les différences sont plus importantes dans le groupe complet ; ceci est dû au fait que les garçons sont davantage affectés par le redoublement. Rappelons qu'il a été impossible de mener l'investigation projetée à son terme en mesurant si la réduction des différences garçons-filles pour le groupe d'élèves à l'heure était de la même ampleur ou d'une ampleur inférieure dans le cas d'une définition de la population en termes de grade fréquenté, comme dans IEARL. La comparaison effectuée pour PISA n'apporte qu'un élément de réponse à la question soulevée. En effet, la toute grande majorité des études de rendement portent soit sur un grade, soit sur un âge. Il est très rare que l'on prenne comme population cible uniquement les élèves à l'heure.

les élèves à l'heure dans leur parcours scolaire. On ne peut l'estimer pour les pays où les élèves à l'heure se répartissent sur deux grades.

7. CONCLUSIONS

Cette étude a mis en évidence une influence assez nette de certaines des caractéristiques de l'évaluation sur l'ampleur des disparités filles-garçons en lecture. Le choix du matériel de lecture, les démarches évaluées, la proportion de questions à choix multiples et de questions ouvertes ont un impact démontré sur l'ampleur des différences en fonction du sexe. La définition de la population en termes d'âge ou de grade est une autre source potentielle d'influence, mais son impact n'a pu être correctement estimé pour des raisons techniques. Il est en tout cas établi que les différences garçons-filles sont plus réduites dans un groupe d'élèves à l'heure que dans un groupe d'âge complet où certains élèves sont à l'heure et d'autres en retard scolaire.

La prudence dans l'interprétation des indicateurs d'équité s'impose donc. Une augmentation ou une diminution de ces écarts filles-garçons dans le temps peut tenir en partie aux caractéristiques du cadre d'évaluation. On ne rappellera jamais assez que les indicateurs d'équité sont relatifs à l'étude qui les a produits et qu'il est hasardeux de les prendre pour argent comptant.

8. RÉFÉRENCES

- Benett, R.E. (1993). On the meaning of constructed response. R.E. Benett & W.C. Ward (Eds). *Construction versus choice in cognitive measurement. Issues in constructed response, performance testing, and portfolio assessment*. Hillsdale : Lawrence Erlbaum Associates. &-29.
- Elley, W.B. (1994). *The IEA Study of Reading Literacy : achievement and instruction in thirty-two school systems*. Londres : Pergamon.
- Kirsch, I., de Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J. & Monseur, C. (2003). *La lecture, moteur de changement. Performances et engagement. Résultats de PISA 2000*. Paris : Ocdé.
- Lafontaine, D. (1996). *Performances en lecture et contexte éducatif. Enquête internationale menée auprès d'élèves de 9 et 14 ans*. Bruxelles : De Boeck.
- Lafontaine, D. (2001). Quoi de neuf en littérature ? Regard sur trente ans d'évaluation de la lecture. *Les Cahiers du Service de Pédagogie expérimentale*, 7-8, 71-95.

Lafontaine, D. (2003). L'engagement des jeunes de 15 ans à l'égard de la lecture : un atout pour la littératie. *Caractères*, 10.

Mullis, I.V.S., Martin, M.O., Fierros, E.G., Goldberg, A.L. & Stemler, S.E.. (2000). *Gender differences in achievement. IEA's third international mathematics and science study*. International Association for the Study of Educational Achievement.

OCDÉ (1999). *Mesurer les connaissances et les compétences des élèves. Un nouveau cadre d'évaluation. PISA*. Paris : OCDÉ.

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

