

REVUE FRANÇAISE
DE
PÉDAGOGIE

Revue française de pédagogie

Recherches en éducation

157 | octobre-décembre 2006

PISA : analyses secondaires, questions et débats
théoriques et méthodologiques

Les évaluations internationales et la recherche en éducation : principaux objectifs et perspectives

Large-scale international comparative achievement studies in education: their primary purposes and beyond

Las valuaciones internacionales y la investigación en educación : principales objetivos y perspectivas

Internationale Vergleichende Analysen und Forschung in der Erziehung: Hauptziele und Perspektiven

Rolf V. Olsen et Svein Lie

Traducteur : Marcel Rieu



Édition électronique

URL : <http://journals.openedition.org/rfp/393>

DOI : 10.4000/rfp.393

ISSN : 2105-2913

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 1 décembre 2006

Pagination : 11-26

ISBN : 978-2-7342-1075-7

ISSN : 0556-7807

Référence électronique

Rolf V. Olsen et Svein Lie, « Les évaluations internationales et la recherche en éducation : principaux objectifs et perspectives », *Revue française de pédagogie* [En ligne], 157 | octobre-décembre 2006, mis en ligne le 01 décembre 2010, consulté le 20 avril 2019. URL : <http://journals.openedition.org/rfp/393> ; DOI : 10.4000/rfp.393

Les évaluations internationales et la recherche en éducation : principaux objectifs et perspectives

Rolf V. Olsen, Svein Lie

*Traduit de l'anglais par Marcel Rieu
avec la collaboration de la rédaction*

Cet article propose une vue structurée de l'ensemble des perspectives de recherches primaires et secondaires en lien avec les études internationales de grande ampleur (PISA et TIMSS). Le but de ces études est tout d'abord examiné avant de décrire de quelle manière leur objectif principal a peu à peu évolué de la recherche en éducation à la constitution de ressources d'indicateurs pédagogiques de contrôle et de développement des politiques nationales. Cette évolution a largement ouvert la voie à d'abondantes possibilités d'analyses secondaires des données en vue de leur diffusion au public. Dans sa majeure partie l'article est dédié à une description des perspectives variées d'analyses secondaires pertinentes et pour chacune de ces perspectives, un certain nombre d'études secondaires est brièvement présenté et des prolongements de recherche proposés. Notre intention est de démontrer le fort potentiel de recherches complémentaires contenu dans ces bases de données internationales.

Descripteurs (TEE) : éducation comparée, évaluation, indicateur, politique de l'éducation, recherche en éducation.

INTRODUCTION

L'objectif de cet article est de démontrer que les études comparatives de grande ampleur à l'échelle internationale sur les résultats en éducation ne comptent pas seulement parmi les outils de décisions les plus importants au plan politique dans de nombreux pays. Elles fournissent aussi des données uniques qui devraient inciter les chercheurs en sciences de l'éducation du monde entier à entreprendre toute une série d'analyses secondaires. Nous consacrerons la

première partie de l'article à l'évolution progressive de l'objectif principal de ces études. Puis nous nous pencherons sur le problème des relations qui existent entre ces études et la recherche pédagogique. Et nous présenterons enfin une vue d'ensemble des types d'analyses secondaires que l'on peut produire à partir des données fournies par ces études.

Les références et exemples concrets ici utilisés sont principalement liés aux études impliquant les sciences qui sont notre principal intérêt de recherche. Le propos de cet article ne se cantonne cependant pas à ce

seul sujet d'étude et les arguments seront tout aussi pertinents pour des études internationales effectuées dans d'autres domaines. Dans un souci de simplification, nous utiliserons donc, dans le cadre cet article le terme d'« études comparatives » (ou d'« études » lorsque le sens est évident) pour indiquer les études comparatives à l'échelle internationale portant sur les résultats en éducation. Nous ferons de plus fréquemment référence à deux études : TIMSS (*Trends in International Mathematics and Science Study* ; traduire) et PISA (*Programme for International Student Assessment* ; « Programme international pour le suivi des acquis des élèves ») qui servent d'exemples de ce genre d'études.

Quelles sont les caractéristiques de ces études ? Études à grande échelle, leur objectif est d'établir un système de mesures généralisables à toutes les écoles et tous les systèmes scolaires. Afin d'obtenir des mesures fiables et donc généralisables et, compte tenu du grand nombre d'écoles, de classes et d'élèves, des procédures rigoureuses d'échantillonnage sont requises, comme l'utilisation d'une méthode permettant les comparaisons à l'échelon international, de telle sorte que les différences entre pays – essentielles dans le cadre d'une enquête internationale – puissent être étudiées en tant que facteurs contextuels.

Bien que les objectifs de ces études, l'emploi de leur données, leur organisation, et la méthodologie appliquée aient progressivement évolué, nous soutiendrons que deux visions globales différentes et partiellement en compétition les sous-tendent. Ces études ont d'abord été conçues comme une méthode ou comme un projet précis pour conduire des recherches en éducation selon une perspective comparatiste internationale. Nous appellerons cette idée de départ « objectif I : objectif de recherche ». L'intérêt des décideurs politiques a progressivement déplacé le débat sur une observation des résultats des systèmes éducatifs et une étude des facteurs possibles agissant sur ces résultats. Cette logique d'étude comparative sera appelée « Objectif II : objectif de politique d'efficacité ».

Les appellations « Objectif I » et « Objectif II » ne sont utilisées que comme éléments heuristiques pratiques afin de comprendre quelques-unes des pressions idéologiques auxquelles ces études sont soumises ; nous nous inspirons ici de la façon dont Roberts (à paraître) utilise les termes « Vision I » et « Vision II » dans son étude du concept des connaissances scientifiques. Toutefois, l'usage de cette dichotomie n'implique nullement, de notre part, qu'« objectif de recherche » et « objectif de politique réelle » soient

incompatibles. Au contraire, les études peuvent être considérées comme des terrains où les chercheurs et les décideurs peuvent échanger des idées, nourrir leur intérêt mutuel pour les problèmes d'éducation et accepter l'engagement de chacun, tant au niveau national qu'international.

L'OBJECTIF PRINCIPAL DES ÉTUDES COMPARATIVES

Objectif I : l'objectif de recherche

Aujourd'hui, l'appellation « études comparatives en éducation » fait référence à différents types de recherches allant de questions d'ordres plus philosophique et méthodologique dans la comparaison interculturelle à des travaux spécifiques portant sur des aspects plus spécifiques de l'éducation selon les pays, les régions ou les salles de classes. De manière générale, cette appellation couvre également les études de conceptions et de grandeurs très différentes (voir par exemple Alexander *et al.*, 1999 & 2000). L'idée d'études comparatives de grande ampleur dont il est ici question a été concrétisée et définie comme un calendrier de recherche lors de la création, en 1961, de l'IEA (*International Association for the Evaluation of Educational Achievement* ; Association internationale pour l'évaluation des résultats éducatifs), sous les auspices de l'Institut pour l'éducation de l'UNESCO (Husén & Tuijnman, 1994 ; Keeves, 1992). L'idée fondamentale des fondateurs de l'IEA est très clairement exprimée par l'un d'eux, Torsen Husén (1973, p. 10) : « Nous, les chercheurs qui avons décidé de coopérer pour élaborer des instruments d'évaluation valables à l'échelle internationale, concevions le monde comme un grand laboratoire d'éducation, dans lequel une grande variété de pratiques, en termes de structures scolaires et de *curriculum* sont mises en œuvre. Nous voulions simplement tirer profit de cette variabilité internationale pour à la fois étudier les résultats des différents systèmes éducatifs et les facteurs explicatifs de leurs différences de résultats. »

Le terme de « laboratoire » est ici métaphorique puisque les conditions de laboratoire pour des expériences contrôlées ne sont pas remplies en recherche pédagogique pour des raisons à la fois éthiques et pratiques. Une alternative serait donc de concevoir des enquêtes dans lesquelles les variables pertinentes seraient étudiées sous un très grand nombre de conditions différentes. De cette façon « les différences observées entre les systèmes éducatifs donne-

raient la possibilité d'examiner l'impact des différentes variables sur les résultats scolaires » (Bos, 2002, p. 5). « Ainsi les études sont abordées dans une perspective de recherche [...] comme ayant des implications politiques » (Kellaghan & Greaney, 2001, p. 92). En d'autres termes, l'hypothèse est soutenue que l'organisation et les pratiques éducatives ont des effets sur les possibilités (ou les potentiels ou chances d'éducation) autant que sur les résultats. Ce pourrait être le but de recherches empiriques « d'aller au-delà d'une identification purement descriptive des facteurs visibles qui rendent compte des différences entre pays pour expliquer comment ils opèrent ? Telle est l'ambition qui prévaut généralement dans les sciences sociales : expliquer, prédire et dégager des généralisations. » (Husén, 1973, p. 10-11)

Ces deux citations de Husén doivent être envisagées comme typiques de cette époque et de l'optimisme qui y règne sur la question de savoir comment les sciences sociales pourraient contribuer au développement d'une meilleure compréhension des relations causales entre les différents types de facteurs sociaux. Aujourd'hui, les chercheurs en sciences sociales sont probablement moins enclins à utiliser des phrases du genre « facteurs qui causaient des différences » ou « l'ambition [...] dans les sciences sociales en général [...] est de prédire, ainsi que de dégager des généralisations ».

Objectif II : objectif de politique efficace

Il est demandé aux décideurs politiques d'établir des projets globaux pour le système éducatif de leur pays comme par exemple :

- d'établir des systèmes de formation et de qualification des enseignants ;
- de décider des moyens à mettre à disposition et de les distribuer ;
- de spécifier les finalités éducatives comme partie intégrante d'un contexte social plus vaste et d'objectifs précis en terme de résultats ;
- de décider de l'organisation de la scolarisation de l'enfance à l'adolescence et au-delà.

Les études et les données comparatives internationales sont largement appréciées par les décideurs comme une source d'information permettant l'évaluation continue de ces projets globaux.

Ce qui, au départ, était l'élaboration d'un espace de recherche comparative en éducation a coïncidé avec une reconnaissance croissante de la part des

politiciens, des responsables industriels et d'autres encore, que l'éducation était un des facteurs les plus importants de la réalisation de visions politiques, sociales ou économiques à long terme telles que :

- développer une société avec une meilleure répartition des ressources entre classes sociales, groupes ethniques, hommes et femmes ou tout autre groupe ;
- répondre aux besoins d'une main d'œuvre hautement qualifiée afin de réussir sur le marché international ;
- mettre en valeur et développer la démocratie en donnant à tous les citoyens une éducation de base et plus encore une éducation pour qu'ils soient capables de vivre pleinement leur vie et d'être des acteurs à part entière du processus démocratique.

Il s'agit là de quelques exemples de conceptions d'une société idéale qui étaient, et sont toujours, largement partagées. Dans le même temps, la période suivant la Seconde Guerre mondiale a vu la constitution des organisations internationales telles que les Nations unies, la Banque mondiale, l'Organisation pour la coopération et le développement économique (OCDE) et l'Union européenne qui gagnèrent rapidement en taille et en influence, organisations ayant des missions différentes voire, sous certains aspects, conflictuelles. Elles investissent néanmoins du temps et beaucoup d'efforts dans l'étude des systèmes et des finalités de l'éducation de leurs pays membres et plusieurs d'entre elles sont liées par des projets communs sur des questions d'éducation.

L'IEA est devenu une source de données et d'analyses pédagogiques non seulement pour les décideurs nationaux, mais aussi pour plusieurs organisations internationales. Outre l'UNESCO qui a pris part à la formation de l'IEA, l'OCDE (avant que PISA ne soit créé) utilisait les données fournies par les études de l'IEA dans ses publications *Education At A Glance* (par exemple l'utilisation des données TIMSS ; cf. OCDE, 1996, 1997 & 1998). Depuis ses premières études (au début des années 1960), l'IEA a été chargée d'un grand nombre d'études comparatives sur différents sujets, lesquelles se sont, au fil des années, étoffées jusqu'à inclure un grand nombre de pays. Dans le même temps, les défis méthodologiques sont devenus un moteur de développement pour de nouveaux projets et de nouvelles procédures psychométriques (Porter & Gamoran, 2002).

Ces deux dernières décennies ont vu la croissance de ces études comparatives renforcée par la réforme des services publics souvent appelée « le nouveau

management public ». Cette réforme se caractérise par une dérégulation du secteur public et une tendance à la privatisation de segments du secteur public qui peuvent être considérées comme l'infrastructure de la société. La dérégulation implique un transfert de responsabilités du pouvoir central vers les autorités locales. Néanmoins, les décisions importantes concernant l'École doivent être prises par les décideurs à un niveau administratif supérieur à celui des collectivités locales ou des écoles de quartier. Il demeure donc un besoin d'informations au niveau national.

Dans la plupart des pays qui ont vécu cette dérégulation, une des conséquences fut donc de renforcer le rôle du gouvernement central en développant un système national d'évaluation. On a observé un glissement de la régulation des « *inputs* » – par exemple la détermination de l'utilisation des ressources ou du nombre d'élèves par classe – vers un contrôle de l'« *output* » – résultats, enquêtes auprès des parents ou des élèves. De cette façon, les fournisseurs de services devinrent responsables [*accountable*] à la fois vis-à-vis du gouvernement central et des usagers de ces services. D'un côté, le gouvernement central pouvait contrôler et diriger les services en liant l'attribution des ressources supplémentaires à l'évaluation des résultats, ou bien intervenir et ajuster le système pour qu'il fonctionne selon les prévisions. D'un autre côté, les usagers pouvaient utiliser l'évaluation des résultats « *output* » dans leurs décisions personnelles concernant ces services publics.

Dans ce contexte, les études fournissent beaucoup d'indicateurs considérés comme pertinents surtout par les décideurs :

- elles permettent de mesurer une partie des *output*, concernant essentiellement les acquis des élèves ;
- elles produisent des indicateurs systémiques qui peuvent être directement liés à la politique choisie – le nombre moyen d'élèves par classe, la disponibilité des ressources (par exemple les ordinateurs), le type d'enseignement, le temps alloué aux différentes matières. Elles offrent aussi la possibilité de relier ces facteurs aux résultats des élèves ;
- elles fournissent des indicateurs de relations entre les variables que la politique cherche à faire évoluer dans une certaine direction, comme par exemple la finalité de l'école d'offrir à tous les mêmes chances d'apprendre, quelle que soit son origine sociale ;
- certaines études, PISA et TIMSS par exemple, produisent des indicateurs permettant de savoir comment les trois points précédents évoluent dans le temps en réitérant régulièrement les enquêtes.

Ce qui est le plus important c'est qu'elles fournissent un contexte pour l'interprétation d'un grand nombre de ces indicateurs. Un concept fondamental dans toute opération de mesure est la comparaison (Andrich, 1988). Si une évaluation est dépourvue de dimension internationale, il est difficile de savoir si les effets observés sont étendus ou limités. Même si les variations internationales ne peuvent servir à établir des inférences causales, elles offrent une description de ce qui est possible et définissent un contexte dans lequel les données nationales peuvent être comparées. Considérons le problème de l'équité. Il est souvent affirmé dans des publications politiques, que des différences systématiques importantes dans les résultats des élèves de niveaux socio-culturels différents, indiquent que les systèmes éducatifs ne donnent pas les mêmes chances à tous. Dès lors une dispersion importante des résultats pour l'ensemble de la population est souvent considérée comme un indicateur d'iniquité. Dans ces deux exemples, le contexte international permet aux décideurs d'évaluer si les différences entre les élèves ou groupes d'élèves sont importantes ou non, en les comparant aux différences observées dans d'autres systèmes éducatifs comparables.

En résumé, le second objectif de développement de politiques d'efficacité est, à bien des égards, compatible avec les buts auxquels avaient adhéré les chercheurs qui ont fondé l'IEA et qui ont entrepris les premières enquêtes (Objectif I). Les problèmes de politique éducative, tels que l'objectif d'amélioration des conditions permettant une éducation équitable et complète, sont aussi des problèmes fondamentaux au plan didactique et pédagogique. La différence est que, pour l'Objectif II, les études comparatives ne sont pas, dans l'ensemble, considérées comme fondamentales pour la recherche en éducation, ce qui ne veut pas dire qu'elles ne doivent plus être utilisées. Cependant, de telles questions sont secondaires par rapport à l'objectif premier qui est de suivre et comparer les résultats pédagogiques des systèmes éducatifs afin d'informer les décideurs. Cette question de la possibilité de recherches secondaires sera reprise de façon plus détaillée un peu plus loin.

LES DÉCIDEURS ET LES CHERCHEURS DANS L'ENSEIGNEMENT DES SCIENCES

La discussion a jusqu'à ici concerné deux protagonistes des enjeux de l'éducation : les chercheurs et les décideurs. On pourrait s'attendre à trouver un lien

étroit entre ces deux acteurs du débat sur l'éducation. Cependant, la recherche en éducation n'est, en général, guère considérée comme pertinente par les décideurs (Tomlinson & Tuijnman, 1994 ; Tooley & Darby, 1998). La même affirmation est valable concernant la question de l'enseignement des sciences (Jenkins, 2000 ; Millar, 2003), une question que nous utiliserons comme illustration dans ce qui suit. Il a même été dit qu'il était difficile de trouver des preuves d'un impact durable de la recherche concernant les pratiques d'enseignement des sciences qui déboucherait sur des changements et des améliorations dans l'enseignement et l'apprentissage des sciences à l'école (Lijnse, 2000). À l'évidence, il s'agit d'une généralisation grossière qui doit être nuancée par le fait que beaucoup de projets de recherche ont été conçus avec la mise en place d'un enseignement pour dessein. De tels projets ont, sans aucun doute, ouvert la voie à des innovations dans les *curriculum* locaux et dans la manière d'enseigner et d'apprendre les sciences dans beaucoup d'écoles. Il n'est cependant pas facile d'évaluer comment ce travail, caractéristique des recherches à petite échelle, s'est diffusé auprès d'un public plus vaste, et s'il n'a pas débouché sur un changement systémique à plus grande échelle. Nous n'avons aucune raison de penser que cette situation est propre à l'enseignement des sciences.

Ceci nous conduit à la première des trois observations-cadres concernant le lien entre la recherche et les décideurs : il semble que *la recherche en éducation est estimée, par beaucoup, comme n'étant guère pertinente pour l'élaboration des politiques éducatives, alors que la communauté des chercheurs estime quant à elle que ses efforts devraient être considérés comme une contribution primordiale au développement de politiques régionales, nationales et internationales.*

Les décideurs s'intéressent en revanche beaucoup aux études comparatives de grande ampleur. Les résultats et analyses de ces travaux ont été, à des degrés divers en fonction des pays, communiqués aux décideurs, aux enseignants et à un public plus large, principalement par le biais de rapports, présentations orales et discussions dans une grande variété de lieux. Le soutien réaffirmé aux études comparatives et la large utilisation des données dans des documents de politiques nationales et internationales sont au fondement de notre seconde observation-cadre : *les études comparatives sont considérées comme très pertinentes par les décideurs.*

Quand ces études comparatives sont débattues par les chercheurs en éducation, elles le sont de manière

critique (au sens négatif du terme). Mais notre propos n'est pas ici de reprocher aux autres chercheurs d'être critiques. Il nous paraît néanmoins étrange que de nombreux chercheurs ne tiennent pas à s'engager de manière plus positive en faveur des conclusions de ces études, qui, selon nous, peuvent être considérées comme pertinentes pour la recherche.

Notre troisième observation pourrait se résumer ainsi : tout compte fait, on peut conclure que *les études comparatives ne sont guère considérées comme pertinentes par la communauté des chercheurs en éducation, alors que les chercheurs impliqués dans celles-ci estiment que leurs travaux devraient être considérés comme une contribution importante aux connaissances partagées par la communauté.*

La première et la troisième observation peuvent sembler identiques : les deux ensembles, chercheurs spécialistes de l'enseignement des sciences et chercheurs en études comparatives à grande échelle, partagent le même défi communicationnel. Ils manquent de lieux pour diffuser les résultats de leurs recherches, pour discuter de leurs significations au-delà de leur propre communauté de recherche. Tous partagent la commune conviction que leurs travaux intéresseraient les autres.

E. W. Jenkins (2000) traite des différences fondamentales qui existent entre la recherche en éducation et l'élaboration de politiques en s'inspirant des travaux de C. C. Loving et W. W. Cobern (2000) ainsi que ceux de M. Huberman (1994). Il suggère que, non seulement les chercheurs en éducation et les décideurs ont des priorités différentes, mais qu'ils évoluent en outre dans des systèmes de connaissances différents : « les connaissances acquises dans un système pour une série définie d'objectifs ne sont normalement pas facilement transférables à un autre. » (Jenkins 2000, p. 18)

Jenkins ne donne cependant aucune définition du concept de « système de connaissances » et ne définit pas plus les aspects spécifiques des deux systèmes réputés très différents ne propose de surcroît aucune solution pour traiter le problème qu'il pose. Son affirmation coïncide néanmoins avec nos propres observations et formule bien l'idée, étayée sur l'expérience et le bon sens, que les décideurs et les chercheurs utilisent des critères différents pour juger de la pertinence des savoirs appropriés, principalement parce que les deux systèmes utilisent les savoirs dans des buts différents.

Les décideurs sont, bien plus que les chercheurs, confrontés au problème de la prise de décision. Cela

implique que deux propriétés des savoirs sont estimées pertinentes. Les décisions sont contraintes par le temps et la célérité de la prise de décision est généralement plus grande que pour la plupart des chercheurs. Il est donc probable que, devant rapidement définir, en un temps très court et sous la pression (avant les prochaines élections par exemple) des politiques éducatives, les décideurs considèrent que les connaissances pouvant être acquises et assimilées en un temps relativement court, sont les plus pertinentes. Confrontés aux réalités de la prise de décision, les décideurs apprécient particulièrement les savoirs considérés *comme* « vrais », comme s'il s'agissait de preuves présentées devant un tribunal.

L'étude PISA de l'OCDE est particulièrement intéressante en ce qui concerne la manière dont on offre aux chercheurs en éducation l'occasion de communiquer directement avec les décideurs. Par le biais de nombreuses initiatives, l'OCDE cherche à établir des procédures et des lieux permettant d'informer les décideurs des recherches en éducation. Réciproquement, par ces mêmes biais, les décideurs peuvent communiquer le besoin d'informations nécessaires à leurs prises de décisions. Il s'agit ici au moins d'une réponse partielle à la question de savoir comment il est possible d'avoir un bon échange, – dans les deux sens – d'informations entre les deux systèmes de connaissances. Cela implique que le but général de l'étude PISA s'aligne largement sur la manière qu'ont les décideurs de définir et justifier leurs objectifs en termes de résultats. Cela veut également dire que les mesures cognitives sont contextualisées avec des variables perçues comme politiquement importantes. Sont incluses les variables directement manipulables au plan politique (la taille des écoles par exemple), les liens que la décision politique veut renforcer ou au contraire affaiblir entre les résultats et les variables contextuelles (par exemple la relation entre résultats scolaires et statuts socio-économiques), et les variables employées pour contrôler ces informations.

Nous avons jusqu'à présent affirmé que les études comparatives peuvent fournir des outils de communication pour échanger des connaissances politiquement pertinentes entre les décideurs et les chercheurs en éducation. Il existe cependant un argument plus essentiel quant à la question de savoir pourquoi les chercheurs en éducation pourraient être grandement intéressés par ces études. Nous allons maintenant aborder cet argument : ces études leur fournissent des données uniques et précieuses.

AU-DELÀ DE L'OBJECTIF PREMIER DES ÉTUDES COMPARATIVES

Les études comparatives en éducation telles que TIMSS et PISA, ont un objectif principal que A. Schleicher (2000, p. 65) condense dans les trois résultats que PISA se propose de fournir : « un profil de connaissances et de capacités de base partagés par les élèves à la fin de leur scolarité obligatoire ; des indicateurs contextuels établissant un lien entre les résultats des élèves et les caractéristiques de leurs écoles ; des indicateurs de tendance montrant l'évolution des résultats dans le temps. »

Il importe de compléter ce programme principal par des travaux de recherches secondaires qui vont des contributions théoriques à une analyse secondaire des données ou des documents accompagnant ces données (par exemple les analyses des instruments et des items, les analyses des cadres théoriques et des logiques sous-tendant ces études). Différentes définitions du terme « analyses secondaires » sont proposées dans la littérature de recherche en sciences sociales. Elles se concentrent sur le fait que les analyses secondaires portent sur des données déjà existantes et que ces analyses sont menées par d'autres chercheurs que ceux qui ont initialement collecté les données et avec d'autres objectifs que ces derniers. La définition qui convient le mieux à la discussion proposée ci-après est probablement celle suggérée par A. Bryman (2004, p. 201), pour qui « l'analyse secondaire est l'analyse des données par des chercheurs qui n'ont probablement pas participé au recueil de ces données et dont le but n'a très probablement pas été envisagé par les responsables de ce recueil. »

Cette définition ouvre donc la possibilité pour la première équipe de chercheurs d'être impliquée dans l'analyse secondaire et cela d'autant plus que l'objectif de l'analyse secondaire peut avoir été inclus dans le projet initial de recherche. Ce dernier point est extrêmement pertinent dans beaucoup d'enquêtes officielles de grande ampleur concernant différents aspects de la vie sociale, beaucoup d'entre elles visant des objectifs multiples (Burton, 2000 ; The BMS, 1994), et considérant que le potentiel d'études secondaires par des chercheurs en sciences sociales est un élément important de leur projet initial.

Il existe un certain nombre de raisons parfaitement légitimes pour lesquelles tant de chercheurs veulent absolument créer leurs propres données plutôt qu'analyser des données déjà existantes. La princi-

pale raison est que « l'approche scientifique » peut, dans une certaine mesure, se définir de façon pragmatique par une méthodologie commençant par la formulation des questions de recherche et des hypothèses. Les données recueillies par d'autres ont été obtenues en ayant à l'esprit d'autres questions spécifiques ou d'autres hypothèses. Il peut donc être difficile d'utiliser ces données pour l'analyse de problèmes différents. Ensuite, l'utilisation de données recueillies par d'autres rencontre beaucoup d'obstacles techniques : les données peuvent ne pas avoir été rendues publiques, la documentation nécessaire à leur compréhension (par exemple le détail du code utilisé) peut manquer, ou bien encore leur lecture requiert des connaissances techniques que la plupart des chercheurs ne possèdent pas. Il peut enfin s'agir aussi de raisons idéologiques pour refuser d'étayer ses recherches sur des données recueillies par des organisations nationales ou internationales selon des perspectives d'analyse politique. Certaines de ces différentes raisons limitent ainsi le potentiel d'utilisation des données d'études comparatives dans les analyses secondaires.

Nous pensons toutefois que les bénéfices de ces analyses secondaires l'emportent largement sur leurs contraintes. Aussi allons-nous désormais brièvement présenter quelques-unes des raisons sous-jacentes pour lesquelles nous estimons que les données issues d'études internationales à grande échelle devraient être plus largement utilisées et par le plus grand nombre.

Tout d'abord, les données provenant de ces études possèdent des qualités, qui ne sont pas si fréquemment observées dans les recherches. Le principal motif de cette affirmation réside en ceci que la grande qualité de ces études est d'être très bien documentées. Dans le rapport technique des enquêtes PISA (Adams & Wu, 2002 ; OECD, 2005b), toutes les procédures de création d'instrument, d'échantillonnage, de marquage et d'interprétation des données sont décrites en détails. Il en va de même pour la documentation de TIMSS (par exemple Martin & Kelly, 1997a & 1997b ; Martin *et al.*, 2004). À la lecture de ces rapports, il apparaît évident que ces études se basent sur :

- des populations très clairement identifiées et des routines adéquates permettant l'échantillonnage de ces populations dans tous les pays ;
- un cadre de travail et des instruments très sophistiqués ; y compris une documentation sur la qualité de traduction dans les différentes langues ;

- des routines performantes et bien contrôlées qui permettent de s'assurer que la passation des tests se fait dans les mêmes conditions dans tous les pays ;
- des routines bien évoluées et un contrôle de qualité sur la façon dont les réponses des élèves sont notées et sur la manière de coder les données puis de les traiter.

Il est généralement difficile de récolter de telles données dans le cadre habituel de recherches bien moins coûteuses, ce qui nous amène à notre second argument en faveur d'une plus large utilisation de ces données. Des millions de dollars ou d'euros ont été dépensés pour produire ces bases de données de haute qualité. Des échantillons représentatifs ont été construits, des outils d'évaluation ont été distribués aux élèves puis renvoyés aux centres de recherches d'une manière qui garantisse qualité et comparabilité, et les données ont été rassemblées et restructurées grâce à un excellent travail d'experts, afin de renforcer la qualité des informations disponibles. Il demeure néanmoins que relativement peu d'argent a servi à l'analyse des données, l'essentiel ayant servi à recueillir les données. Il serait de toute évidence profitable de plus investir dans des analyses complémentaires de ces données.

Troisièmement, des données provenant d'études telles que PISA et TIMSS ont été mises à la disposition du public (bien que certains items de résultats soient protégés pour une utilisation future), et les chercheurs intéressés peuvent les obtenir par de multiples canaux (1). Afin de faciliter l'accès à sa base de données, PISA a créé un manuel expliquant comment analyser les données (OCDE, 2005a) ; mieux il est possible de dialoguer avec le centre national et, par ce biais, d'obtenir des conseils et d'accéder à des documents dont l'accès n'est pas toujours aisé.

Un grand nombre de données de grande ampleur recueillies constituent d'importantes sources activement utilisées pour définir les politiques éducatives dans la plupart des pays. Elle devraient être, attentivement et en tant que telles, examinées selon toutes les perspectives possibles. Même si on peut subodorer que ces données peuvent être idéologiquement biaisées ou utilisées, les analyses secondaires peuvent précisément aider à le comprendre et à le montrer (Pole & Lampard, 2002). Ce ne peut être qu'à partir de recensions ou d'analyses précises et documentées des données et des publications PISA ou TIMSS que d'autres chercheurs peuvent les

considérer et les analyser différemment et donc en proposer de nouvelles interprétations, fussent-elles critiques. Certaines de ces critiques concernent les aspects idéologiques des études (par exemple Atkin & Black, 1997 ; Brown, 1998 ; Goldstein, 2004a ; Keitel & Kilpatrick, 1999 ; Kellaghan & Greaney, 2001 ; Orpwood, 2000 ; Reddy, 2005) ; d'autres sont plus spécialement liées à des problèmes méthodologiques (Blum *et al.*, 2001 ; Bonnet, 2002 ; Freudenthal, 1975 ; Goldstein, 1995 & 2004b ; Harlow & Jones, 2004 ; Wang, 2001).

Ces questions méritent d'être discutées et les contributions produites jusqu'à présent sont des points de départ féconds pour le débat académique. Le développement et l'amélioration méthodologiques des études comparatives internationales de ces dix dernières années, peuvent, dans une certaine mesure, être considérés comme ayant été suscités par les critiques portées sur les méthodes utilisées (Porter & Gamoran, 2002). Il serait néanmoins souhaitable que beaucoup de ces critiques soient développées plus avant puisque certaines des affirmations avancées sont peu claires voire fondées sur un manque de compréhension de ces études, et/ou demeurent à l'état d'hypothèses non encore vérifiées.

Finalement, puisque les résultats de ces études sont principalement utilisés pour instruire des politiques à l'échelon national, il est nécessaire de discuter sur la manière dont les résultats peuvent être utilisés dans l'évaluation des systèmes éducatifs nationaux. Afin que les études comparatives puissent fournir une meilleure base d'information à cette discussion, il est sans doute nécessaire de développer un projet national précis, garantissant l'obtention d'informations considérées comme vitales dans le contexte national. L'Allemagne est un parfait exemple de pays qui a mis l'accent sur la dimension nationale tout en élaborant plusieurs prolongements intra-nationaux à l'étude PISA (2). Les élèves allemands ont répondu à des tests conçus à l'échelon international mais l'Allemagne a également élaboré un échantillon étendu afin de couvrir le système éducatif dans chacun de ses *Länder*, partiellement autonomes (Stanat *et al.*, 2002). Ces efforts ont accru, en Allemagne, l'implication des chercheurs concernant l'analyse des données, comme l'atteste le nombre d'articles dans les publications académiques (allemandes) en éducation. Cela a aussi suscité une prise de conscience du public et de nombreux débats sur les problèmes d'éducation en général. À un moindre degré, on trouve une situation identique en Norvège.

CIBLER LES QUESTIONS DE RECHERCHE EN ÉDUCATION

La discussion précédente a principalement présenté les raisons liées aux études elles-mêmes pour lesquelles les données, issues des études comparatives internationales sur les résultats scolaires, devraient faire l'objet d'analyses secondaires. Toutefois, le principal motif pour lequel les chercheurs en éducation pourraient avoir envie de consacrer leur temps et leurs moyens à ces analyses secondaires, réside dans le fait que les dites données permettent de travailler d'importantes questions de recherche. Nous allons donc à présent aborder la question, plus précise, de savoir comment ces données peuvent être utilisées pour cibler des questions de recherche en éducation. Au travers des exemples qui suivent, nous n'avons aucune intention de présenter une liste exhaustive des questions possibles que ces travaux (d'analyses secondaires) pourraient aborder. Notre but est de définir un cadre général de travail sur la manière dont les travaux de recherche secondaire rattachés à des études comme TIMSS et PISA peuvent être menés à bien. Notre formation explique pourquoi ce sont les études sur l'enseignement des sciences qui ont prévalu dans le choix des références qui sont proposées dans la discussion.

Nous avancerons ici l'idée que l'on peut classer la plupart des travaux de recherche secondaire utilisant ces données, en six types génériques d'objectifs de recherche ou d'approches méthodologiques. L'idée principale est de structurer notre tentative de présenter les différentes manières dont les chercheurs peuvent utiliser les données fournies par ces études. Il ne s'agit pas d'établir ici une liste exhaustive des problèmes que peut aborder la recherche secondaire sur ces données, pas plus que nous ne suggérons que ces types génériques forment une typologie de catégories mutuellement exclusives.

Utiliser les données, les résultats ou les interprétations comme contexte

L'analyse secondaire de données existantes, de leurs résultats ou de leurs interprétations peut être considérée comme le point de départ des projets de recherche et peut fournir le contexte ou le référent majeur suscitant des hypothèses ou des questions de recherche. Elle peut également fournir des données ou des conclusions à comparer avec d'autres données et d'autres conclusions. Dans ce type de recherches, l'objectif est généralement d'aller au-delà des données

et d'entreprendre une description et des analyses plus riches et approfondies des problèmes soulevés par les conclusions de ces études internationales.

Un exemple de ce type de travaux est, dans un contexte norvégien, le projet de recherche intitulé *PISA+* (3). Les chercheurs impliqués dans ce projet utilisent comme principale source de données, des transcriptions de cassettes vidéo faites en salle de classe et correspondant à plusieurs heures d'activités. Il ne s'agit évidemment pas d'une analyse secondaire des données de PISA mais, comme l'indique l'intitulé du projet de recherche, il a été suscité par certaines conclusions de PISA qui, dans le contexte norvégien, ont fait écho à celles d'une autre étude norvégienne basée sur des observations en classe (Klette *et al.*, 2003).

D'autres types de recherches centrées sur la manière dont les phénomènes évoluent avec le temps, ou sur la manière dont un groupe de personnes testées peut être comparé à un autre, peuvent aussi utiliser les données ou les conclusions d'études comparatives comme contexte. Dans certains cas, les études comparatives internationales peuvent fournir des données qui seront utilisées comme base et comme critère [*benchmark*] auxquels rapporter les données recueillies par d'autres chercheurs. Dans ce but, il serait nécessaire, à rigoureusement parler, de se servir d'une part de tests identiques et d'autre part, de procédures similaires de recueil et de traitement des données. L'utilisation d'items de TIMSS 1995 dans l'évaluation des résultats en sciences en Norvège avant et après la réforme des programmes scolaires de 1997 (Almendingen *et al.*, 2003) en est un exemple précis.

Analyses en profondeur de variables

Le propos qui suit concerne deux types d'analyses secondaires, liées au fait que les principales unités d'analyse des études comparatives consistent en un ensemble d'indicateurs caractérisant les deux principales composantes – personnes et items – des matrices de données. Les personnes sont choisies pour constituer un échantillon représentatif de la population de référence, et décrites à l'aide de mesures larges et composites, produits de l'agrégation de plusieurs items. Ces « construits » ou caractéristiques sont les mesures des résultats des élèves dans des domaines définis de façon large (sciences, mathématiques, lecture) ainsi que des descripteurs contextuels (statut socioéconomique, intérêt, motivation, stratégies d'apprentissages). Pour les analyses secondaires,

nous proposons donc deux classes d'objectifs respectivement liées à la déconstruction des deux axes de la matrice des données.

Parmi les analyses secondaires les plus fréquemment citées, se trouvent celles qui visent à présenter une image plus nette en étudiant en détail des caractéristiques définies plus étroitement, voire des items isolés. Ce type d'analyse utilise des informations se trouvant dans les données qui ne sont pas incluses dans les analyses de l'ensemble des résultats des tests (Olsen, 2005). Quelques exemples pertinents peuvent être cités. A. Turmo (2003b) fait état des aspects qualitatifs des réponses des élèves sur un petit nombre d'items cognitifs particuliers de l'étude PISA 2000 (qui portent sur le problème environnemental de la diminution de la couche d'ozone), en reliant les types de réponses aux travaux de recherche publiés sur l'enseignement des sciences. Il existe de nombreuses études similaires concernant les données de TIMSS 1995. J. A. Dossey *et al.* (2002) et M. Kjærnsli *et al.* (2002) ont analysé plusieurs items, respectivement en mathématiques et en sciences, à la lumière de ce que l'on connaît des conceptions erronées qui sont communément partagées par les élèves, en montrant combien l'importance relative de certaines conceptions spécifiques varie selon les pays. Un développement de ce type de recherche a été entrepris par C. Angell (2004), avec l'objectif de rendre compte des conceptions précises que les élèves exprimaient dans leurs réponses pour ensuite, et à un niveau général, mieux comprendre si les conceptions alternatives des élèves traduisent des structures quasi-conceptuelles ou si elles sont en grande partie dépendantes de la situation ou du contexte créés par les items, manifestant ainsi des idées plutôt intuitives.

Pareillement, les données des questionnaires soumis aux élèves peuvent être étudiées en profondeur, en sélectionnant une ou plusieurs variables pour une analyse plus étroitement ciblée et en y incluant alors les discussions et les interprétations alternatives faites à la lumière d'autres positions théoriques ou méthodologiques. En se fondant sur des données de PISA 2000, E. C. Papanastasiou *et al.* (2003) ont par exemple effectué une analyse en profondeur de la relation entre les usages des ordinateurs et les connaissances scientifiques aux États-Unis. S. Gorard & E. Smith (2004) ont utilisé d'autres données de la même étude pour calculer plusieurs indices de ségrégation au sein des pays de l'Union européenne. Ces indices viennent compléter les sélections d'indicateurs citées dans les publications officielles de l'OCDE sur les données de

PISA. Thématiquement parlant, ces deux derniers travaux comptent parmi les principales intentions de l'étude comparative en question et en tant que tels, ils illustrent bien l'idée que la ligne de démarcation entre « analyse secondaire » et « analyse principale » n'est pas facile à tracer.

Analyses en profondeur d'un sous-échantillon

Un autre type de recherche secondaire réside, nous l'avons brièvement indiqué plus haut, dans les analyses où l'axe « personne » de la matrice des données est déconstruit. Il s'agit d'une approche féconde pour cibler des problèmes précis dans la recherche en éducation. Beaucoup de séries de données provenant de ces études sont si volumineuses que les chercheurs peuvent en extraire des sous-séries avec des caractéristiques semblables.

On pourra, par exemple, mener des analyses approfondies relativement aux groupes minoritaires comme cela a été fait avec les données de TIMSS 1995 (Heesch *et al.*, 1998 & 2000) ou de PISA 2000 et 2003 (Hvistendahl & Roe, 2004 ; Roe & Hvistendahl, 2006). C'est du reste le principal objectif de ces études et l'OCDE a récemment publié (2006) un supplément thématique sur cette question. Une autre étude qui illustre les possibilités de cette approche a été réalisée par I. V. S. Mullis et S. E. Stemler (2002). Il s'agit d'une étude approfondie des différences garçons-filles parmi les bons élèves. Si de tels groupes aussi bien définis peuvent être étudiés, c'est assurément que les échantillons utilisés dans les études sources sont très importants. Il est donc possible de sélectionner les élèves au-delà du soixante-quinzième centile et de les diviser encore par genre tout en conservant des tailles d'échantillons adéquates, comme cela a été fait dans cette étude. Ainsi, en utilisant des données définissant le parcours scolaire, le milieu social, les attitudes et/ou les résultats des élèves, il est possible de construire un certain nombre de sous-groupes correspondant au problème sur lequel porte la recherche en question.

Une perspective de comparaison plus précise et plus fine des données

La quatrième classe d'analyses secondaires comprend les recherches qui visent une vision comparative plus fine en prenant comme unité d'analyse un petit nombre de pays, voire un seul pays. Ces études se rattachent à une longue tradition dans la recherche comparatiste en éducation.

En comparant une sélection de pays plus restreinte, deux stratégies de recherche assez différentes pour sélectionner les pays se sont révélées fécondes. Dans l'une d'elles, les pays sont sélectionnés pour représenter des systèmes éducatifs divergents puisque étudier une variation est le noyau de toute étude comparative. L'étude sur les cassettes vidéo de TIMSS (Stigler & Hiebert, 1999) compte comme l'une des plus heuristiques parmi les études internationales qui utilisent cette stratégie. Même s'il ne s'agit pas d'une étude secondaire des données de TIMSS, mais plutôt d'une étude indépendante, conduite et analysée en parallèle, elle n'en reste pas moins un parfait exemple montrant comment l'étude de systèmes éducatifs divergents peut dévoiler des présupposés et des éléments implicites dans les pratiques pédagogiques des pays participants. On peut citer à titre d'exemple d'analyse secondaire de PISA, la comparaison de l'enseignement des mathématiques en Finlande et en France, à laquelle une récente conférence a été récemment entièrement consacrée (4). Un autre exemple nous est donné par la comparaison des résultats en mathématiques de PISA au Brésil, au Japon et en Norvège (Güzel & Berberoglu, 2005).

L'autre stratégie consiste à comparer des systèmes éducatifs convergents. Ces études sont naturellement, et bien souvent des études régionales, concernant des pays voisins. Les rapports rédigés par des chercheurs de pays nordiques travaillant sur les données de PISA (Lie *et al.*, 2003 ; Mejdning & Roe, 2006) ainsi qu'un rapport rédigé par des chercheurs de plusieurs pays d'Europe de l'Est et basé sur des données de TIMSS 1995 (Vári, 1997) en constituent de bons exemples. Plusieurs raisons fondent l'intérêt de tels rapports. Tout d'abord, comparer des pays qui ont des traits culturels communs au sens large (historiques, politiques et/ou linguistiques) implique la possibilité de contrôler plus de facteurs, ce qui est assurément primordial dans toute comparaison. Deuxièmement, lors de comparaisons entre pays voisins ou de langues communes, les erreurs de mesure liées aux interactions « un item/un pays » se trouvent réduites (Wolfe, 1999). Par conséquent, d'un point de vue politique, de telles comparaisons ont plus de chances de produire des recommandations pertinentes pour la prise de décision politique puisque des pays voisins comme les pays nordiques ont toujours eu des échanges continus et réciproques quant à leurs politiques éducatives.

On peut encore réduire la base de comparaisons à des études de cas dans un seul pays : les rapports nationaux qui sont faits dans la plupart des pays participants

en sont de bons exemples. Cependant, ces rapports sont principalement des rapports publics à destination d'un très large public. Une large part de ses analyses figurant dans ce type de rapports devraient donc faire l'objet d'un retravail pour satisfaire aux exigences d'expertise par les pairs qui sont celles des publications internationales de recherche. Nous incluons également dans ce type d'analyses des études cherchant à relier les études internationales aux programmes et idéologies en vigueur dans différents pays.

De nombreux exemples de ce genre d'études existent, dont une récente contribution française (Bodin, 2005) présentée lors de la conférence sur la comparaison de l'enseignement des mathématiques en Finlande et en France. A. Bodin a ainsi analysé, pour la France, le degré de correspondance entre le programme de mathématiques, l'examen national de fin de 3^e (le Brevet) et les résultats obtenus en mathématiques avec PISA. Bien qu'il ait découvert que les épreuves de mathématiques de PISA ne couvrent officiellement qu'une petite partie du programme français (15 %), il met en relation ce constat avec celui selon lequel les épreuves du Brevet ne couvre également qu'une partie du programme (35 %). Mais, ce qui est sans doute plus frappant, c'est que si l'on ordonne les épreuves de PISA et du Brevet par niveaux d'exigences cognitives, celles de PISA couvrent un bien plus large éventail que celles du Brevet, lesquelles correspondent pour une majeure partie aux plus bas niveaux d'exigence.

Ces deux exemples mettent en lumière en quoi l'analyse d'un échantillon national de données issues des enquêtes internationale permet d'élargir et d'approfondir une perspective de recherche. O. Ichilov (2004) utilise des données de l'IEA *Civic Education Study* (CivEd) dans son rapport sur les orientations civiques dans les écoles hébraïques et arabes d'Israël, problème qui (hélas) est extrêmement important pour la communauté internationale. S. J. Howie (2004) et V. Reddy (2005) ont utilisé le cas de TIMSS en Afrique du Sud pour rendre compte et remettre en question l'intérêt de participer à des études comparatives internationales pour les pays en voie de développement, en particulier lorsque la langue maternelle des élèves n'est pas la langue des tests.

Combiner des données provenant de plusieurs études

Un autre objectif générique pour les analyses secondaires de données issues de ces études consiste à combiner des données de plusieurs d'entre elles.

Cependant, cela peut être une question techniquement redoutable au plan méthodologique, voire, dans certains cas, insoluble, puisque ces études différentes concernent des populations différentes et construisent des échantillons différents, ne permettant pas de trouver un dénominateur ou une unité d'analyse communs.

La possibilité la plus évidente pour mettre en relation des enquêtes internationales différentes est d'utiliser les résultats en les rassemblant par pays. L'étude exposée par B. Kirkcaldy *et al.* (2004) sur les relations entre l'efficacité du secteur médical, les acquisitions scolaires et le bien-être en est une preuve éclatante. Cette étude combine des données de PISA avec des données provenant de l'Organisation mondiale de la santé, des Nations unies, et d'autres sources encore.

Une autre possibilité pour combiner des données, et peut-être la principale compte tenu de la discussion précédente, serait de trouver un moyen de combiner les données de PISA et de TIMSS pour traiter de questions concernant l'enseignement des mathématiques ou des sciences, et les données de PISA et les compétences de PIRLS (*Progress in International Reading Literacy Study*) pour les compétences en lecture. Ce n'est pas une tâche évidente puisque ces études diffèrent sur de nombreux points, y compris l'âge des élèves. Il devrait cependant être en principe possible d'utiliser des données combinées par pays pour explorer et décrire les traits caractéristiques des résultats des élèves ainsi que leurs attitudes, leur motivation et leurs acquis dans différents pays.

Il est en outre fortement recommandé de recueillir des données complémentaires pour aider à établir des liens entre les différentes études. Ceci a déjà été fait, par exemple, dans une étude danoise : les mesures de niveau de lecture de PISA 2000 ont été méthodiquement rapportés à l'étude sur le niveau de lecture de l'IEA de 1991. Ceci a permis à la fois de comparer les deux mesures et, ce qui est le plus important, de rendre possible la création d'instruments de mesure des progrès en lecture pour les élèves danois de 1991 à 2000 (Allerup & Mejdning, 2003).

Une façon moins rigoureuse de relier des études entre elles serait de comparer les documents décrivant les études et de comparer les points communs entre les différents cadres théoriques et ensembles d'items. Par exemple, plusieurs comparaisons de TIMSS et PISA ont été faites en étudiant en quoi ces deux études différaient quant à leur conceptualisation des mathématiques et des sciences (Nohara, 2001 ; Olsen, 2005).

Aborder les données avec d'autres outils méthodologiques

Évoquons pour finir, non pas une approche méthodologique précise comme pour les catégories précédentes, mais plutôt une catégorie employée pour regrouper les nombreuses études usant d'approches méthodologiques différentes dans le traitement des données. Bien que ces études aboutissent souvent à des interprétations alternatives, leur objectif est également, et souvent, de discuter les conséquences des méthodes utilisées. Ces dernières années ont vu, par exemple, une prise en considération croissante des structures hiérarchiques dans l'analyse des données concernant les résultats scolaires : les résultats des élèves sont analysés dans le contexte des classes, lesquelles sont à leur tour étudiées dans le contexte des écoles, lesquelles le sont dans celui des régions, etc. (Malin, 2005 ; O'Dwyer, 2002 ; Ramírez, 2006 ; Schagen, 2004). Par l'utilisation d'outils statistiques sophistiqués, il est possible de prendre en considération l'emboîtement structurel de ces différents contextes tout en modélisant les données.

Pour ce qui est des études comparatives, certains observateurs ont remis en question l'exigence d'unidimensionnalité des mesures et ont analysé les séries de données de certaines de ces études afin de prendre en considération leurs multiples dimensions (Blum *et al.*, 2001 ; Gustafsson & Rosén, 2004). Un certain nombre d'études, conduites par des collègues ainsi que par nous-mêmes, ont choisi d'utiliser des méthodologies d'analyse permettant d'étudier les différents profils de réussite aux items, de différents groupes de pays partageant les mêmes caractéristiques (Angell *et al.*, 2006 ; Grønmo *et al.*, 2004 ; Lie & Roe, 2003 ; Olsen, 2006). D'autres ont préféré, pour traiter les données, des techniques de modélisation de variables ou de groupes latents (Hansen *et al.*, 2004 ; Papanastasiou & Papanastasiou, 2006 ; Wolff, 2004). Ces différentes études constituent seulement quelques exemples de l'utilisation d'approches méthodologiques alternatives. Leur fécondité est double : elles peuvent utiliser d'autres aspects des séries de données que ce qui a été initialement prévu et peuvent aussi être reconnues comme proposant des hypothèses alternatives quant à la manière de modéliser et d'interpréter ce genre de données.

CONCLUSION

Nous avons soutenu l'idée que les données fournies par les études comparatives internationales devraient

être appréciées en tant que source importante de données pour les chercheurs en éducation. L'argument principal en faveur de cette affirmation est que ces données sont complexes et portent sur de nombreux aspects et qu'il devrait donc être possible de pointer des problèmes fondamentaux de la recherche en éducation par le biais d'analyses secondaires de ces données. Nous avons donné l'exemple d'un certain nombre d'études spécifiques pour étayer cet argument. Ce dernier est renforcé par le fait que les données des enquêtes internationales sont de très haute qualité et que leur visée consistant à mettre en rapport mesures d'« *input* » et d'« *output* » est largement reconnue pour sa pertinence politique et qu'enfin, ces données ont généralement été mises à la disposition du public.

Jusqu'à présent, la plupart de ces recherches secondaires disponibles ont fait l'objet de publications sous formes d'ouvrages ou de présentations lors de conférences thématiques relatives à une ou plusieurs de ces études – et beaucoup de nos références aux analyses secondaires en sont précisément issues (Howie & Plomp, 2006 ; Mejding & Roe, 2006 ; Papanastasiou, 2004 ; Robitaille & Beaton, 2002 ; Shorrocks-Taylor & Jenkins, 2000). Un certain nombre d'articles ont également été publiés dans des revues académiques très diverses ; nous affirmons, pour notre part, que ces articles pourraient être beaucoup plus nombreux. Pour ne prendre qu'un exemple dans notre domaine (l'enseignement des sciences), on compte trois revues internationales prépondérantes : *Science Education*, *International Journal of Science Education* et *Journal of Research in Science Teaching*. Jusqu'à présent, le nombre d'articles de ces revues qui portent sur des analyses de données d'études comparatives est assez faible. Il en va de même pour les plus grandes conférences internationales dans ce domaine précis. Le nombre de communications, de posters, etc., concernant PISA et TIMSS est faible. Nos connaissances sont moins précises concernant les mathématiques ou la lecture mais nous subodorons que ce doit être également le cas. C'est, selon nous, un enjeu essentiel de savoir et de comprendre pourquoi les études comme TIMSS et PISA n'ont pas encore été plus largement reconnues comme des contributions majeures à la communauté de recherche en éducation.

Afin de tirer le meilleur parti de leur participation à ces études, les gouvernements devraient débloquent des moyens pour analyser plus avant ces bases de données et surtout conduire des analyses qui aideraient à mettre ces données en rapport avec les contextes nationaux. En Norvège par exemple, des

fonds ont été débloqués pour que les chercheurs impliqués dans les analyses principales puissent consacrer du temps au développement de recherches et à la publication de travaux, qui aillent au-delà des rapports institutionnels qui leur sont demandés. En outre, ces fonds ont été utilisés pour faciliter l'emploi des données provenant de TIMSS et PISA par les étudiants pour qu'ils puissent les utiliser comme support de travaux universitaires (masters ou doctorat). Dans notre pays, plusieurs thèses de doctorat ont ainsi été élaborées en s'étayant sur ce que l'on peut nommer des analyses secondaires de données de TIMSS et PISA (Angell, 1996 ; Isager, 1996 ; Kind, 1996 ; Olsen, 2005 ; Turmo, 2003a).

L'engagement des chercheurs à conduire des analyses secondaires de données issues d'études comparatives internationales recèle un double défi. Les chercheurs chargés de travailler sur ces études dans leur pays devraient être encouragés à publier beaucoup plus d'articles dans les revues académiques et à plus participer à des conférences, dans la mesure où ces études font l'objet de publications officielles sous la forme de rapports nationaux et internationaux. Et quoique ces rapports fassent l'objet d'une attention toute particulière de la part des personnes assurément compétentes, ils n'ont cependant pas suivi le processus d'expertise qui est l'une des caractéristiques les plus importantes de la production et de l'évaluation scientifiques.

De telles publications constituent selon nous des contributions vitales pour la recherche en éducation

et nous estimons, de surcroît, que l'occasion devrait être donnée aux chercheurs qui ne sont pas quotidiennement impliqués dans ces études, de s'impliquer davantage dans des projets d'analyses secondaires des données, y compris pour en développer une critique argumentée. Ces critiques pourraient ainsi être moins fondées sur la rhétorique ou l'anecdote mais, au contraire, étayées sur les analyses secondaires des données et/ou sur une présentation détaillée des arguments théoriques.

Nous souhaitons que l'argumentaire que nous avons développé incitera les chercheurs à explorer les possibilités qu'offre l'utilisation de ces données. Nous souhaitons aussi que le cadre systématique de types d'analyses que nous avons proposé, allant de l'utilisation des données comme simples référents pour obtenir de nouvelles données, à une modélisation plus sophistiquée, servira de guide pour savoir quelles formes peuvent prendre de telles analyses secondaires. Les nombreux exemples donnés devraient aussi être pris comme première source d'inspiration pour les chercheurs qui voudraient relever notre défi.

Rolf V. Olsen
r.v.olsen@ils.uio.no
Université d'Oslo (Norvège)
Faculté des sciences de l'éducation

Svein Lie
svein.lie@ils.uio.no
Université d'Oslo (Norvège),
Faculté des sciences de l'éducation

NOTES

- (1) Les données TIMSS sont disponibles sur internet à l'adresse : <http://timss.bc.edu/timss2003.html> et pour PISA à l'adresse <http://www.pisa.oecd.org> [liens consultés le 4 janvier 2007].
- (2) Cf. l'étude de M. Prenzel & K. Zimmer (2006) [ndlr].
- (3) On trouvera une description sur Internet à l'adresse suivante : <http://www.pfi.uio.no/forskning/forskningsprosjekter/pisa/> (consulté le 10 janvier 2007).

- (4) Voir les documents disponibles sur le site Internet de la Société mathématique de France : <http://smf.emath.fr/Vie/Societe/Rencontres/France-Finlande-2005/> (consulté le 4 janvier 2007).

BIBLIOGRAPHIE

ADAMS R. & WU M. [éd.] (2002). *PISA 2000 Technical Report*. Paris : OCDE.

ALEXANDER R. ; BROADFOOT P. & PHILLIPS D. [éd.] (1999). *Learning From Comparing : new directions in comparative educational research. Volume 1 : Contexts, Classrooms and Outcomes*. Oxford : Symposium Books.

ALEXANDER R. ; OSBORN M. & PHILLIPS D. [éd.] (2000). *Learning from comparing : new directions in comparative educational research. Volume 2 : Policy, Professionals and Developments*. Oxford : Symposium Books.

ALLERUP P. & MEJDING J. (2003). « Reading Achievement in 1991 and 2000 ». In S. Lie, P. Linnakylä & A. Roe (éd.),

- Northern Lights on PISA. Oslo : University of Oslo: Department of Teacher Education and School Development, p. 133-145.
- ALMENDINGEN S. B. M. F. ; TVEITA J. & KLEPAKER T. (2003). *Tenke det, ønske det, ville det med, men gjøre det... ? En evaluering av natur- og miljøfag etter Reform 97*. Nesna : Høgskolen i Nesna.
- ANDRICH D. (1988). *Rasch Models for Measurement*. London : Sage.
- ANGELL C. (1996). *Elevers fysikkforståelse. En studie basert på utvalgte fysikkoppgaver i TIMSS*. Oslo : Universitetet i Oslo : Det matematisk-naturvitenskapelige fakultet.
- ANGELL C. (2004). « Exploring Students' Intuitive Ideas Based on Physics Items in TIMSS 1995 ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 TIMSS*. Nicosia : Cyprus University Press, p. 108-123.
- ANGELL C. ; KJÆRNLSLI M. & LIE S. (2006). « Curricular and cultural effects in patterns of students' responses to TIMSS science items ». In S. J. Howie & T. Plomp (éd.), *Contexts of learning mathematics and science : Lessons learned from TIMSS*. London : Routledge, p. 277-290.
- ATKIN J. M. & BLACK P. (1997). « Policy Perils of International Comparisons : The TIMSS Case ». *Phi Delta Kappan*, vol. 79, n° 1, p. 22-28.
- BLUM A. ; GOLDSTEIN H. & GUERIN-PACE F. (2001). « International adult literacy survey (IALS) : an analysis of international comparisons of adult literacy ». *Assessment in Education*, vol. 8, n° 2, p. 225-246.
- BODIN A. (2005). *What does PISA really assess ? What it doesn't ? A French View*. Communication présentée à la conférence franco-finnoise « Enseigner les mathématiques : au-delà de l'enquête PISA », Paris : 6-8 octobre. Disponible sur Internet à l'adresse : <http://smf.emath.fr/VieSociete/Rencontres/France-Finlande-2005/BodinGB.pdf> (consulté le 8 janvier 2007).
- BONNET G. (2002). « Reflections in a Critical Eye [1] : on pitfalls of international assessment ». *Assessment in Education*, vol. 9, n° 3, p. 387-399.
- BOS K. T. (2002). *Benefits and Limitations of Large-Scale International Comparative Achievement Studies : The Case of IEA's TIMSS Study*. Unpublished PhD : University of Twente [Pays-Bas].
- BROWN M. (1998). « The Tyranny of the International Horse Race ». In R. Slee, G. Weiner & S. Tomlinson (éd.), *School Effectiveness for Whom ? Challenges to the School Effectiveness and School Improvement Movements*. London : Falmer Press, p. 33-47.
- BRYMAN A. (2004). *Social Research Methods*. Oxford : University Press.
- BURTON D. (2000). « Secondary Data Analysis ». In D. Burton (éd.), *Research Training for Social Scientists*. London : Sage, p. 347-360.
- DOSSEY J. A. ; JONES C. O. & MARTIN T. S. (2002). « Analyzing Student Responses in Mathematics Using Two-Digit Rubrics ». In D. F. Robitaille & A. E. Beaton (éd.), *Secondary Analysis of the TIMSS Data*. Dordrecht : Kluwer, p. 21-45.
- FREUDENTHAL H. (1975). « Pupils' achievements internationally compared - The IEA ». *Educational Studies in Mathematics*, vol. 6, p. 127-186.
- GOLDSTEIN H. (1995). *Interpreting international comparisons of student achievement*. Paris : UNESCO, vol. 63.
- GOLDSTEIN H. (2004a). « Education for All : the globalization of learning targets ». *Comparative Education*, vol. 40, n° 1, p. 7-14.
- GOLDSTEIN H. (2004b). « International comparative assessment : how far have we really come ? » *Assessment in Education*, vol. 11, n° 2, p. 227-234.
- GORARD S. & SMITH E. (2004). « An international comparison of equity in education systems ». *Comparative Education*, vol. 40, n° 1, p. 15-28.
- GRØNMO L. S. ; KJÆRNLSLI M. & LIE S. (2004). « Looking for cultural and geographical factors in patterns of response to TIMSS items ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 TIMSS*. Nicosia : Cyprus University Press, vol. 1, p. 99-112.
- GUSTAFSSON J.-E. & ROSEN M. (2004). « The IEA 10-Year Trend Study of Reading Literacy : A multivariate reanalysis ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004*. Nicosia : Cyprus University Press, vol. 1, p. 99-112.
- GÜZEL C. I. & BERBEROGLU G. (2005). « An analysis of the programme for international student assessment 2000 (PISA 2000) mathematical literacy data for brazilian, japanese and norwegian students ». *Studies in Educational Evaluation*, vol. 31, n° 4, p. 283-314.
- HANSEN K. Y. ; ROSEN M. & GUSTAFSSON J.-E. (2004). « Effects of Socio-Economic Status on Reading Achievement at Collective and Individual Levels in Sweden in 1991 and 2001 ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 PIRLS*. Nicosia : Cyprus University Press, vol. 3, p. 123-139.
- HARLOW A. & JONES A. (2004). « Why Students Answer TIMSS Science Test Items the Way They Do ». *Research in Science Education*, vol. 34, n° 2, p. 221-238.
- HEESCH E. J. ; STORAKER T. & LIE S. (1998). *Språklige minoriteters prestasjoner i matematikk og naturfag*. Oslo : Universitetet i Oslo : Institutt for lærerutdanning og skoleutvikling.
- HEESCH E. J. ; STORAKER T. & LIE S. (2000). *Språklige minoritetselever og realfag*. Oslo : Universitetet i Oslo : Institutt for lærerutdanning og skoleutvikling.
- HOWIE S. J. (2004). « TIMSS in South Africa : The Value of International Comparative Studies for a Developing Country ». In D. Shorrocks-Taylor & E. W. Jenkins (éd.), *Learning from Others*. Dordrecht : Kluwer, vol. 8.
- HOWIE S. J. & PLOMP T. [éd.]. (2006). *Contexts of learning mathematics and science : Lessons learned from TIMSS*. London : Routledge.
- HUBERMAN M. (1994). « The OERI/CERI Seminar on educational research and development : a synthesis and commentary ». In T. M. Tomlinson & A. C. Tuijnman (éd.), *Education research and reform : an international perspective*. Washington [D. C.] : OECD Centre for Educational Research and Innovation ; US Department of Education, p. 45-66.
- HUSEN T. (1973). « Foreword ». In L. C. Comber & J. P. Keeves (éd.), *Science Achievement in Nineteen Countries*. Stockholm/New York : Almqvist & Wiksell ; J. Wiley, p. 13-24.
- HUSEN T. & TUIJNMAN A. (1994). « Monitoring Standards in Education : Why and How it Came About ». In A. C. Tuijnman

- & T. N. Postlethwaite (éd.), *Monitoring the standards of education. Papers in honor of John P. Keeves*. Oxford : Pergamon, p. 1-21.
- HVISTENDAHL R. & ROE A. (2004). « The Literacy Achievement of Norwegian Minority Students ». *Scandinavian Journal of Educational Research*, vol. 48, n° 3, p. 307-324.
- ICHILOV O. (2004). « Becoming Citizens in Israel : A deeply divided society. Civic orientations in Hebrew and Arab schools ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 CivEd-Sites*. Nicosia : Cyprus University Press, vol. 4, p. 69-86.
- ISAGER O. A. (1996). *Den norske grunnskolenes biologi i et historisk og komparativt perspektiv*. Oslo : Universitetet i Oslo : Det matematisk-naturvitenskapelige fakultet.
- JENKINS E. W. (2000). « Research in Science Education : Time for a Health Check ? » *Studies in Science Education*, vol. 35, p. 1-25.
- KEEVES J. P. [éd.]. (1992). *The IEA Study of Science III : Changes in Science Education and Achievement : 1970 to 1984*. New York : Pergamon Press.
- KEITEL C. & KILPATRICK J. (1999). « The Rationality and Irrationality of International Comparative Studies ». In G. Kaiser, E. Luna & I. Huntley (éd.), *International Comparisons in Mathematics Education*. London : Falmer Press, p. 241-256.
- KELLAGHAN T. & GREANEY V. (2001). « The globalisation of Assessment in the 20th Century ». *Assessment in Education*, vol. 8, n° 1, p. 87-102.
- KIND P. M. (1996). *Exploring Performance Assessment in Science*. Oslo : Universitetet i Oslo : Det matematisk-naturvitenskapelige fakultet.
- KIRKCALDY B. ; FURNHAM A. & SIEFEN G. (2004). « The Relationship Between Health Efficiency, Educational Attainment, and Well-Being Among 30 Nations ». *European Psychologist*, vol. 9, n° 2, p. 107-119.
- KJÆRNSLI M. ; ANGELL C. & LIE S. (2002). « Exploring Population 2 Students' Ideas about Science ». In D. F. Robitaille & A. E. Beaton (éd.), *Secondary Analysis of the TIMSS Data*. Dordrecht : Kluwer, p. 127-144.
- KLETTE K. ; AUKRUST V. G. ; HAGTVET B. E. & HERTZBERG F. (2003). *Synteserapport : klasserommets praksisformer etter Reform97*. Oslo : Norges forskningsråd.
- LIE S. ; LINNAKYLA P. & ROE A. [éd.]. (2003). *Northern Lights on PISA : Unity and diversity in the Nordic countries in PISA 2000*. Oslo : University of Oslo : Department of Teacher Education and School Development.
- LIE S. & ROE A. (2003). « Unity and diversity of reading literacy profiles ». In S. Lie, P. Linnakylä & A. Roe (éd.), *Northern Lights on PISA*. Oslo : University of Oslo : Department of Teacher Education and School Development, p. 147-157.
- LJUNSE P. (2000). « Didactics of science : the forgotten dimension in science education research ? » In R. Millar, J. Leach & J. Osborne (éd.), *Improving science education : the contribution of research*. Buckingham : Open University Press.
- LOVING C. C. & COBERN W. W. (2000). « Invoking Thomas Kuhn : What citation analysis reveals about science education ». *Science & Education*, vol. 9, n° 1-2, p. 187-206.
- MALIN A. (2005). *School Differences and Inequities in Educational Outcomes*. Jyväskylä : Jyväskylä University Press.
- MARTIN M. O. & KELLY D. L. [éd.]. (1997a). *Technical Report Volume I : Design and Development*. Chestnut Hill : Boston College : Center for the Study of Testing, Evaluation and Educational Policy.
- MARTIN M. O. & KELLY D. L. [éd.]. (1997b). *Technical Report Volume II : Implementation and Analysis*. Chestnut Hill : Boston College : Center for the Study of Testing, Evaluation and Educational Policy.
- MARTIN M. O. ; MULLIS I. V. S. & CHROSTOWSKI S. J. [éd.]. (2004). *TIMSS 2003 Technical Report*. Chestnut Hill : Boston College : TIMSS & PIRLS International Study Center.
- MEJDING J. & ROE A. [éd.]. (2006). *Northern Lights on PISA 2003 : a reflection from the Nordic countries*. Copenhagen : Nordic Council of Ministers.
- MILLAR R. (2003). « Presidential Address : What can we reasonably expect of research in science education ». In D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos & M. Kallery (éd.), *Science Education Research in the Knowledge-Based Society*. Dordrecht : Kluwer, p. 3-8.
- MULLIS I. V. S. & STEMLER S. E. (2002). « Analyzing Gender Differences for High Achieving Students on TIMSS ». In D. F. Robitaille & A. E. Beaton (éd.), *Secondary Analysis of the TIMSS Data*. Dordrecht : Kluwer, p. 287-290.
- NOHARA D. (2001). *A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)*. « Working Paper » n° 2001-07, disponible sur Internet à l'adresse : <http://nces.ed.gov/pubs2001/200107.pdf> (consulté le 8 janvier 2007).
- O'DWYER L. M. (2002). « Extending the Application of Multi-level Modelling to Data from TIMSS ». In D. F. Robitaille & A. E. Beaton (éd.), *Secondary Analysis of the TIMSS Data*. Dordrecht : Kluwer, p. 359-373.
- OECD (1996). *Education at a Glance*. Paris : OECD Publications.
- OECD (1997). *Education at a Glance*. Paris : OECD Publications.
- OECD (1998). *Education at a Glance*. Paris : OECD Publications.
- OECD (2005a). *PISA 2003 Data Analysis Manual*. Paris : OECD Publishing.
- OECD (2005b). *PISA 2003 : Technical Report*. Paris : OECD Publications.
- OECD (2006). *Where immigrant students succeed. A comparative review of performance and engagement in PISA 2003*. Paris : OECD Publications.
- OLSEN R. V. (2005). *Achievement tests from an item perspective. An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students' knowledge and thinking in science*. Oslo : Unipub forlag.
- OLSEN R. V. (2006). « A Nordic Profile of Mathematics Achievement : Myth or Reality ? » In J. Mejdning & A. Roe (éd.), *Northern Lights on PISA 2003 : a reflection from the Nordic countries*. Copenhagen : Nordic Council of Ministers, p. 33-45.

- ORPWOOD G. (2000). « Diversity of Purpose in International Assessments : Issues arising from the TIMSS test of Mathematics and Science ». In D. Shorrocks-Taylor & E. W. Jenkins (éd.), *Learning from Others : International Comparisons in Education*. Dordrecht : Kluwer, p. 49-62.
- PAPANASTASIOU C. [éd.] (2004). *Proceedings of the IRC-2004*. Nicosia : Cyprus University Press, 4 vol.
- PAPANASTASIOU C. & PAPANASTASIOU E. C. (2006). « Modelling Mathematics Achievement in Cyprus ». In S. J. Howie & T. Plomp (éd.), *Contexts of Learning Mathematics and Science*. London : Routledge.
- PAPANASTASIOU E. C. ; ZEMBYLAS M. & VRASIDAS C. (2003). « Can Computer Use Hurt Science Achievement ? The USA Results from PISA ». *Journal of Science Education and Technology*, vol. 12, n° 3, p. 325-332.
- POLE C. & LAMPARD R. (2002). *Practical Social Investigation : Qualitative and Quantitative Methods in Social Research*. Essex : Pearson.
- PORTER A. C. & GAMORAN A. [éd.] (2002). *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington [D. C.] : National Academy Press.
- PRENZEL M. & ZIMMER K. (2006). « PISA 2003 en Allemagne : principaux résultats et utilité des tests nationaux complémentaires ». *Revue française de pédagogie*, n° 157, p. 55-70.
- RAMIREZ M. J. (2006). « Factor Related to Mathematics Achievement in Chile ». In S. J. Howie & T. Plomp (éd.), *Contexts of Learning Mathematics and Science*. London : Routledge, p. 97-111.
- REDDY V. (2005). « Cross-national achievement studies : learning from South Africa's participation in the Trends in International Mathematics and Science Study (TIMSS) ». *Compare*, vol. 35, n° 1, p. 63-77.
- ROBERTS D. A. (à paraître). « Scientific literacy/science literacy ». In *NARST Handbook of Research in Science Education*.
- ROBITAILLE D. F. & BEATON A. E. [éd.] (2002). *Secondary Analysis of the TIMSS Data*. Dordrecht : Kluwer.
- ROE A. & HVISTENDAHL R. (2006). « Nordic Minority Students' Literacy Achievement and Home Background ». In J. Mejding & A. Roe (éd.), *Northern Lights on PISA 2003 : a reflection from the Nordic countries*. Copenhagen : Nordic Council of Ministers.
- SCHAGEN I. (2004). « Multilevel Analysis of PIRLS Data for England ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 PIRLS*. Nicosia : Cyprus University Press, vol. 3, p. 82-102.
- SCHLEICHER A. (2000). « Monitoring Student Knowledge and Skills : The OECD Programme for International Student Assessment ». In D. Shorrocks-Taylor & E. W. Jenkins (éd.), *Learning from Others*. Dordrecht : Kluwer, p. 63-77.
- SHORROCKS-TAYLOR D. & JENKINS E. W. [éd.] (2000). *Learning From Others*. Dordrecht : Kluwer.
- PRENZEL M. *et al.* (2002). *PISA 2000 : Overview of the Study. Design, Method and Results*. Berlin : Max Planck Institute for Human Development.
- STIGLER J. W. & HIEBERT J. (1999). *The teaching gap : best ideas from the world's teachers for improving education in the classroom*. New York : Free Press.
- The BMS (1994). « Correspondence Analysis : A history and French Sociological Perspective ». In M. J. Greenacre & J. Blasius (éd.), *Correspondence Analysis in the Social Sciences*. London : Academic Press, p. 128-137.
- TOMLINSON T. M. & TUIJNMAN A. C. [éd.] (1994). *Education research and reform : an international perspective*. Washington : OECD Centre for Educational Research and Innovation ; US Department of Education.
- TOOLEY J. & DARBY D. (1998). *Educational research : a critique*. London : Office for Standards in Education.
- TURMO A. (2003a). *Naturfagdidaktikk og internasjonale studier. Store internasjonale studier som ramme for naturfagdidaktisk forskning : En drøfting med eksempler på hvordan data fra PISA 2000 kan belyse sider ved begrepet naturfaglig allmenndannelse*. Oslo : Unipub AS.
- TURMO A. (2003b). *Understanding a newsletter article on ozone : a cross-national comparison of the scientific literacy of 15-year-olds in a specific context*. Communication présentée à la 4^e conférence ESERA « Research and the Quality of Science Education », Noordwijkerhout [Pays-Bas] : Août 2003.
- VÄRI P. [éd.] (1997). *Are We Similar in Math and Science ? A Study of Grade 8 in Nine Central and Eastern European Countries*. Amsterdam : International Association for the Evaluation of Educational Achievement.
- WANG J. (2001). « TIMSS Primary and Middle School Data : Some Technical Concerns ». *Educational Researcher*, vol. 30, n° 6, p. 17-21.
- WOLFE R. G. (1999). « Measurement Obstacles to International Comparisons and the Need for Regional Design and Analysis in Mathematics Surveys ». In G. Kaiser, E. Luna & I. Huntley (éd.), *International Comparisons in Mathematics Education*. London : Falmer Press.
- WOLFF U. (2004). « Different Patterns of Reading Performance : A latent profile analysis ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 PIRLS*. Nicosia : Cyprus University Press, vol. 3, p. 188-202.