

ARTICLES**PISA : analyses secondaires, questions et débats théoriques et méthodologiques**

(dossier coordonné par Jean-Yves Rochex avec la collaboration d'Andrée Tiberghien)

- Jean-Yves Rochex* – Présentation p. 5
- Rolf V. Olsen, Svein Lie* – Les évaluations internationales et la recherche en éducation : principaux objectifs et perspectives p. 11
- Pierre Vrignaud* – La mesure de la littéracie dans PISA : la méthodologie est la réponse, mais quelle était la question ? p. 27
- Jean Moreau, Christian Nidegger, Anne Soussi* – Définition des compétences, choix méthodologiques et retombées sur la politique scolaire en Suisse p. 43
- Manfred Prenzel, Karin Zimmer* – Études complémentaires de PISA 2003 en Allemagne : principaux résultats et enseignements p. 55
- Martine Rémond* – Éclairages des évaluations internationales PIRLS et PISA sur les élèves français p. 71
- Élisabeth Bautier, Jacques Crinon, Patrick Rayou, Jean-Yves Rochex* – Performances en littéracie, modes de faire et univers mobilisés par les élèves : analyses secondaires de l'enquête PISA 2000 p. 85

Varia

- Virginie Albe* – Procédés discursifs et rôles sociaux d'élèves en groupes de discussion sur une controverse socio-scientifique p. 103
- Monique Loquet* – Analyse des gestes professionnels : illustration de « l'œil du maquignon » chez une formatrice en expression corporelle p. 119
- Marie-Paule Poggi, Nathalie Wallian, Mathilde Musard* – La construction sociale du jugement inspectoral : la définition du métier d'enseignant vue à travers une analyse de rapports d'inspection en EPS p. 131

NOTE DE SYNTHÈSE

- Philippe Sarrazin, Damien Tessier, David Trouilloud* – Climat motivationnel instauré par l'enseignant et implication des élèves en classe : l'état des recherches p. 147

NOTES CRITIQUES

- É. Bautier (dir.)* – Apprendre à l'école, apprendre l'école : des risques de construction d'inégalités dès la maternelle (E. Mangez) p. 179
- P. Caspard, J.-N. Luc, P. Savoie* – Lycées, lycéens, lycéennes : deux siècles d'histoire (F. Jacquet-Francillon) p. 180
- L. Chalmel* – Oberlin : le pasteur des Lumières (L. Loeffel) p. 184
- M. Grandière* – La formation des maîtres en France : 1792-1914 (A.-M. Chartier) p. 186

Sommaire

<i>R. Malet, E. Brisard (dir.) – Modernisation de l'école et contextes culturels : des politiques aux pratiques en France et en Grande-Bretagne (R. Normand)</i>	p. 189
<i>J.-A. Rondal – Expliquer l'acquisition du langage : caveats et perspectives (F. François)</i>	p. 191
LA REVUE A REÇU	p. 195
SUMMARIES	p. 199
RESUMENES	p. 203
ZUSAMMENFASSUNGEN	p. 207

Introduction

Jean-Yves Rochex

« L'un des principaux changements qui sont intervenus dans les systèmes éducatifs au cours des vingt dernières années dans le monde concerne l'analyse comparative de la qualité de ces systèmes sur la base d'une évaluation de leurs performances ». Ainsi François Orivel inaugure-t-il sa préface à l'ouvrage intitulé *Le pilotage des résultats des élèves* (1) que Neville Postlethwaite a publié en 2005 dans la collection « Principes de la planification de l'éducation » éditée par l'UNESCO. Cette affirmation, de même que l'ouvrage et la collection, rendent bien compte de la montée en puissance, durant les dernières décennies, et plus encore depuis le début de ce siècle, des grandes enquêtes internationales visant à évaluer et à comparer les « performances » des différents systèmes éducatifs et, en leur sein, de leurs différents segments, écoles ou établissements, à l'aide d'épreuves mesurant les acquis des élèves. Initiées au début des années 1960 par l'IEA – *International Association for the Evaluation of Educational Achievement* dont Neville Postlethwaite fut, avec Torsten Husén, un des principaux fondateurs, avant d'en assurer la Présidence de 1978 à 1986 –, poursuivies par cette association internationale (enquêtes TIMSS, 1995 et 1999, et PIRLS, 2001), ces enquêtes sont aujourd'hui de plus en plus commanditées et réalisées dans le cadre d'organismes internationaux tels que l'OCDE ou l'UNESCO, et concernent un nombre de plus en plus grand de pays et d'élèves. Tel est le cas de l'enquête PISA (Programme international

pour le suivi des acquis des élèves), mise en œuvre par l'OCDE en 2000, réitérée tous les trois ans, et visant à évaluer les acquis des élèves de quinze ans des différents systèmes éducatifs, principalement dans les pays de l'OCDE dans les domaines de la littéracie, de la culture mathématique et de la culture scientifique. La première version de l'enquête PISA a concerné trente-deux pays en 2000, et onze pays supplémentaires deux ans plus tard. La seconde a été initialement réalisée dans quarante et un pays, dont les trente appartenant à l'OCDE. Trente pays non membres de l'OCDE sont aujourd'hui concernés, à un titre ou à un autre, par l'enquête PISA. Quant à l'enquête SACMEQ, mise en œuvre par le consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation, et visant à évaluer, lors de leur sixième année de scolarité, les acquis des élèves en lecture et en mathématiques, elle a concerné sept pays lors de sa première édition en 1995, et quatorze lors de la seconde édition en 2000. Si certains des pays concernés par ces enquêtes internationales disposaient – et disposent toujours – d'une expérience en matière d'évaluation des acquis des élèves et de données produites par des services ou des organismes nationaux, données auxquelles peuvent être confrontés les résultats des enquêtes internationales, tel n'est pas, tant s'en faut, le cas de tous les pays, et pour certains d'entre eux les données produites par les enquêtes internationales sont les seules qui leur permettent d'avoir une vision d'ensemble des

« performances » de leurs systèmes éducatifs, ce qui n'est pas sans poser problème quand on sait que ces données ont été produites à l'aide d'épreuves et selon des méthodologies d'enquête sur lesquelles les pays ne disposant pas ou guère d'expérience, ni de compétences en la matière, n'ont pu avoir que très peu d'influence.

Ces enquêtes et les données qu'elles permettent de produire ne sont pas considérées, par les organismes intergouvernementaux et gouvernementaux, par les décideurs et les experts qui les promeuvent et les réalisent, comme de seuls outils de connaissance, mais aussi, si ce n'est d'abord, comme des moyens et des occasions d'évaluer et d'infléchir les politiques éducatives et le fonctionnement des systèmes éducatifs des pays concernés. Le présupposé est qu'évaluer et comparer les « performances » des différents systèmes éducatifs (et, au sein de ceux-ci, de leurs différentes composantes et unités), au regard d'épreuves et d'indicateurs communs, serait le moyen le plus pertinent d'améliorer leur efficacité et, par là, d'améliorer la compétitivité économique des pays concernés. Les technologies de l'évaluation et de l'*accountability* sont dès lors considérées et mises en œuvre comme instruments au service de la gouvernance et de l'amélioration des systèmes éducatifs, à l'échelle nationale et internationale, de même qu'au sein des différents pays, les résultats produits par les travaux s'inscrivant dans le courant de la *School Effectiveness Research* et visant à évaluer l'efficacité des établissements sont pensés et utilisés comme devant permettre d'améliorer cette efficacité. Ces travaux, leurs fondements théoriques ou idéologiques, leurs méthodologies, leurs résultats, leurs effets positifs ou pervers sur les systèmes éducatifs ou les établissements, ont été et demeurent au centre de nombreuses publications et de nombreux débats, dont le dossier que nous publions, consacré à PISA ne prétend bien sûr pas rendre compte de manière exhaustive, et sur lesquels la *Revue française de pédagogie* se propose de revenir (2).

On pourra néanmoins lire dans ce dossier six articles abordant et rendant compte sous différents angles : de questions et débats théoriques et méthodologiques portant sur l'enquête PISA elle-même ; des effets que la publication et les commentaires des résultats de PISA ont pu avoir sur les politiques scolaires de différents pays ; ou encore des possibilités d'analyses secondaires auxquelles ces données peuvent donner ou ont déjà donné lieu, que ces analyses concernent la confrontation des résultats de PISA à ceux produits par d'autres enquêtes internationales, par des enquê-

tes et des tests complémentaires menées à l'échelle nationale, ou par d'autres modes d'investigation visant à saisir, au-delà des scores obtenus, les modes de faire des élèves avec les épreuves qui leur étaient soumises (3). Trois de ces articles concernent plus spécifiquement la littéracie et son évaluation, auxquelles la première édition de PISA, en 2000, accordait un poids prépondérant, qui sont de longue date l'objet de nombreuses autres enquêtes nationales ou internationales, et qui dès lors ont pu être au cœur de différentes analyses secondaires et d'importants débats théoriques et méthodologiques.

Ainsi Rolf V. Olsen et Svein Lie s'efforcent-ils de présenter un panorama structuré des perspectives de recherches, primaires et secondaires, ouvertes par les enquêtes internationales telles que PISA ou TIMSS. Plaidant pour que les données produites par ce type d'enquêtes soient mieux prises en considération par les chercheurs, et pour un meilleur dialogue entre chercheurs et décideurs, ainsi qu'entre objectif de recherche et objectif d'amélioration de l'efficacité des systèmes éducatifs, ils dessinent un cadre général permettant de classer les perspectives et méthodologies d'analyses secondaires en six types d'objectifs et d'approches de recherche, en donnant de nombreux exemples concernant l'enseignement des sciences, lequel est leur principal domaine de recherche, tout en souhaitant que ces différents types d'analyses secondaires soient mieux soutenus et financés par les gouvernements nationaux.

Pour autant, comme le reconnaissent ces auteurs, les données des enquêtes internationales sont non seulement d'une grande richesse mais d'une extrême complexité, tant elles sont le produit de l'interaction entre la manière dont sont conceptualisées et définies les compétences que l'on vise à mesurer, et les méthodologies fort sophistiquées que l'on met en œuvre pour ce faire. L'article de Pierre Vrignaud s'efforce de faire le point sur ces interactions concernant la mesure de la littéracie dans PISA. Présentant les théories psychométriques et le modèle de mesure utilisés dans PISA, il insiste non seulement sur l'aspect très technique de ces choix et procédures – lesquels rendent difficile la discussion des modalités et des résultats de l'enquête par les non-spécialistes –, mais aussi sur les conséquences de cette approche sur l'analyse des résultats. Il interroge pour finir le postulat d'unidimensionnalité de la compétence évaluée et la conception même de la littéracie qui sous-tend l'enquête PISA, ses épreuves et consignes de correction et de traitement statistique, faisant ainsi écho à certaines critiques d'ordre méthodologique (Bonnet,

2002 ; Goldstein, 1995 & 2004 ; Rocher, 2003) et à certaines analyses qui se demandent si le souci et l'exigence psychométriques n'ont pas pris le dessus sur le souci et l'exigence d'élaboration théorique de ce qu'est la littéracie, voire s'ils ne masquent pas la faiblesse ou « l'absence d'une théorie élaborée organisant les différentes dimensions de la compréhension de l'écrit en tant qu'activité langagière et proposant un point de vue déterminé sur les opérations de lecture » (Bain, 2003 ; Goody, 2001 ; cf. également sur ce point Rochex, 2006).

Les deux articles suivants rendent compte des effets ou des retombées de l'enquête PISA, et des prolongements auxquels elle a donné lieu en Suisse et en Allemagne. La publication des résultats de PISA a en effet suscité dans chacun de ces deux pays de très vives réactions et de très importants débats politiques, médiatiques, pédagogiques et institutionnels (concernant l'Allemagne, cf. Gruber, 2006), liés sans doute au fait que ces résultats tendaient à montrer que les systèmes éducatifs de ces deux pays seraient à la fois moins efficaces et moins équitables que la moyenne des pays de l'OCDE. Ces constats et les débats auxquels ils ont donné lieu ont entraîné des investigations complémentaires et contribué à des modifications structurelles ou curriculaires dont rendent en partie compte ces deux articles. Le premier, que nous devons à Jean Moreau, Christian Nidegger et Anne Soussi, s'interroge sur l'adéquation du modèle et des variables statistiques utilisés par PISA pour rendre compte de spécificités locales, à partir du constat selon lequel certains items considérés comme étant de difficulté identique obtenaient néanmoins des pourcentages de réussite très différents dans des populations scolaires pourtant comparables, ou encore du constat selon lequel le lien statistique entre niveau de performance des élèves (ici en mathématiques) et indice socio-économique s'avère être très variable d'un canton à l'autre. Les auteurs présentent ensuite les grandes lignes d'un projet d'harmonisation structurelle et curriculaire à l'échelle fédérale, qui s'inspire largement de démarches mises en œuvre dans PISA, malgré des objectifs différents.

Manfred Prenzel et Karin Zimmer, quant à eux, présentent les investigations complémentaires auxquelles a donné lieu l'enquête PISA 2003 en Allemagne, concernant particulièrement l'enseignement des sciences et la résolution de problèmes : élargissement de l'échantillon permettant de situer les seize *Länder* et leurs spécificités, au regard des données internationales et nationales, mais aussi permettant de disposer de données sur des catégories de

population originaires de pays différents ou arrivées en Allemagne à des époques différentes ; tests et questionnaires complémentaires permettant d'élargir l'éventail et la nature des données recueillies ; répétition de certains de ces tests permettant de recueillir des données sur l'évolution des compétences des individus ou des classes au terme d'une année scolaire. L'ensemble de ces données ont rendu possible l'évaluation de la conformité des tests nationaux et internationaux aux programmes des différentes filières de scolarisation dans les différents *Länder*, ou l'étude des disparités entre *Länder* concernant leur niveau moyen dans les différents domaines évalués, les rapports entre les différents types de tests ou encore le lien, variable d'un *Land* à l'autre, entre origine sociale et performances des élèves.

Quant à Martine Rémond, elle met en rapport les résultats des élèves français à deux enquêtes internationales portant sur la *reading literacy*, l'enquête PISA 2000 portant sur des élèves de quinze ans, et l'enquête PIRLS, réalisée par l'IEA, et portant sur des élèves d'une dizaine d'années, résultats qui sont également rapprochés de ceux des dernières évaluations-bilans réalisées à l'échelle nationale par la Direction de l'évaluation et de la prospective du ministère de l'Éducation nationale. Ces mises en rapport et analyses croisées permettent de mettre au jour des profils de résultats et des caractéristiques propres à la population des élèves français, des écarts ou des comportements obérant leur réussite aux épreuves d'évaluation qui leur ont été soumises. La confrontation de ces données avec les réponses fournies par les maîtres ayant en charge les élèves évalués par PIRLS à un questionnaire les interrogeant sur leurs pratiques en matière d'exposition de leurs élèves à l'écrit, permet ensuite à l'auteur de s'interroger sur l'efficacité de telles pratiques déclarées, et de formuler quelques recommandations sur les contenus et les stratégies d'enseignement susceptibles de contribuer à l'amélioration des compétences des élèves en compréhension et en production d'écrit.

Enfin, Élisabeth Bautier, Jacques Crinon, Patrick Rayou et Jean-Yves Rochex rendent compte d'un travail d'analyse secondaire et complémentaire des épreuves de PISA 2000 portant sur les compétences en littéracie. Cette recherche a conjugué un travail d'analyse *a priori* de ces épreuves et de leurs modalités de correction et de cotation, un re-traitement statistique des réponses de plus de 800 élèves, mené à partir d'hypothèses théoriques différentes de celles qui ont présidé à la réalisation de l'enquête PISA, et un recueil et une analyse d'entretiens visant à mieux

comprendre les modes de faire de différentes catégories d'élèves face aux épreuves considérées, à mieux identifier les univers auxquels ces élèves se réfèrent pour comprendre les textes et répondre aux questions qui leur étaient proposés. L'ensemble des analyses ainsi produites ont conduit leurs auteurs à mettre au jour le caractère problématique et incertain des rapports entre résultats et modes de faire des élèves ; à mettre en évidence que le postulat de stabilité et d'homogénéité des compétences que les épreuves de PISA permettaient d'évaluer n'était vérifié que pour les élèves les plus performants ou les plus faibles, mais qu'il ne pouvait être soutenu concernant plus de 40 % des élèves ; et par conséquent, à interroger de manière critique tant les présupposés méthodologiques que les conceptions théoriques concernant la littéracie qui ont présidé à l'élaboration et à la mise en œuvre de l'enquête PISA.

Pour conclure, il convient de dire que les articles composant ce dossier ne représentent qu'un éventail partiel et fort incomplet des recherches et analyses auxquelles PISA et les enquêtes internationales ont donné et peuvent donner lieu, ou encore des travaux, débats et réflexions que suscitent ce type d'enquêtes ainsi que leurs (més-)usages sociaux, politiques, institutionnels ou médiatiques. Questions de recherche, débats scientifiques et enjeux sociaux et politiques qui appellent le développement de travaux, d'échanges et de controverses dont ce dossier n'est qu'un jalon, et auxquels la *Revue française de pédagogie* entend bien contribuer de nouveau dans l'avenir.

Jean-Yves Rochex
Université Paris 8
Équipe ESCOL

NOTES

- (1) Le titre de la version anglaise de ce même ouvrage est *Monitoring Educational Achievement*.
(2) Parmi les publications s'inscrivant dans ce débat, citons, sans souci d'exhaustivité, Ball, 1990 ; Goldstein, 2001 & 2002 ; Goldstein & Myers, 1997 ; Goldstein & Woodhouse, 2000 ; Normand, 2003

- & 2004 ; Reynolds & Teddlie, 2000 ; Teddlie & Reynolds, 2000 ; Thrupp, 2001.
(3) Pour un autre exemple d'analyse secondaire, cf. Duru-Bellat, Mons & Suchaut, 2004.

BIBLIOGRAPHIE

- BAIN D. (2003). « PISA et la lecture : un point de vue de didacticien ». *Revue suisse des sciences de l'éducation*, vol. 25, n° 1, p. 59-78.
- BALL S. J. (1990). « Management as moral technology : a Luddite analysis ». In S. J. Ball (éd.), *Foucault and Education : disciplines and knowledge*. London : Routledge, p. 153-166.
- BONNET G. (2002). « Reflexion in the critical eye : on the pitfalls of international assessment ». *Assessment in Education*, vol. 9, p. 387-399.
- BOTTANI N. & VRIGNAUD P. [éd.] (2005). *La France et les évaluations internationales*. Rapport établi à la demande du Haut conseil d'évaluation de l'école. Paris : Haut conseil de l'évaluation de l'école.
- DURU-BELLAT M. ; MONS N. & SUCHAUT B. (2004). « Caractéristiques des systèmes éducatifs et compétences des jeunes de 15 ans. L'éclairage des comparaisons entre pays ». *Les Cahiers de l'IREDU*, n° 66.
- GEWIRTZ S. (1998). « Can All Schools be Successful ? An exploration of the determinants of "school success" », *Oxford Review of Education*, vol. 24, n° 4, p. 439-457.
- GEWIRTZ S. ; BALL S. J. & BOWE R. (1995). *Markets, Choice and Equity in Education*. Buckingham : Open University Press.
- GOLDSTEIN H. (1995). « Interprétation des comparaisons internationales des résultats scolaires ». In UNESCO, *Études et documents d'éducation*. Paris : UNESCO, p. 11-32.
- GOLDSTEIN H. (2001). « Using pupil performance data for judging schools and teachers : scope and limitations ». *British Educational Research Journal*, vol. 27, n° 4, p. 433-442.
- GOLDSTEIN H. (2004). « International comparisons of students attainment : some issues arising from the PISA study ». *Assessment in Education*, vol. 11, n° 3, p. 319-330.
- GOLDSTEIN H. & MYERS K. (1997). « School effectiveness research : a bandwagon, a hijack or a journey towards enlightenment ? ». Paper presented at *British Educational Research Association* meeting, York, 11-14 september.
- GOLDSTEIN H. & WOODHOUSE G. (2000). « School Effectiveness Research and Educational Policy », *Oxford Review of Education*, vol. 26, n° 3-4, p. 353-363.
- GOODY J. (2001). « Competencies and Education : Contextual Diversity ». In D. S. Rychen & L. S. Salganik (éd.), *Defining and Selecting Key Competencies*. Göttingen : Hogrefe & Huber, p. 175-190.
- GRUBER K. H. (2006). « The German "PISA-shock" : some aspects of the extraordinary impact of the OECD's PISA study on the German education system ». In H. Ertl

- (ed.), *Cross-national Attraction in Education. Accounts from England and Germany*. Oxford: Symposium Books, p. 195-208.
- MASTERS G. N. (1995). « Élaboration d'échelles de mesure et agrégation dans les études de l'IEA : critique du rapport du Professeur Goldstein ». In UNESCO, *Études et documents d'éducation*. Paris : UNESCO, p. 33-39.
- NORMAND R. (2001). « Le mouvement de la *school effectiveness* et sa critique dans le monde anglo-saxon ». *Revue de l'institut de sociologie [de l'Université libre de Bruxelles]*, n° 1-4, p. 135-166.
- NORMAND R. (2004). « Les comparaisons internationales de résultats : problèmes épistémologiques et questions de justice ». *Éducation et sociétés : revue internationale de sociologie de l'éducation*, n° 12, p. 73-89.
- ORIVEL F. (2005). « Foreword ». In N. Postlethwaite, *Monitoring Educational Achievement*. Paris : UNESCO.
- POSTLETHWAITE N. (2005). *Monitoring Educational Achievement*. Paris : UNESCO.
- REYNOLDS D. & TEDDLIE C. (2000). « Reflections on the critics and beyond them ». Paper presented at *American Educational Research Association Annual Meeting*, New Orleans, 24-28 april.
- ROCHER T. (2003). « La méthodologie des évaluations internationales de compétences ». *Psychologie et psychométrie*, vol. 24, n° 2-3, p. 117-147.
- ROCHEX J.-Y. (2006). « Social, Methodological and Theoretical Issues Regarding Assessment Lessons From a Secondary Analysis of PISA 2000 Literacy Tests ». *Review of Research in Education*, vol. 30, p. 163-212.
- TEDDLIE C. & REYNOLDS D. (2000). « Responses to the criticisms of school effectiveness research contained in Slee, Weiner & Tomlinson (1998) and Thrupp (1999) ». Paper presented at *American Educational Research Association Annual Meeting*, New Orleans, 24-28 april (quoted in Thrupp, 2001).
- THRUPP M. (2001). « Recent School Effectiveness Counter-critiques : problems and possibilities », *British Educational Research Journal*, vol. 27, n° 4, p. 443-457.

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, concevons 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 débloquer 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'anecdotique 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. ; KJÆRNLI 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. ; KJÆRNLI 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.

La mesure de la littéracie dans PISA : la méthodologie est la réponse, mais quelle était la question ?

Pierre Vrignaud

L'objectif de cet article est de montrer l'interaction entre les aspects méthodologiques et la manière dont est conceptualisée et définie la littéracie dans l'enquête PISA. Pour introduire le thème on évoque d'abord les changements conceptuels dont a été l'objet l'évaluation des acquis des élèves dans les enquêtes internationales depuis les premières enquêtes de l'IEA jusqu'à l'enquête PISA. Après avoir rappelé que toute mesure est un construit, on expose les théories psychométriques qui fondent les modèles de mesure utilisés pour l'évaluation des acquis des élèves. La présentation des approches classiques permet d'introduire les concepts de difficulté et de discrimination des items ainsi que le concept de consistance globale d'une épreuve. On insiste sur les conditions de validité des modèles psychométriques qui posent des contraintes fortes sur la façon dont la variable mesurée est construite. On présente ensuite le modèle de mesure utilisé dans PISA qui appartient à la famille des modèles de réponse à l'item (MRI). PISA a privilégié le modèle dit de Rasch, on discute des conséquences de ce choix. On aborde ensuite un point très technique peu souvent abordé : celui de l'algorithme utilisé pour l'estimation des paramètres. La procédure utilisée aboutit à estimer non pas une valeur par sujet mais à construire la distribution des compétences de chaque sujet, on insiste sur les conséquences de cette approche sur la manière de conduire l'analyse des résultats. Ensuite, on discute du caractère réellement unidimensionnel de la variable du point de vue des contenus et du point de vue méthodologique. On s'interroge sur la contradiction apparente qu'il peut y avoir à considérer simultanément des sous-échelles et une échelle globale. En conclusion, on revient sur la manière d'interpréter la littéracie telle qu'elle est mesurée dans PISA en la mettant en relation avec d'autres enquêtes visant à évaluer le même type de compétence.

Descripteurs (TEE) : littéracie, comparaisons internationales, psychométrie, MRI.

Les enquêtes internationales sur l'évaluation des acquis des élèves et parmi elles, l'enquête PISA témoignent des changements conceptuels profonds qui ont modifié l'objet de ces enquêtes au cours de la dernière décennie. À la différence des enquêtes

internationales conduites pendant un demi-siècle par l'IEA (*International Association for the Evaluation of Educational Achievement* ; Association internationale pour l'évaluation du rendement scolaire), principalement centrées sur les acquis définis à partir des

curricula, PISA (OECD, 1999) a introduit l'idée qu'il est plus pertinent d'évaluer les compétences pour travailler et vivre dans une société post-industrielle, compétences considérées comme devant être le produit, l'*output* des systèmes éducatifs (Bottani & Vrignaud, 2005). Ce choix s'inscrit tout à fait dans la logique des enquêtes américaines : NAEP (*National Assessment of Education in Progress*), YALS (*Young Adult Literacy Assessment*), NALS (*National Adult Literacy Assessment*) – cf. Johnson, 1992, et de la première enquête internationale sur la littéracie, IALS (*International Adult Literacy Assessment*) – cf. Murray, Kirsch & Jenkins, 1998. Cette approche en termes de compétences plutôt que d'acquis peut apparaître comme l'œuf de Colomb des enquêtes internationales. Elle a permis d'éviter les difficultés que posait la construction d'un « méta-curriculum » – opération indispensable aux premières enquêtes mises en place par l'IEA à partir des années 1950 – en faisant glisser la mesure du contenu des enseignements vers une compétence suffisamment large pour considérer que tous les systèmes scolaires ont pour objectif de l'enseigner.

Dans cette optique, on argumente l'intérêt des enquêtes internationales en général et de l'enquête PISA en particulier par le fait qu'elles fournissent des informations sur des compétences très générales – transversales – qui interviennent dans la plupart des situations de la vie quotidienne et plus particulièrement de la vie professionnelle. Telles qu'elles sont présentées, ces compétences peuvent être considérées comme des interfaces entre la formation initiale dont elles sont le produit et le monde professionnel où elles sont mises en œuvre. Ces compétences sont porteuses de sens pour des utilisateurs divers, tant les chercheurs (économistes, psychologues, sociologues, sciences de l'éducation) que les décideurs des politiques éducatives ainsi que les médias. Si l'avantage de cette approche est de présenter une validité écologique importante de par son large spectre d'utilisation, son principal inconvénient est le danger de réification auquel cette compétence peut donner lieu. Dans le cadre de l'évaluation psychométrique qui est celui de ces enquêtes, les compétences sont avant tout des construits et ne sont pas séparables de la manière dont ces construits ont été opérationnalisés. Perdre de vue cette caractéristique laisse la porte ouverte à des généralisations abusives voire à des extrapolations que ne soutient pas réellement l'interprétation des résultats.

Ce danger est bien identifié et historiquement connu en psychologie à partir des problèmes liés à

la mesure de l'intelligence. On a d'abord critiqué les tests d'aptitudes intellectuelles pour leur caractère réducteur (mesure d'une intelligence plutôt académique) et socialement biaisé (en particulier les débats autour de la possibilité de construire des tests *culture free* ou *culture fair* ; sur ce point voir Vrignaud, 2002). Puis, les apports de la psychologie cognitive ont montré que l'approche unitaire de l'intelligence, véhiculée par l'emploi d'un score unique (par exemple le QI), ne reflétait pas ou, du moins, reflétait imparfaitement le fonctionnement psychologique des sujets (sur ce thème voir Huteau & Lautrey, 1999). La pluralité des processus et des stratégies mis en œuvre par les sujets pour résoudre les problèmes proposés n'était pas prise en compte par le score global. Les variabilités inter- comme intra-individuelles obéraient fortement la pertinence de l'interprétation d'un score unique.

Ces critiques et débats font parfois oublier que la construction de tests a permis le développement des méthodes et des concepts qui fondent la mesure en sciences humaines, en particulier, la psychométrie. La pierre de touche de l'évaluation en psychologie et en éducation est de distinguer entre une performance et une compétence. On observe une performance d'un sujet à une épreuve et on infère des conclusions sur sa compétence (Mislevy, 1994). Loin des représentations naïves que l'idée de calcul d'un score peut véhiculer, on cherche à estimer la compétence de ces sujets. Il s'agit d'un processus faisant appel à de nombreux concepts psychométriques et l'estimation de la compétence obtenue est aussi éloignée du simple calcul d'un score par sommation des bonnes réponses que peuvent l'être les premiers travaux sur les localisations cérébrales des recherches actuelles bénéficiant des avancées les plus récentes des techniques d'imagerie cérébrale.

Les enquêtes internationales ont été, depuis leur origine, un laboratoire d'essai et de développement des méthodologies psychométriques les plus sophistiquées (on trouvera une présentation très complète en français de la méthodologie des enquêtes internationales dans Rocher, 2003). Il faut dire qu'en plus du désir de l'ensemble des participants de fournir aux utilisateurs des résultats présentant les meilleures garanties de fiabilité, ces enquêtes présentaient des problèmes méthodologiques redoutables comme celui d'assurer l'équivalence de la mesure dans de multiples contextes linguistiques et nationaux. Les méthodes utilisées comme on peut le constater à la lecture du *technical manual* de PISA (Wu & Adams, 2002) sont complexes. Il paraît donc utile de donner

aux lecteurs intéressés un aperçu schématique de quelques uns des principaux points méthodologiques comme la construction de l'échelle de compétence et l'algorithme d'estimation des scores à ces échelles. Cet exposé est évidemment technique mais c'est justement l'un des problèmes cruciaux de ces enquêtes, que la compréhension des résultats et surtout de leurs limites est liée à des questions méthodologiques complexes.

De fait, l'enquête PISA est un dispositif de mesure de la littéracie et l'interprétation de ses résultats doit se faire en gardant présent à l'esprit la manière dont cette compétence a été construite. Il est donc utile pour expliciter ce qu'est et n'est pas la compétence évaluée dans PISA, de donner et de discuter les éléments méthodologiques qui valident ce passage entre la performance à un ensemble de tests et la compétence de populations de nombreux pays. L'objectif de cet article est de montrer les importantes avancées méthodologiques qui ont été intégrées dans PISA pour construire un dispositif de mesure solide et, en même temps, de montrer que l'accent mis sur le dispositif de mesure a peut-être laissé dans l'ombre d'autres interrogations sur la nature et la mesure de la compétence. Cet exposé nécessitera quelques rappels historiques qui montreront que les méthodes utilisées dans PISA sont le produit d'une longue histoire : celle de la psychométrie et des enquêtes sur l'évaluation des acquis des élèves en particulier aux États-Unis. On présentera d'abord les principaux concepts psychométriques selon l'approche classique. Puis, on présentera le modèle de mesure utilisé dans les enquêtes internationales en général et dans PISA en particulier. On insistera à la fois sur les avancées réalisées pour la mise au point de ce dispositif et sur les difficultés qui peuvent se rencontrer dans sa mise en œuvre et sa bonne compréhension par les utilisateurs. En conclusion, on reviendra sur les relations entre le dispositif de mesure et la nature de la compétence.

L'APPROCHE CLASSIQUE DE LA MESURE PSYCHOMÉTRIQUE

La majeure partie des méthodes utilisées pour les enquêtes internationales sur les acquis des élèves ont été élaborées au sein de la psychologie ou plutôt de la psychométrie. On parle aujourd'hui de « l'édu-métrie » pour définir un champ équivalent à celui de la psychométrie dans le domaine de l'évaluation en éducation. Cette distinction reste cependant une dis-

tinction de surface dans la mesure où les méthodes et les concepts sont largement similaires et où, bien souvent, les chercheurs qui travaillent et publient dans l'un de ces deux champs travaillent et publient également dans l'autre.

Les modèles de mesure

Pour introduire cette présentation des concepts de base de la psychométrie, on peut rappeler que la mesure, c'est à dire l'assignation de grandeurs à des objets en respectant certaines propriétés de ceux-ci, a posé en psychologie des problèmes particuliers qui ont abouti au développement de solutions originales au sein de cette discipline. Ces méthodes se sont trouvées rassemblées dans la psychométrie qui définit les méthodes à mettre en œuvre, depuis les dispositifs de collecte des données jusqu'à la définition de normes de fiabilité ; pour une présentation des théories et méthodes psychométriques, on se reportera, en français, à des ouvrages comme ceux de Dickes *et al.*, 1994 ou de Laveault & Grégoire, 2002. La démarche de validation de la mesure en psychométrie repose sur le principe selon lequel toute mesure est un construit. On parlera ici d'un modèle de mesure, et la démarche hypothético-déductive consiste à tester l'adéquation de ce modèle de mesure aux données. Plusieurs approches peuvent être mises en œuvre pour tester cette adéquation (on en trouvera une présentation dans les ouvrages cités plus haut). Les trois modèles de mesure les plus généralement utilisés sont l'approche classique (formalisée par Lord & Novick, 1969), les *modèles de réponse à l'item* (MRI) et les modèles structuraux. Les traitements des données des enquêtes internationales comme PISA utilisent majoritairement les MRI. Cependant, il est commode d'introduire les principaux concepts psychométriques à partir de l'approche classique.

La théorie classique des tests

On peut résumer le principe essentiel de la psychométrie par la formule de l'équation n° 1 :

$$\text{Score observé} = \text{Score vrai} + \text{Erreur de mesure.}$$

On cherche à distinguer performance (les résultats observés) et compétence (l'aptitude, le trait qui a produit cette performance et que l'on cherche à évaluer). L'étude de la fidélité interne est de s'assurer que le passage des items à la variable évaluée est fiable. Elle garantit que le score calculé à partir des items, généralement obtenu en faisant la somme des points accordés pour des réponses correctes a une

signification univoque. Ce qui ne serait pas le cas, par exemple, dans le cas où les items mesureraient des compétences différentes. C'est pourquoi on parle ici d'homogénéité ou de consistance interne. L'analyse interne se fait à deux niveaux : local, celui des items et global, celui du score. Au niveau des items, on s'intéresse principalement à deux de leurs caractéristiques : leur difficulté et leur discrimination.

Indice de difficulté de l'item

Dans le cas d'un score dichotomique (bonne ou mauvaise réponse), la difficulté de l'item est souvent estimée par la proportion d'élèves de l'échantillon qui donnent une réponse correcte à cet item. Le score moyen est une variante pour des items polytomiques (réponses multiples ordonnées). L'utilisation de cette proportion observée comme estimation de la difficulté peut être biaisée lorsque la représentativité de l'échantillon n'a pu être démontrée. À l'extrême dans le cas d'échantillons de compétence très faible ou très élevée, l'estimation de la difficulté des items peut conduire à des estimations très différentes. Cette dépendance entre l'estimation de la difficulté des items et l'estimation de la compétence des sujets a été la source de nombreuses réflexions visant à obtenir des estimations indépendantes. Les MRI ont été souvent présentés comme fournissant une solution à ce problème.

La gestion des non réponses ou plutôt des réponses manquantes est un autre problème pour l'estimation de la difficulté des items et de la compétence des sujets. On peut identifier au moins trois types différents de réponses manquantes : 1) les omissions intermédiaires ; 2) les omissions finales ; 3) les réponses manquantes structurelles. Les omissions intermédiaires ou finales correspondent à des items présents dans le protocole du sujet mais auxquels il n'a pas répondu. En général, on interprète les omissions intermédiaires, comme une déclaration d'ignorance et/ou une absence de prise de risque ; les omissions terminales comme un manque de temps. La distinction entre ces deux types de non réponse est importante car elle renvoie à la distinction entre test de puissance ou de vitesse (la rapidité du sujet à accomplir la tâche fait partie de la compétence évaluée). Le codage des réponses manquantes comme échecs ou comme items non examinés est donc fondamentale pour l'estimation de la difficulté des items. Le codage des omissions terminales comme des items non examinés ou comme des échecs aboutit à une estimation différente de la difficulté. Si la proportion de réussite est estimée à partir des seuls élèves de l'échantillon qui ont répondu à l'item, cela évite

d'interpréter comme absence de maîtrise du domaine ce qui dépend en fait de la vitesse de travail et du temps de passation. Les omissions structurelles proviennent, elles, de l'organisation du plan de collecte des données. L'utilisation de la méthode dite des « cahiers tournants » dans PISA produit des données structurelles manquantes. Pour concilier deux exigences : recueillir de l'information sur de nombreux exercices sans trop augmenter le temps de passation, on va répartir les exercices (items) en plusieurs blocs de longueur (temps de passation) à peu près égale. Chaque sujet ne passera qu'un nombre de blocs correspondant au temps de passation choisi. Pour permettre de traiter les données, il faut que toutes les paires de blocs soit présentes dans le dispositif expérimental. Il s'agit alors de réduire le nombre de combinaisons des paires de blocs pour maîtriser l'explosion combinatoire que pourrait engendrer la nécessité de construire toutes les combinaisons de paires de blocs. En général, on a choisi de construire des cahiers comprenant trois blocs pour s'appuyer sur une méthode de construction des plans expérimentaux bien connue : celle des triades. Pour neutraliser les effets liés à l'apprentissage et à la fatigabilité, on va contrôler l'ordre de passation des blocs en les contrebalançant. Chaque bloc apparaîtra au moins une fois dans les différentes positions de l'ordre de passation d'où le nom de « cahiers tournants » sous lequel ce dispositif expérimental est souvent désigné en français. Les protocoles contiennent donc des données manquantes structurelles et peuvent contenir des omissions finales et intermédiaires.

La recherche de solutions satisfaisantes pour la gestion de ces trois types réponses manquantes a été un des moteurs qui ont fait évoluer les méthodes employées pour traiter les données. Les réflexions ayant abouti à ces évolutions seront présentées dans le cadre des MRI.

Indice de discrimination de l'item

La discrimination de l'item renseigne sur la qualité et la quantité d'information apportées par l'item pour déterminer la compétence du sujet. Un item au pouvoir discriminant élevé apporte beaucoup d'information sur la compétence du sujet, un item peu discriminant renseigne peu sur la compétence du sujet. Leur pouvoir discriminant est un des principaux critères de sélection des items pour la construction définitive d'une épreuve. L'indice utilisé pour estimer le pouvoir discriminant de l'item se fonde sur la corrélation entre l'item et le critère évalué (en général le score au test). On fait l'hypothèse qu'un item est discriminant si les sujets qui le réussissent ont, en moyenne,

un score plus élevé que les sujets qui y échouent. La prise en compte de l'indice de discrimination est importante pour s'assurer de la fiabilité des items de l'épreuve (suppression des items peu discriminants donc peu informatifs). Il faut souligner que le modèle de mesure retenu pour le traitement des données de PISA requiert que tous les items présentent une discrimination égale.

La consistance au niveau global

De la même manière qu'on s'est intéressé à la validité des items, on va étudier la fiabilité de l'épreuve au niveau global. On parle d'homogénéité ou de consistance interne. Dans la théorie classique des tests, celle-ci est estimée par le coefficient α de Cronbach (Cronbach & Meehl, 1955). Cet indicateur répond à la question « l'ensemble des items est-il suffisamment homogène pour que le calcul d'un score soit valide ? » La valeur de l' α dépend à la fois de l'homogénéité des items (appréciée à partir de leurs intercorrélations) et de leur nombre. À homogénéité donnée, on peut augmenter la consistance interne du test en augmentant sa longueur (Cortina, 1993). Ce point est important dans la mesure où les épreuves pour les évaluations internationales sont en général plutôt longues.

Le modèle de mesure classique repose, comme les autres modèles, sur plusieurs conditions de validité. Les plus connues sont l'unidimensionnalité et l'indépendance conditionnelle des items et des sujets. Ces conditions seront davantage développées dans la présentation des MRI. On peut cependant signaler ici un problème posé par le format des épreuves de littéracie par rapport à la condition d'indépendance conditionnelle. L'indépendance conditionnelle se traduit par l'hypothèse selon laquelle la réponse d'un sujet à un item ne dépend pas de ses réponses aux autres items de l'épreuve. La réussite d'un sujet à un item ne dépend que de sa compétence sur le trait latent mesuré par l'item et de rien d'autre (en particulier pas de ses réponses aux items qu'il a examinés avant celui-ci). Il est souvent difficile de tester l'hypothèse d'indépendance conditionnelle.

On peut, par contre, identifier de nombreuses situations de *testing* où, par construction, la condition d'indépendance conditionnelle n'est pas respectée (Vrignaud, 2003). Ainsi, dans l'évaluation de la littéracie, on demande souvent de répondre à plusieurs questions posées sur le même texte. Cette manière de procéder se justifie par le fait que l'investissement du sujet, tant cognitif que temporel, pour s'approprier des objets complexes, ici un texte, doit être rentabi-

lisé au mieux. On utilise en anglais l'expression de *test/let* pour de tels exercices comprenant plusieurs items. En général, on ne tient pas compte des biais induits par cette dépendance dans le traitement des résultats des enquêtes internationales sur la littéracie (Dicke & Vrignaud, 1995). Ces biais ont pourtant des effets non négligeables comme l'ont montré les quelques recherches réalisées sur les *test/lets* (par exemple Wainer & Thissen, 1996). Les indicateurs psychométriques classiques tels que l' α de Cronbach sont biaisés dans le sens d'une surestimation.

La référence

On sait qu'un score brut à une épreuve n'est pas interprétable puisqu'il dépend de la difficulté des items profondément intriquée dans la compétence de l'échantillon. En psychologie, on a privilégié l'utilisation d'une population de référence pour situer les performances des sujets. La compétence du sujet va être estimée faible, moyenne ou forte selon que sa performance se situe, respectivement, en dessous de, égale ou supérieure à la moyenne de la distribution de la population de référence. Plusieurs solutions peuvent être adoptées pour situer un score dans une distribution de référence : le calcul d'une note standardisée en utilisant les paramètres (moyenne et écart type) de la distribution de référence – ce calcul s'accompagne souvent d'un changement d'échelle, l'exemple le plus connu étant celui du QI ; le recours à un étalonnage ; et enfin la référence à un critère de maîtrise. En éducation, on a plutôt privilégié le recours à un critère traduisant la maîtrise du domaine évalué par l'épreuve. L'approche la plus simple consiste à calculer le pourcentage des items réussis par le sujet et à considérer qu'au-delà d'un seuil donné (en général 75 ou 80 %) le sujet maîtrise le programme évalué par l'épreuve.

Cette façon de procéder peut inciter à des interprétations erronées. En effet, le fait que les scores à différents tests se trouvent ainsi standardisés laisse penser qu'ils sont comparables. Or, comme on l'a rappelé plus haut, la difficulté d'un item donc d'un test ne peut être appréciée qu'en relation avec la compétence de l'échantillon sur les résultats desquels cette difficulté a été estimée. Pour placer les résultats obtenus à différentes versions – ici linguistiques et/ou nationales – d'un même test il faut procéder à une opération dite « de parallélisation » afin de placer les résultats sur une même échelle (pour une présentation de ces procédures, voir Kolen & Brennan, 1995). Dans le cadre des enquêtes internationales, la procédure de parallélisation est gérée par le modèle de mesure employé (les MRI).

LES MODÈLES DE RÉPONSE À L'ITEM (MRI)

Présentation

Ces modèles regroupés sous l'appellation générique de *modèles de réponse à l'item* (MRI) – *Item Response Modeling* (IRM) en anglais (1) – ont été créés il y a une trentaine d'années (voir, pour une présentation, Hambleton & Swaminathan, 1985 ou, en français, Dickes *et al.*, 1994 ; Vignaud, 1996). Il faut signaler qu'ils ont été « inventés » à peu près simultanément et de manière indépendante au Danemark par le mathématicien Georg Rasch (1960) qui cherchait un modèle permettant de comparer des compétences d'élèves en lecture à plusieurs années d'intervalle et, aux États-Unis, par le statisticien Allan Birnbaum (1959, cité dans Birnbaum, 1968) qui cherchait à améliorer les modèles de mesure en psychométrie. Ces modèles ont profondément renouvelé l'approche psychométrique car d'une part ils offrent un cadre unitaire pour penser l'ensemble des concepts psychométriques (exposés plus haut à propos du modèle classique) et d'autre part, ils offrent un nouveau cadre d'interprétation des résultats aux tests en situant la performance des sujets par rapport à des tâches et non plus par rapport à la performance d'autres sujets. Ces modèles dont le principe est présenté dans l'équation n° 2 sont probabilistes. On postule que la probabilité qu'un sujet j donne une réponse correcte à un item (i) est fonction de la compétence (θ_j) du sujet et de la difficulté de l'item (d_i) :

$$\Pr(X=x) = f(d_i, \theta_j)$$

Dans le cas d'items dichotomiques, X prend les valeurs échec « 0 » ou réussite « 1 », on obtient donc la probabilité d'un échec ou d'un succès.

Les modèles MRI sont basés sur la recherche d'un modèle mathématique du fonctionnement de l'item permettant de représenter la relation entre difficulté de l'item et compétence du sujet. On utilise en général la fonction logistique. Le modèle le plus général comprend trois paramètres pour modéliser le fonctionnement de l'item : « b_i » la difficulté de l'item « a_i » la pente (discrimination de l'item), « c_i » le paramètre de réponse « au hasard » (2).

On peut les rapprocher des paramètres classiques : « b_i », la difficulté de l'item de la fréquence de réussite ; « a_i », la pente (discrimination de l'item) de la corrélation item/score ; « c_i » de l'étude des distracteurs. Le paramètre de compétence « θ_j » est une estimation de la mesure vraie de la compétence du sujet, c'est-à-dire que les MRI permettent de sépa-

rer performance et compétence. L'explication de la compétence et de la difficulté de l'item par une même variable latente justifie explicitement la comparaison entre items et entre sujets. Les paramètres de difficulté vont permettre de comparer les items entre eux. Les paramètres de compétences autorisent la comparaison des sujets et des groupes de sujets. Toutes les opérations de construction de tests et d'interprétation des résultats demandant d'assurer l'équivalence des items et des tests ou la comparaison de différentes populations vont se trouver ainsi facilitées.

Combien de paramètres utiliser pour modéliser la compétence ?

La question du nombre de paramètres du modèle a été souvent discutée. Les options retenues ayant des conséquences sur les conditions de validité des statistiques et la présentation des résultats, ces choix ont un retentissement sur le traitement des enquêtes internationales. Ainsi, pour les traitements de l'enquête PISA, ACER (*Australian Council for Educational Research*) (3) utilise un modèle dérivé du modèle de Rasch implanté dans son logiciel CONQUEST, modèle qui ne comprend, pour expliquer le fonctionnement de l'item, que le paramètre de difficulté alors qu'ETS (*Educational Testing Service*) s'appuie sur un modèle à deux paramètres (difficulté et discrimination) en utilisant des algorithmes d'estimation implantés dans le logiciel BILOG (Zimowski *et al.*, 1996) : voir pour un exemple les traitements de l'enquête IALS : Yamamoto, 1998. Cette différence de choix s'explique par au moins quatre raisons. En premier lieu des raisons historiques puisque les travaux sur les MRI s'étaient inscrits à ETS dans la suite des travaux de Birnbaum (1968) repris et enrichis par Lord (1980) qui avait introduit d'emblée un modèle à deux paramètres alors que les travaux d'ACER s'inscrivaient dans le cadre de l'approche de Rasch comme le montrent les logiciels construits par cette organisation (Titan puis Quest : Adams & Khoo, 1994). Ensuite des raisons liées au format des items, PISA comprend des items polytomiques (les réponses peuvent faire l'objet d'un codage ordonné selon des niveaux de réussite). Ce format d'item est facile à traiter par le modèle de Rasch (on sépare le paramètre de difficulté en une partie représentant la difficulté générale de l'item et une autre partie représentant le passage d'un niveau de difficulté à un autre) alors que l'estimation des paramètres de difficulté de tels items n'est pas aussi aisément accessible par le modèle à deux paramètres. Une troisième raison peut trouver son origine dans la détermination des niveaux de compétence

dont le rationnel sera présenté plus loin. La procédure de classement des items en niveau de difficulté est plus cohérente si la discrimination des items est identique. L'existence de différences de discrimination entre items peut rendre ce classement moins univoque. Enfin, une des phases essentielles de l'étude de l'équivalence en fonction des différentes versions linguistiques et/ou nationales est l'identification des *fonctionnements différentiels des items* en abrégé FDI (pour une présentation en français : Vrignaud, 2002 ou, dans le cadre des enquêtes internationales, Rocher, 2003). Le FDI est une différence de réussite à un item entre deux groupes de sujets comparables quant au construit mesuré par le test. Le FDI (4) peut porter sur chacune des caractéristiques de l'item : sa difficulté et sa discrimination. Le recours à un modèle à un seul paramètre simplifie l'approche de cette question. En revanche, l'utilisation du modèle de Rasch nécessite une condition de validité supplémentaire : l'hypothèse d'égale discrimination des items. Cette condition est en général vérifiée *a posteriori* dans la mesure où les tests d'adéquation au modèle de Rasch permettent de retenir l'hypothèse que ce modèle rend bien compte des données sans qu'il soit besoin d'introduire un paramètre supplémentaire afin de prendre compte la discrimination.

Dans le cadre des MRI, l'estimation des valeurs des paramètres de difficulté se fait sous cette hypothèse d'indépendance conditionnelle. Si on ne peut pas retenir l'hypothèse d'indépendance conditionnelle, alors il faudrait introduire un paramètre spécifique représentant la dépendance conditionnelle entre ces deux items comme la probabilité particulière de réussite à ces deux items, leur interaction comme le suggérait le statisticien anglais H. Goldstein (1980). Par exemple E. T. Bradlow, H. Wainer et H. L. Lang (1998) proposent un MRI incluant des paramètres représentant la dépendance locale et proposent un algorithme permettant l'estimation de ces paramètres.

ÉVALUER LA COMPÉTENCE DANS LE CADRE DES MRI

Les modèles MRI ont été présentés par leurs avocats comme renouvelant la théorie de la mesure. G. Rasch argumentait que l'estimation de la difficulté des items et de la compétence des sujets étaient indépendantes, ce qui fondait, selon lui, le concept d'objectivité spécifique (Rasch, 1977). Quels que soient les items passés par un sujet, on obtiendra

une même estimation de sa compétence. Quels que soient les groupes de sujets auxquels l'item a été administré, on obtiendra une même estimation de sa difficulté. Cette idée a été souvent considérée comme peu « réaliste » et semble d'ailleurs ne pas avoir donné lieu à de nombreuses études comme on le constate dans un ouvrage de synthèse sur les développements du modèle de Rasch (Fischer & Molenaar, 1995).

Les MRI définissent la compétence du sujet comme sa probabilité de résoudre des items d'une difficulté donnée. La compétence se définit donc par rapport à des tâches et non par rapport à d'autres sujets. Le paramètre de compétence du sujet définit sa zone de compétence qui peut être mise en relation avec les paramètres de difficulté des items. La définition de la zone de compétence nécessite de décider du seuil de probabilité de réussite retenu pour considérer que le sujet maîtrise l'item. Peut-on considérer qu'un seuil supérieur à 50 % est signe que l'item peut être résolu par le sujet ou vaut-il mieux considérer que seul un seuil proche de 100 % peut refléter la réelle maîtrise par le sujet ? Par exemple dans les évaluations éducatives aux États-Unis, le seuil de 80 % est généralement retenu (Kirsch, 1995). Ce seuil a l'avantage de garantir une probabilité quasi certaine de réussite, mais sa sévérité peut être trompeuse quant aux réussites réelles des sujets. En effet, les probabilités sont fortes que les sujets réussissent d'autres items de difficulté plus grande que celle comprise dans leur zone de compétence. Un second problème est celui de la définition de la compétence en fonction du contenu des items. Dire qu'un sujet est capable de résoudre des items d'une difficulté donnée renvoie à la définition opérationnelle de ces items. Cette définition peut paraître simple quand le contenu des items s'y prête : par exemple la complexité d'opérations arithmétiques, le nombre d'inférences à effectuer pour conduire un raisonnement. Néanmoins, ce type d'analyse apparaît souvent simplificatrice au regard des modèles de résolution proposés par la psychologie cognitive (Rémond, 2006).

La construction de l'échelle de compétence dans les enquêtes utilisant les MRI est essentiellement basée sur les regroupements d'items à partir de leurs indices de difficulté. Ainsi, dans la plupart des enquêtes internationales on définit plusieurs niveaux (en général cinq) de compétences. L'interprétation de chacun de ces niveaux est ensuite enrichie par l'analyse cognitive des items classés dans ce niveau. Ce système de définition d'une compétence est essentiellement psychométrique même s'il reçoit un habillage de psychologie cognitive. Un tel système a

été particulièrement développé par Kirsch et collaborateurs dans les enquêtes NAEP puis IALS et PISA (voir par exemple Kirsch, Jungeblut & Mosenthal, 1998). Cette approche présente deux inconvénients majeurs.

Le premier est d'être partiellement tautologique : cet item est facile puisqu'il est réussi par un grand nombre de sujets et qu'il correspond donc à des opérations de niveau faible. Un second inconvénient est la difficulté de déterminer le niveau auquel appartient un item. En effet, on prend en compte le paramètre de difficulté, non pas en lui-même, mais en recherchant quel niveau de compétence est nécessaire pour maîtriser un item de ce niveau de difficulté. Un item sera donc classé dans la catégorie correspondant au niveau de compétence permettant d'avoir une probabilité (en général 75 ou 80%) de le réussir. Mais, les sujets qui ont un niveau de compétence inférieur ont encore une probabilité élevée de le réussir si leurs compétences sont proches de la coupure séparant les classes de niveau. La qualité de cette séparation peut être appréciée à partir du pouvoir discriminant des items. L'information donnée par ces niveaux apparaît donc relativement floue et imprécise dans la mesure où les coupures sont par nature arbitraires : le fait d'être classé dans un niveau de compétence ne veut en aucun cas dire que le sujet n'est pas capable de fonctionner à des niveaux de compétence plus élevés. L'interprétation des niveaux n'est pas toujours facile car certains niveaux possèdent parfois peu d'items (en général les niveaux supérieurs). Et, surtout, l'interprétation en termes de fonctionnement cognitif n'est pas fondée sur l'analyse des tâches et des processus mais apparaît plutôt comme un produit dérivé du modèle de mesure psychométrique.

Dans PISA, les différents niveaux de compétence ont été définis de telle manière que les sujets dont le paramètre de compétence a une valeur proche de la borne inférieure ont une probabilité de 50% de réussir les items de ce niveau, et ceux dont le paramètre de compétence a une valeur proche de la borne supérieure, une probabilité de 80% de réussir ces mêmes items. Par construction, il est donc certain qu'un sujet ne réussit pas uniquement tous les items correspondant à son niveau et a – au moins pour les sujets proches de la borne supérieure – une probabilité non négligeable de réussir ceux du niveau supérieur. Encore une fois, il ne s'agit pas de pointer les insuffisances de la méthode sans en voir les avantages, en premier lieu, ceux de définir la compétence en relation avec des tâches et non plus en relation avec d'autres sujets comme dans l'approche psycho-

métrique classique. Il faut également souligner la prudence avec laquelle ces opérations ont été effectuées et la clarté avec laquelle elles sont exposées dans le *technical manual* (Turner, 2002). Mais, on ne peut passer sous silence le risque d'aboutir à une réification de la notion de niveaux de compétence qui, dans les représentations d'utilisateurs n'ayant pas eu accès à l'ensemble des sources techniques, peuvent paraître plus objectifs qu'ils ne le sont en réalité.

L'estimation des paramètres

La mise en œuvre de l'estimation des paramètres des MRI n'est pas une opération anodine – on trouvera une excellente présentation exhaustive de cette question dans l'ouvrage de F. B. Baker (1992). L'appréciation de l'adéquation des modèles MRI se pose aux différentes étapes de l'estimation des paramètres de difficulté des items et de compétence des sujets. En amont, les modèles MRI reposent sur des conditions de validité nombreuses : unidimensionnalité, indépendance conditionnelle des items, et, pour le modèle de Rasch, égal pouvoir discriminant des items. Ces conditions sont parfois difficiles à tenir et à vérifier. Ainsi R. K. Hambleton, H. Swaminathan & H. J. Rogers (1991) recensent une vingtaine de procédures à mettre en œuvre pour s'assurer de la possibilité d'application du modèle aux données. On peut citer également l'ensemble de travaux menés par l'équipe de Stout (Bolt & Stout, 1996 ; Shealy & Stout, 1993 ; Nandakumar, 1994) à l'Université de Chicago qui a permis de trouver des cadres conceptuels plus performants pour tester certaines hypothèses (unidimensionnalité, indépendance conditionnelle, fonctionnement différentiel des items). On peut regretter que les travaux de cette équipe soient totalement absents des traitements des enquêtes internationales.

L'algorithme d'estimation utilisé dans PISA est issu des travaux du statisticien américain D. Rubin sur l'algorithme dit « EM » (*Expectation-Maximization* ; Dempster, Laird & Rubin, 1977 ; Rubin, 1987 & 1991) (5). Rubin a clarifié le concept de valeur manquante en identifiant trois types de situations. La distribution des valeurs manquantes peut être représentée par une distribution complètement aléatoire (MCAR, *Missing Completely At Random*). Par exemple dans le cas des enquêtes internationales, l'utilisation de la méthode dite « des cahiers tournants », les réponses manquantes sont dites MCAR puisque les blocs qui n'ont pas été présentés à l'élève résultent d'une affectation au hasard d'un cahier à chaque élève. Le second type de situation est celui où on peut faire

l'hypothèse que la distribution des données manquantes peut être représentée par une distribution aléatoire (*Missing At Random*, MAR) mais peuvent dépendre des réponses des sujets à d'autres variables utilisées dans l'enquête. Enfin, le dernier cas dit *Missing Not At Random* ou *not ignorable* est celui où les données manquantes résultent d'un processus dépendant de la variable elle-même par exemple la non réponse à une question sur le niveau de revenus est plus fréquente dans les classes de revenu élevé.

Cette réflexion sur les données manquantes a conduit Rubin à opérer un renversement de perspective concernant l'estimation de la compétence des sujets. Rubin considère que la valeur manquante fondamentale est la position du sujet sur la variable latente. En effet, la compétence n'est connue que conditionnellement aux réponses du sujet à un nombre réduit de questions : celles qui sont incluses dans le test qu'il a passé y compris dans le cas où il a répondu à toutes les questions du test. Dans le cadre des MRI, cette formulation a conduit à repenser l'algorithme d'estimation des paramètres en utilisant l'algorithme EM (Bock & Aitkin, 1981), procédure implantée dans les logiciels BILOG dédiés à l'estimation des paramètres des MRI (Mislevy & Bock, 1990 ; Zimowski *et al.*, 1996). R. J. Mislevy et ses collaborateurs (Mislevy, 1987 ; Sheehan & Mislevy, 1990 ; Mislevy *et al.*, 1992) ont perfectionné cette approche en introduisant dans l'algorithme d'estimation les données descriptives du contexte du sujet (*background variables*) afin de rendre l'estimation du paramètre de compétence des sujets plus robuste.

Il s'agit d'estimer la compétence des sujets conditionnellement aux réponses qu'ils ont données aux items auxquels ils ont répondu (donc sans inclure les items manquant par construction des cahiers tournants et les omissions terminales) et conditionnellement aux variables décrivant le contexte socio-économique des sujets. Il faut préciser que le score de compétence du sujet est conceptuellement une valeur non observée et que son estimation renvoie non pas à un seul paramètre mais à une distribution. Conditionnellement aux réponses et aux caractéristiques de ce sujet, on infère, avec une plus ou moins bonne garantie, la distribution du paramètre de compétence d'un sujet ayant ces caractéristiques et ce patron de réponses aux items. On ne connaît pas la valeur vraie du paramètre de compétence mais sa distribution. Pour renforcer la robustesse de cette estimation, on va procéder à plusieurs tirages dans cette distribution de valeurs dites plausibles dont la moyenne sera une meilleure estimation de la compé-

tence de ce sujet. On trouvera le détail de cette procédure dans le *technical manual* (Adams, 2002).

On peut faire plusieurs commentaires par rapport à cette approche. En premier lieu, il est certain qu'elle prend au sérieux et qu'elle pousse, de manière particulièrement élégante, à l'extrême les concepts théoriques de la psychométrie. Sur le plan théorique, il est également certain que ces procédures permettent d'assurer une estimation plus rapide (convergence accélérée) et plus robuste des paramètres de compétence des sujets. On a pu également montrer qu'elle permet une estimation plus fidèle des moyennes des pays dans le cas des enquêtes internationales. Les points forts de cet algorithme sont la source de ses points faibles : la distribution des paramètres dépendant de plus nombreuses informations, cela introduit de nouvelles sources de biais dans l'estimation (par exemple les caractéristiques des sujets). Il va falloir s'assurer de la fidélité de toutes les informations portant sur les caractéristiques des sujets et de leur équivalence dans les différents contextes nationaux. Elle multiplie également les conditions de validité. Enfin, *last but not least*, cette procédure d'estimation aboutit à un ensemble (cinq dans PISA) de valeurs plausibles pour chaque sujet. D'après les publications sur cette approche, la théorie réalise un apport majeur à la réflexion psychométrique et les procédures semblent donner des résultats robustes pour l'estimation des paramètres des MRI. Il est, d'ailleurs, à noter que cette procédure élaborée par les chercheurs d'ETS pour les enquêtes américaines de type NALS et YALS (en ajoutant des procédures spécifiques au logiciel BILOG MG) puis pour les enquêtes internationales (voir par exemple IALS : Yamamoto, 1998) a été ensuite implantée dans le logiciel Conquest édité par ACER (Wu, Adams & Wilson, 1997) lorsque ce groupe a été chargé du traitement des données PISA. Le recours à la distribution de valeurs plausibles est maintenant généralisé dans les enquêtes internationales (voir par exemple PIRLS : Gonzalez, 2001).

Le fait d'estimer la compétence d'un sujet par cinq valeurs plausibles et non par un score unique a des implications importantes sur la manière de conduire les analyses. La dispersion de ces valeurs plausibles est aussi importante que leur moyenne. Toutes les analyses statistiques devraient donc être élaborées à partir des différentes valeurs plausibles et non d'une seule ou d'une agrégation de celles-ci. Par exemple, si on souhaite calculer la corrélation entre une variable de contexte (la PCS de l'élève) et la compétence, il faudra calculer cette corrélation pour chacune des cinq valeurs plausibles fournies pour chaque sujet

puis réaliser une agrégation des cinq valeurs obtenues pour la corrélation. La dispersion des valeurs de l'indicateur devra être utilisée pour les tests de signification. On trouvera des descriptions des procédures permettant de réaliser cette agrégation dans les publications traitant des méthodes d'imputations multiples (voir par exemple Schafer & Graham, 2002 pour une revue récente). Il n'est pas certain que les chercheurs réalisant des analyses secondaires à partir des données de PISA aient complètement intégré l'importance d'utiliser ces procédures pour obtenir des estimations sans biais des indicateurs dans le cadre de leurs analyses. Ces éléments sont présentés très explicitement et très clairement dans le *technical manual* (Adams, 2002).

L'unidimensionnalité de la littéracie : artefact ou réalité ?

Les MRI ont été l'objet de nombreuses critiques. La plus fondamentale porte sur leur réalisme pour représenter le fonctionnement des sujets répondant à des items. Ainsi, M. Reuchlin (1996) conteste le caractère continu du modèle qui présuppose qu'un sujet peut toujours réussir un item. La réponse à un item a un caractère discret. La réussite à un item difficile n'est pas peu probable pour un sujet peu compétent, elle est tout simplement impossible. Une contestation moins radicale porte sur certaines de leurs propriétés au premier rang desquelles l'unidimensionnalité.

L'unidimensionnalité de la variable latente laisse présupposer que les différences interindividuelles ne sont que des différences de puissance, que les différences de difficulté entre items ne sont que des différences quantitatives. On accrédite ainsi l'idée que quel que soit le niveau de compétence des sujets, ceux-ci mettent en œuvre des processus et des stratégies similaires pour répondre aux items. Cette critique a déjà été souvent portée à l'encontre des scores dont le caractère globalisant n'informe pas sur les processus sous-jacents (Huteau & Lautrey, 1999). Le nombre de variables à introduire dans un modèle pour rendre compte d'un ensemble de comportements est une question classique en psychologie.

La question centrale est la prise en compte de différentes dimensions et par conséquent de plusieurs compétences expliquant la performance des sujets aux items. Si l'on considère par exemple trois échelles, les relations entre leurs scores peuvent se situer entre deux situations extrêmes : 1) il n'existe aucune relation entre elles ; 2) la relation entre les dimensions est tellement élevée qu'il n'y a pas lieu de

les distinguer : elles mesurent la même chose. Dans le cas n° 1, les dimensions sont orthogonales (les corrélations sont nulles), il faut présenter et interpréter les résultats de chacune des échelles séparément. Dans le cas n° 2, les corrélations sont proches de 1, il n'y a pas lieu d'interpréter séparément les dimensions, les compétences mesurées sont complètement redondantes et si l'on devait les distinguer ce serait par un artefact sémantique qui consisterait à les nommer différemment. La plupart du temps, les données se situent entre ces deux pôles. La question est alors de décider à partir de quel seuil la liaison entre les dimensions peut être estimée comme suffisamment faible pour considérer que les dimensions mesurées correspondent à des compétences différentes ? Cette question a été au cœur de la plupart des débats autour des modèles psychologiques des aptitudes.

La dimensionnalité des compétences en littéracie s'inscrit dans un tel débat. On cherche à savoir si les résultats peuvent être présentés sur une ou plusieurs échelles. Cependant, la pertinence d'une discussion apparaît, dans le cas des enquêtes internationales sur la littéracie, comme faussée car pour des raisons de fiabilité de la mesure, on s'attache au fait que les épreuves soient fortement unidimensionnelles. On a montré *supra* que cette condition est requise par le modèle de mesure employé : le MRI. L'unidimensionnalité est à la fois la structure recherchée et la condition de validité (l'hypothèse au sens de l'*assumption*) des MRI. En effet, les modèles de base des MRI nécessitent la condition d'unidimensionnalité : on doit rendre compte des relations entre items (estimés par leurs paramètres) et entre les sujets ainsi qu'entre items et sujets par une seule variable latente.

La solution retenue pour l'interprétation de PISA est de considérer cinq échelles : trois de littéracie, une de mathématiques et une de science. On s'intéressera uniquement aux échelles de littéracie. Ces trois échelles se distinguent selon les auteurs du dispositif par les opérations auxquelles elles font appel (sur ce point voir Rémond, 2006) : 1) retrouver de l'information ; 2) développer une interprétation ; 3) réfléchir sur le contenu du texte. La distinction entre ces trois échelles et le rattachement des items à chacune d'elle a été fait à partir de jugements d'experts et des résultats de l'analyse des données. Les valeurs des corrélations entre échelles publiées pour les trois échelles de littéracie dans PISA 2000 sont très élevées ($> .89$; cf. Adams & Caspersen, 2002) et dans bien des cas seraient considérées comme suffisantes pour rassembler les trois échelles en une seule. Ce

qui est d'ailleurs le cas puisque certains résultats sont estimés sur une échelle globale qui est, elle-même, considérée par hypothèse comme unidimensionnelle puisqu'elle présente une bonne adéquation à un modèle de Rasch. On peut donc légitimement s'interroger sur le bien fondé de distinguer trois échelles puisqu'un modèle comprenant une seule échelle rend parfaitement compte des données (selon les décisions prises par les statisticiens quant à l'adéquation du modèle de mesure aux données).

CONCLUSION

Ce tour d'horizon du modèle de mesure et de l'estimation des paramètres dans les enquêtes internationales en général et dans PISA en particulier fait ressortir plusieurs points. D'abord la sophistication des méthodes utilisées, le soin apporté à résoudre des problèmes délicats posés par l'évaluation psychométrique. Bien que tous ces éléments soient présentés dans le *technical manual* (Adams & Wu, 2002), on peut s'interroger sur la réalité de leur accessibilité à l'ensemble des utilisateurs potentiels de PISA dans la mesure où la psychométrie, du moins à ce niveau de complexité, ne fait pas forcément partie du socle commun de connaissances de l'ensemble de la communauté scientifique francophone des sciences humaines. Ceci peut conduire certains utilisateurs à des erreurs dans l'utilisation des données comme cela a été souligné à propos de la prise en compte des valeurs plausibles dans les analyses secondaires.

Un second point est que, malgré le soin apporté à ces questions méthodologiques, certaines solutions restent encore insatisfaisantes au regard de la sophistication du reste de l'édifice. On a signalé parmi les aspects les plus techniques la violation de la condition d'indépendance conditionnelle. La question de la dimensionnalité apparaît plus centrale et donc plus gênante dans la mesure où elle est en prise directe avec la présentation et l'interprétation des résultats. Ceci conduit à une interrogation plus générale sur la nature conceptuelle de la compétence évaluée.

À ce sujet, il faut signaler que H. Goldstein et ses collaborateurs (Goldstein, 2004 ; Goldstein *et al.*, soumis) ont montré, en appliquant les modèles d'équations structurales aux données anglaises et françaises de PISA qu'elles n'étaient pas unidimensionnelles, mais à tout le moins bidimensionnelles. L'écart à l'unidimensionnalité est révélateur de failles dans le dispositif de mesure et ses conséquences sur

la définition de la compétence doivent être prises en considération.

Il est certain que cette compétence est bien une compétence largement transversale dont la plus ou moins grande maîtrise peut être considérée comme le produit des systèmes éducatifs. Mais une telle variable relativement décontextualisée puisqu'elle ne doit pas être sensible aux différents contextes linguistiques et culturels, n'est-elle pas une sorte de facteur général de réussite protéiforme susceptible de recevoir de multiples dénominations et interprétations ? Le résultat d'une étude conduite sur la comparaison entre une enquête précédente sur la littéracie auprès d'adultes IALS et PISA conduit également à s'interroger sur la nature des échelles de PISA. L'enquête IALS comprenait trois échelles définies d'après le contenu du support (prose, document et littéracie quantitative). Plusieurs items (15) de l'échelle « Prose » de IALS ont été intégrés dans PISA. Il était donc possible de comparer les deux types d'approches de la littéracie celle de IALS et celle de PISA. Cette étude comparative a été conduite par Yamamoto (2002). Malgré les nombreux biais conduisant à rendre difficile la comparaison entre les deux échelles, Kentaro Yamamoto aboutit à la conclusion que la corrélation entre l'échelle de *prose literacy* de IALS et de PISA est de .83. Ce qui correspond à peu près à l'ordre de grandeur des corrélations entre les sous-échelles de IALS. On peut en conclure que ces deux enquêtes bien que constituées de sous échelles interprétées différemment mesurent globalement la même compétence.

On peut également s'interroger sur le fait que ce facteur peut s'apparenter dans une large mesure à des variables du type des aptitudes intellectuelles, en particulier, du raisonnement verbal. Dans une autre enquête menée dans le cadre d'un projet européen (Vrignaud & Rémond, 2002), on observe une corrélation proche de .70 entre un test de vocabulaire (subtest de vocabulaire du WISC III) et des épreuves nationales d'évaluation de la lecture pour deux pays (l'Angleterre et l'Italie). Bien que l'intensité de ces corrélations ne soit pas suffisamment forte pour assimiler les compétences évaluées par les deux types de tests, elle est néanmoins suffisamment élevée pour faire l'hypothèse qu'une partie relativement importante (près de la moitié de la variance) est expliquée par un test de vocabulaire. Les tests de vocabulaire sont les meilleurs indicateurs du raisonnement verbal et même du raisonnement en général (corrélation élevée avec la mesure globale du QI). Ces tests de lecture mesurent donc également une compétence

verbale très générale. On pourrait s'interroger, au moins pour les niveaux supérieurs de PISA qui, selon leur définition, requièrent que les sujets réalisent des opérations d'inférence, sur le fait qu'on mesure autant la capacité au raisonnement verbal que la capacité à tirer de l'information d'un texte écrit.

La seconde question porte sur l'unidimensionnalité du construit mesuré. Le recours à trois dimensions, même s'il est intéressant d'un point de vue conceptuel, n'apparaît pas pleinement convainquant du point de vue psychométrique. L'agrégation de l'ensemble des items dans une seule variable latente est un point qui ne plaide pas particulièrement en faveur de l'utilisation de plusieurs sous échelles. Les contraintes du modèle de mesure sont telles qu'elles conduisent à éliminer toutes les causes éventuelles d'écart à l'unidimensionnalité qui seraient en violation avec l'utilisation des MRI. On peut considérer que cette réduction va s'opérer dès la sélection des items. Par conséquent, l'univers des items risque d'éliminer des informations porteuses de différences qualitatives supportant d'autres dimensions et non plus seulement des différences quantitatives consistant à ordonner les moyennes des pays sur un axe.

On peut également s'interroger sur la pertinence d'expliquer les différences entre sujets de manière uniquement quantitative pour les sujets faiblement compétents dont la situation est mieux qualifiée par le terme d'illettrisme que par celui de niveau faible de littéracie. Il est plus heuristique de chercher à qualifier ces situations d'illettrisme en identifiant leurs causes plutôt que de les quantifier. L'enquête sur les compétences en littéracie des adultes français « Information et vie quotidienne » (Murat, 2005) comportait un module particulier pour les sujets identifiés comme étant en situation d'illettrisme visant à diagnostiquer les causes de cet illettrisme.

Le choix fait par des enquêtes de type PISA d'évaluer des compétences n'est pas exempt de tout questionnement scientifique et idéologique. En effet, on se souvient des débats sur la mesure de l'intelligence et de la boutade de Binet. On court ici le risque de

déclarer « la compétence ? c'est ce que mesure notre test ! ». Comment être sûr que l'on échantillonne les items (les tâches) de manière à réellement balayer le domaine ? Ne court-on pas le risque comme dans les tests d'intelligence de sur-représenter voire de ne représenter que les tâches en relation avec les apprentissages scolaires et le milieu culturel dominant tels qu'ils sont conçus et valorisés dans certains pays et d'assister aux terribles dérives apparues dans le domaine des aptitudes avec les travaux de Terman comme l'évoquent A. Blum et F. Guérin-Pace (2000) ? Il y a un risque de dérive idéologique à considérer ces compétences comme dotées d'une réalité autonome et objective alors qu'elles sont étroitement dépendantes d'un modèle de mesure.

Si l'on choisit une approche des compétences, alors, il est nécessaire de définir les compétences en termes de domaines, opération qui seule pourra valider l'interprétation de la mesure psychométrique puisqu'elle permettra de vérifier la couverture du domaine de la compétence par les épreuves construites. Cette approche a été l'objet d'une enquête internationale pilotée par l'OCDE : le programme DESECO – « Définitions et sélections des compétences » (OCDE, 2002). Il s'agissait de demander à différents experts : philosophes (Canto-Sperber & Dupuy, 2001), ethnologue (Goody, 2001), psychologue (Haste, 2001), économistes (Levy & Murnane, 2001), spécialistes des sciences de l'éducation (Perrenoud, 2001) comment on pourrait définir les compétences nécessaires pour vivre et réussir dans le monde moderne. Ce type de travaux pourrait permettre de définir les compétences évaluées sur des bases théoriques et non uniquement psychométriques. La validité du construit et son interprétation s'en trouveraient davantage validées. Il ne semble pas malheureusement que les résultats de DESECO aient été injectés dans les réflexions sur les enquêtes internationales d'évaluation des compétences.

Pierre Vrignaud
vrignaud.pierre@wanadoo.fr
Université Paris 10-Nanterre
Laboratoire « Travail et évolution professionnelle »

NOTES

(1) En anglais, le terme d'*Item Response Theory* (IRT), est plus largement utilisé. Le terme de modèle paraît plus approprié dans la mesure où il s'agit de rendre compte du comportement du sujet répondant à un item plutôt que de construire une théorie psychologique du comportement du sujet comme le font remarquer H. Goldstein & R. Wood (1989).

(2) L'anglais utilise le terme de *guessing* (parfois traduit en français par « pseudo-chance ») pour désigner, principalement dans les QCM, la probabilité de « deviner » la bonne réponse ou de la donner par hasard. On a jugé utile d'introduire ce paramètre dans les MRI pour rendre compte du fait que la probabilité de bonne réponse d'un sujet ne devient pas infiniment petite au

fur et à la mesure que la compétence de ce sujet est estimée faible mais peut rester dans une zone nettement plus élevée. Par exemple dans le cas d'un QCM comprenant quatre possibilités de réponse, la possibilité de donner la bonne réponse au hasard serait de 25 %. Dans ce cas, le paramètre de *guessing* estimerait la probabilité à ce seuil même pour des sujets de compétence faible.

- (3) ACER est l'organisation principale en charge du consortium qui a géré PISA, ETS a été l'organisation en charge du traitement des données des enquêtes américaines (NAEP, etc.) ainsi que de plusieurs enquêtes internationales, en particulier, IALS.
- (4) Lorsque la différence de réussite à l'item est de même sens en faveur ou en défaveur du même groupe dans toutes les classes de sujets le FDI est dit « uniforme ». Le FDI uniforme porte uniquement sur la difficulté de l'item. Il existe un écart en faveur du même groupe à tous les niveaux de compétence. Lorsque la différence de réussite change de sens selon le niveau de performance des sujets (par exemple la différence est en faveur d'un groupe pour les classes de performance faibles et en défaveur du même

groupe pour les classes de performance élevée) on parle de FDI « croisé ». Le FDI croisé porte sur la discrimination de l'item celui-ci est plus discriminant dans un groupe que dans l'autre. Si on se représente aisément la signification psychologique d'un FDI uniforme, celle d'un FDI croisé peut être plus délicate.

- (5) L'algorithme EM estime, selon la méthode du maximum de vraisemblance, les paramètres de distributions expliquant un échantillon de données lorsqu'on est en présence de données manquantes, en complétant les données par une variable aléatoire rendant compte de la relation entre les données observées (les réponses aux items) et les données manquantes (ici les paramètres du MRI). Dans une première phase on va calculer l'espérance de la vraisemblance (expectation) et, dans une seconde phase, on va opérer une maximisation (maximisation) de l'espérance obtenue. Puis, on utilise les valeurs trouvées à l'étape de maximisation pour une nouvelle étape d'espérance. On répétera ce processus de manière itérative dont chaque phase augmente la vraisemblance jusqu'à ce qu'on atteigne un critère d'arrêt (en général un écart faible entre la vraisemblance à deux étapes consécutives).

BIBLIOGRAPHIE

- ADAMS R. J. (2002). « Scaling PISA cognitive data ». In M. L. Wu & R. J. Adams (éd.), *PISA 2000 : Technical Report*. Paris : OECD, p. 99-108.
- ADAMS R. J. & CARSTENSEN C. (2002). « Scaling outcomes ». In M. L. Wu & R. J. Adams (éd.), *PISA 2000 : Technical Report*. Paris : OECD, p. 149-162.
- ADAMS R. J. & KHOO S. J. (1994). *QUEST : The Interactive Test Analysis System Version 2.0*. Hawthorn : ACER.
- ADAMS R. J. & WU M. L. (2002). *PISA : Technical report*. Paris : OECD.
- BAKER F. B. (1992). *Item Response Theory : parameter, estimation techniques*. New York : M. Dekker.
- BEATON A. E. & JOHNSON E. G. (1992). « Overview of the scaling methodology used in the national assessment ». *Journal of Educational Measurement*, vol. 29, p. 163-175.
- BLUM A. & GUÉRIN-PACE F. (2000). *Des lettres et des chiffres*. Paris : Fayard.
- BOCK R. D. & AITKIN M. (1994). « Marginal maximum likelihood of item parameters : Application of an EM algorithm ». *Psychometrika*, vol. 46, p. 443-459.
- BOLT D. & STOUT W. (1996). « Differential item functioning : Its multidimensional model and resulting subtest detection procedure ». *Behaviormetrika*, vol. 23, p. 67-95.
- BOTTANI N. & VRIGNAUD P. [dir.] (2005). *La France et les évaluations internationales*. Rapport établi à la demande du Haut conseil de l'évaluation de l'école. Paris : Haut conseil de l'évaluation de l'école. Disponible sur au format PDF sur Internet à l'adresse : http://cisad.adc.education.fr/hcee/documents/rapport_Bottani_Vrignaud.pdf (consulté le 8 janvier 2007).
- BRADLOW E. T. ; WAINER H. & WANG H. (1998). « A bayesian random effects model for testlets ». In *ETS Research Report RR-98-3*. Princeton : Educational Testing Service.
- CANTO-SPERBER M. & DUPUY J.-P. (2001). « Competencies for the good life and the good society ». In D. S. Rychen & L. H. Salganik (éd.), *Defining and Selecting Key Competencies*. Göttingen : Hogrefe & Huber, p. 67-92.
- CORTINA J. M. (1993). « What is coefficient alpha : An examination of theory and application ». *Journal of Applied Psychology*, vol. 78, p. 98-104.
- CRONBACH L. J. & MEEHL P. E. (1955). « Construct validity in psychological tests ». *Psychological Bulletin*, vol. 52, 281-302.
- DEMPSTER A. P. ; LAIRD N. M. & RUBIN D. B. (1977). « Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion) ». *Journal of the Royal Statistical Society, Series B*, vol. 39, p. 1-38.
- DICKES P. & FLIELLER A. (1997). *Analyse secondaire des données françaises de la première enquête internationale sur la littéracie des adultes (enquête IALS)*. Rapport pour le ministère de l'Éducation nationale. Paris : ministère de l'Éducation nationale ; Nancy : université de Nancy 2 : Laboratoire de psychologie : équipe GRAPCO.
- DICKES P. & VRIGNAUD P. (1995). *Rapport sur les traitements des données françaises de l'enquête internationale sur la littéracie*. Rapport pour le ministère de l'Éducation nationale. Paris : ministère de l'Éducation nationale.
- DICKES P. ; TOURNOIS J. ; FLIELLER A. & KOP J.-L. (1994). *Psychométrie*. Paris : PUF.
- FISCHER G. H. & MOLENAAR I. W. [éd.] (1995). *Rasch models : Foundations, recent developments, and applications*. New-York : Springer.
- GARDNER H. (1983). *Frames of mind*. New-York : Basic Books ; trad. fr. *Les formes de l'intelligence*. Paris : O. Jacob, 1997.
- GOLDSTEIN H. (1980). « Dimensionality, bias, independence and measurement scale problems in latent trait test score models ». *The British Journal of Mathematical and Statistical Psychology*, vol. 33, p. 234-246.
- GOLDSTEIN H. (2004). « International comparisons of student attainment : some issues arising from the PISA study ». *Assessment in Education*, vol. 11, p. 319-330.
- GOLDSTEIN H. & WOOD R. (1989). « Five decades of item response modelling ». *The British Journal of Mathematical and Statistical Psychology*, vol. 42, p. 139-167.

- GOLDSTEIN H. ; BONNET G. & ROCHER T. (soumis). « Multi-level multidimensionnal structural equation models for the analysis of comparative data on Educational performance ». *Journal of Educational and Behavioural Statistics*.
- GONZALEZ E. J. (2003). « Scaling the PIRLS reading assessment data ». In I. V. S. Mullis, M. O. Martin, E. Gonzalez & A. Kennedy, *PIRLS 2001 International Report*. Boston : International Study Center.
- GOODY J. (2001). « Competencies and Education : Contextual Diversity ». In D. S. Rychen & L. H. Salganik (éd.), *Defining and Selecting Key Competencies*. Göttingen : Hogrefe & Huber, p. 175-190.
- HAMBLETON R. K. & SWAMINATHAN H. (1985). *Item Response Theory. Principles and applications*. Boston : Kluwer-Nijhoff.
- HAMBLETON R. K. ; SWAMINATHAN H. & ROGERS H. J. (1991). *Fundamentals of item response theory*. Newbury Park : Sage.
- HASTE H. (2001). « Ambiguity, Autonomy and Agency ». In D. S. Rychen & L. H. Salganik (éd.), *Defining and Selecting Key Competencies*. Göttingen : Hogrefe & Huber, p. 93-120.
- HUTEAU M. & LAUTREY J. (1999). *Évaluer l'intelligence : psychométrie cognitive*. Paris : PUF.
- JOHNSON E. G. (1992). « The design of the National Assessment of Educational Progress ». *Journal of Educational Measurement*, vol. 29, p. 95-110.
- KIRSCH I. S. ; JUNGEBLUT A. & MOSENTHAL P. B. (1998). « The measurement of adult literacy ». In T. S. Murray, I. S. Kirsch & L. B. Jenkins (éd.), *Adult Literacy in OECD countries. Technical report on the first international adult literacy survey*. Washington [D. C.] : US Department of Education : National Center for Education Statistics, p. 105-134.
- KOLEN M. J. & BRENNAN R. L. (1995). *Test Equating. Methods and practices*. New-York : Springer
- LAVEAULT D. & GREGOIRE J. (2002). *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck.
- LEVY F. & MURNANE R. J. (2001). « Key Competencies Critical to Economic Success ». In D. S. Rychen & L. H. Salganik (éd.), *Defining and Selecting Key Competencies*. Göttingen : Hogrefe & Huber, p. 151-174.
- LORD F. & NOVICK M. R. [éd.] (1968). *Statistical theories of mental test scores*. Reading : Addison-Wesley.
- MISLEVY R. J. (1987). « Exploiting auxiliary information about examinees in the estimation of item parameters ». *Applied Psychological Measurement*, vol. 11, p. 81-91.
- MISLEVY R. J. (1994). « Evidence and inference in educational assessment ». *Psychometrika*, vol. 59, p. 439-483.
- MISLEVY R. J. ; BEATON A. E. ; KAPLAN B. & SHEEHAN K. M. (1992). « Estimating population characteristics from sparse matrix samples of item responses ». *Journal of Educational Measurement*, vol. 29, p. 133-161.
- MISLEVY R. J. & BOCK R. D. (1990). *BILOG 3 : Item analysis and test scoring with binary logistic models*. Mooresville : Scientific Software [2^{de} éd.]
- MURAT F. (2005). « Les compétences des adultes à l'écrit, en calcul et en compréhension orale ». *INSEE Première*, n° 1044, 4 p. Document disponible au format PDF sur Internet à l'adresse : http://www.insee.fr/fr/ffc/docs_ffc/IP1044.pdf (consulté le 25 janvier 2007).
- MURRAY T. S. ; KIRSCH I. S. & JENKINS L. B. [éd.] (1998). *Adult Literacy in OECD countries. Technical report on the first international adult literacy survey*. Washington [D. C.] : US Department of Education : National Center for Education Statistic, p. 105-134.
- NANDAKUMAR R. (1994). « Assessing dimensionality of a set of item responses-Comparison of different approaches ». *Journal of Educational Measurement*, vol. 31, p. 17-35.
- OCDE (1999). *Mesurer les connaissances et compétences des élèves : un nouveau cadre d'évaluation*. Paris : OCDE.
- OCDE (2002). *Définitions et sélections des compétences (DESECO) : fondements théoriques et conceptuels : document de stratégie*. Neuchâtel : OFS. Document disponible au format PDF sur Internet à l'adresse : http://www.portal-stat.admin.ch/desecco/desecco_doc_strategique.pdf (consulté le 8 janvier 2007).
- PERRENOUD P. (2001). « The key to social fields : Competences of an Autonomous Actor ». In D. S. Rychen & L. H. Salganik (éd.), *Defining and Selecting Key Competencies*. Göttingen : Hogrefe & Huber, p. 121-150.
- RASCH G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen : Nielsen & Lydiche.
- RASCH G. (1977). « On specific objectivity. An attempt at formalizing the request for generality and validity of scientific statements ». *Danish Yearbook of Philosophy*, vol. 14, p. 58-94.
- REMOND M. (2006). « Éclairages des évaluations internationales PIRLS et PISA sur les élèves français ». *Revue française de pédagogie*, n° 157, p. 71-84.
- REUCHLIN M. (1996). *Psychologie différentielle*. Paris : PUF [n^{lle} éd. aug.]
- ROCHER T. (2003). « La méthodologie des évaluations internationales de compétences ». *Psychologie et Psychométrie*, n° 24, p. 117-146.
- RUBIN D. B. (1991). « EM and beyond ». *Psychometrika*, vol. 56, p. 241-254.
- RUBIN D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York : Wiley.
- RYCHEN D. & SALGANIK L. [éd.] (2003). *Key Competencies for a Successful Life and a Well-Functioning Society*. Göttingen : Hogrefe & Huber.
- SALGANIK L. H. ; RYCHEN D. S. ; MOSER U. & KONSTANT J. W. (1999). *Projects on competencies in the OECD context : Analysis of theoretical and conceptual foundations*. Neuchâtel : Office fédéral des statistiques.
- SALGANIK L. & RYCHEN D. [éd.] (2001). *Defining and Selecting Key Competencies*. Seattle : Hogrefe & Huber.
- SALINES M. & VRIGNAUD P. (2001). *Apprécier et certifier les acquis des élèves en fin de collège : diplôme et évaluations-bilans*. Rapport établi à la demande du Haut conseil pour l'évaluation de l'école. Paris : Haut conseil de l'évaluation de l'école.
- SCHAFFER J. L. & GRAHAM J. W. (2002). « Missing data : Our view of the state of the art ». *Psychological Methods*, vol. 7, p. 147-177.
- SCHAFFER J. L. & OLSEN M. K. (1998). « Multiple imputation for multivariate missing-data problems : A data analyst's

- perspective ». *Multivariate Behavioral Research*, vol. 33, p. 545-571.
- SHEALY R. T. & STOUT W. F. (1993). « A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DIF as well as item bias/DIF ». *Psychometrika*, vol. 58, p. 159-194.
- SHEEHAN K. & MISLEVY R. J. (1990). « Integrating cognitive and psychometric models to measure document literacy ». *Journal of Educational Measurement*, vol. 27, p. 255-272.
- TURNER R. (2002). « Constructing the proficiency scales ». In M. L. Wu & R. J. Adams (éd.), *PISA 2000 : Technical Report*. Paris : OECD, p. 195-216.
- VAN DER LINDEN W. J. & HAMBLETON R. K. [éd.] (1997). *Handbook of modern item response theory*. New-York : Springer.
- VRIGNAUD P. (1996). « Les tests au XXI^e siècle : que peut-on attendre des évolutions méthodologiques et technologiques dans le domaine de l'évaluation psychologique des personnes ? » *Pratiques psychologiques*, vol. 2, p. 5-28.
- VRIGNAUD P. (2002). « Les biais de mesure : savoir les identifier pour y remédier ». *Bulletin de psychologie*, vol. 55, n° 6, p. 625-634.
- VRIGNAUD P. (2003). « Objectivité et authenticité dans l'évaluation : avantages et inconvénients des questions à choix multiples [QCM] et des questions à réponses complexes [QRC] : importance du format de réponse pour l'évaluation des compétences verbales ». *Psychologie et psychométrie*, vol. 24, n° 2-3, p. 147-188.
- VRIGNAUD P. & RÉMOND M. (2002). « The Use of National Reading Tests for International Comparisons : Results from a feasibility study », communication présentée au *XXV^e International Congress of Applied Psychology* (Singapour, juillet 2002).
- WAINER H. & THISSEN D. (1996). « How is reliability related to the quality of test scores ? What is the effect of local dependence on reliability ? », *Educational Measurement : Issues and Practice*, vol. 15, p. 22-29.
- WEINERT F. E. (1999). *Concepts of competence*. Neuchâtel : Office fédéral des statistiques.
- WU M. L. ; ADAMS R. J. & WILSON M. R. (1997). *ConQuest. Generalized item response modelling software*. Hawthorn [Australia] : ACER.
- ZIMOWSKI M. F. ; MURAKI E. ; MISLEVY R. J. & BOCK R. D. (1996). *BLOG-MG. Multiple-Group IRT analysis and test maintenance for binary items*. Chicago : Scientific Software International.

Définition des compétences, choix méthodologiques et retombées sur la politique scolaire en Suisse

Jean Moreau, Christian Nidegger, Anne Soussi

Nous essayerons dans le présent article de mettre en évidence, en illustrant par quelques exemples, les choix méthodologiques et conceptuels pris en compte dans l'enquête PISA. Dans quelle mesure les compétences définies et opérationnalisées permettent-elles la comparaison des systèmes éducatifs ou n'introduisent-elles pas de biais culturels et linguistiques ? Les méthodes statistiques sophistiquées utilisées permettent-elles la transparence et l'appropriation des données par les destinataires des résultats : décideurs politiques et acteurs de l'école ? Dans une deuxième partie, nous montrerons les effets des résultats de PISA sur la politique scolaire en Suisse et plus particulièrement sur la mise en place d'un projet d'harmonisation scolaire national (HarmoS) qui s'inspire largement de démarches mises en œuvre dans l'enquête PISA, notamment approche par compétences, malgré des objectifs différents tels que le pilotage et la régulation des systèmes éducatifs.

Descripteurs (TEE) : compétence, évaluation, mesure, politique de l'éducation, Suisse.

L'enquête PISA vise à mesurer, tous les trois ans, les compétences dans trois domaines, littérature, culture mathématique et culture scientifique. En 2003, cette enquête a concerné plus de quarante pays dont la Suisse. Avec des résultats relativement moyens en littéracie et des résultats au-dessus de la moyenne en mathématiques, ce pays qui se caractérise par la coexistence de trois régions linguistiques et de vingt-six systèmes scolaires différents, a subi un électrochoc (comme d'autres pays, par ailleurs) En effet, la Suisse qui avait jusqu'alors peu participé aux enquêtes internationales était persuadée de l'excellence de son système scolaire. Des mesures ont

été mises en place au plan national suite aux résultats de l'enquête PISA 2000.

Nous allons dans le présent article essayer de mettre en évidence, en illustrant par quelques exemples, les choix méthodologiques et conceptuels pris en compte dans l'enquête PISA. Ainsi, un choix a été fait par les concepteurs, de mesurer des compétences qui seront utiles « tout au long de la vie ». Ce choix pose un certain nombre de questions : dans quelle mesure les compétences définies et opérationnalisées permettent-elles la comparaison de systèmes éducatifs différents et n'introduisent-elles pas

de biais culturels et linguistiques ? Les méthodes statistiques sophistiquées utilisées permettent-elles la transparence et l'appropriation des données par les destinataires des résultats : décideurs politiques, acteurs de l'école (directions des écoles, enseignants, parents) ?

Dans une deuxième partie, nous montrerons les effets des résultats de PISA sur la politique scolaire en Suisse et plus particulièrement sur la mise en place d'un projet d'harmonisation scolaire national (« HarmoS ») se caractérisant notamment par l'entrée à l'école à quatre ans et l'élaboration de standards minimaux de compétences dans quatre domaines scolaires, Langue I (langue d'enseignement), Langue II (première langue étrangère), mathématiques et sciences, dont trois sont identiques à ceux de PISA, et ceci à trois moments-clés du cursus scolaire, 2^e, 6^e et 9^e années. Il est à noter que ce projet s'inspire largement de démarches utilisées lors de la mise en œuvre de l'enquête PISA, approche par compétences et méthodologie statistique de même nature.

PISA : PROBLÈMES LIÉS À UNE UTILISATION LOCALE

Construction du matériel de test

Pour constituer le matériel nécessaire à l'enquête, et assurer une diversité linguistique et culturelle, des directives précises ont été définies afin que chaque pays puisse proposer des items. En fait, pour l'enquête principale 2003 dans le domaine des mathématiques, plus de 60 % des items proviennent de trois instituts (australien américain et néerlandais) chargés de l'enquête, quelques items issus de l'enquête TIMSS. Les 40 % restant proviennent d'une douzaine de pays. L'ensemble des items proposés par ces pays a été adapté par ces instituts. Un groupe d'experts internationaux a assuré la supervision de l'ensemble. On notera que plus de 70 % des items ont été proposés en anglais, 14 % en néerlandais (provenant d'un des instituts pilotant l'enquête). Les autres items ont été proposés en allemand, espagnol, japonais, tchèque et français. L'ensemble du matériel retenu a été soumis aux pays participants afin qu'ils puissent faire part de leurs remarques et de leurs éventuels désaccords.

Un point important à respecter pour la mise en œuvre d'une enquête internationale à visée comparative est la question de la traduction et de l'adaptation

du matériel. Pour ce faire, PISA a conçu un système de traduction à partir de deux sources (l'une en anglais et l'autre en français). Les pays étaient ensuite invités à harmoniser les deux traductions. Cependant, la traduction à partir de deux sources n'a pas toujours pu être respectée. De plus, chaque pays pouvait proposer des adaptations nationales afin de s'assurer que les textes proposés soient plus proches du contexte linguistique local. Toutefois, c'est bien le même matériel qui est soumis aux élèves. Ainsi, dans le but de la comparaison, la diversité culturelle est bien envisagée comme une difficulté à surmonter et à réduire et non comme un élément constitutif du domaine à mesurer. Cela peut avoir pour conséquence que pour garantir les possibilités de comparaison, on définisse un plus petit dénominateur commun qui soit éloigné de toutes les situations locales existantes.

Lors de chaque prise d'information, un des trois domaines est particulièrement étudié. À cet effet, un certain nombre d'échelles sont construites afin de mieux décrire les compétences des élèves dans ce domaine. Pour l'enquête 2003, dont le domaine principal était les mathématiques, il avait été prévu au départ de constituer des échelles de type « compétences » manifestées par les élèves (reproduction, connexions, réflexion), comparables aux sous échelles utilisées pour la littéracie en 2000 (retrouver l'information, développer une interprétation et réfléchir sur le texte). Cependant, il s'est avéré que ces échelles étaient non seulement trop fortement corrélées mais aussi hiérarchiques. C'est pourquoi les concepteurs ont opté pour des échelles qui reflétaient différents aspects des contenus mathématiques. Dans le tableau (page suivante), on notera que si ces différentes échelles comprennent à peu près le même nombre d'items, le type de questions varie toutefois, considérablement d'une échelle à l'autre. Les résultats des élèves pourraient alors dépendre plus du type de question que du contenu de l'échelle. Or, on sait d'expérience qu'il est potentiellement plus facile de répondre à une question à choix multiple que de formuler soi-même une réponse. Dans l'enquête PISA 2000 sur la littéracie, nous avons pu vérifier cet effet (Broi *et al.*, 2003).

Adéquation du modèle statistique aux analyses au niveau local

La validation du matériel a été réalisée à partir d'un double processus : 1) empirique par des experts qui ont classé l'ensemble du matériel selon les objectifs visés et la difficulté attendue des items ; 2) statistique sur la

**Tableau 1. – PISA 2003 : items de mathématiques
(répartition selon le type de contenu et le format des items)**

Type de contenu	Format des items			Total
	Items à choix multiple	Items à réponse construite fermée	Items à réponse construite ouverte	
Espace et formes	8	6	6	20
Quantité	6	2	15	23
Variations et relations	3	4	15	22
Incertitude	11	1	8	20
Total	28	13	44	85

Source : OCDE, PISA 2003, Technical Report, p. 29 (traduit et adapté par nos soins).

base des résultats au pré-test réalisé dans l'ensemble des pays participants à l'enquête, l'année précédant le recueil de données principal. Une série d'aller-retour entre ces processus a permis de sélectionner le matériel qui sera retenu pour l'enquête principale.

La mesure des compétences des élèves implique la prise en compte de nombreux aspects. Ces aspects concernent bien sûr l'instrument lui-même : le choix des textes et des différents items proposés aux élèves, mais aussi la manière dont les réponses aux questions vont être utilisées pour produire des scores. Le choix d'un modèle par les concepteurs du projet (le modèle de Rasch) a des conséquences sur les évaluations des compétences moyennes des pays participant à cette enquête. Rappelons que ce modèle (cas particulier des modèles de réponse à l'item) définit la relation entre la compétence de l'élève, la difficulté de l'item et sa probabilité de réussir correctement cet item. Cette relation est exprimée par une fonction mathématique (la fonction caractéristique de l'item) et elle est représentée par une courbe en forme de S. Tous les items sont supposés avoir le même pouvoir discriminatif. Les compétences des élèves et les difficultés des items alors sont situées sur le même *continuum* (1).

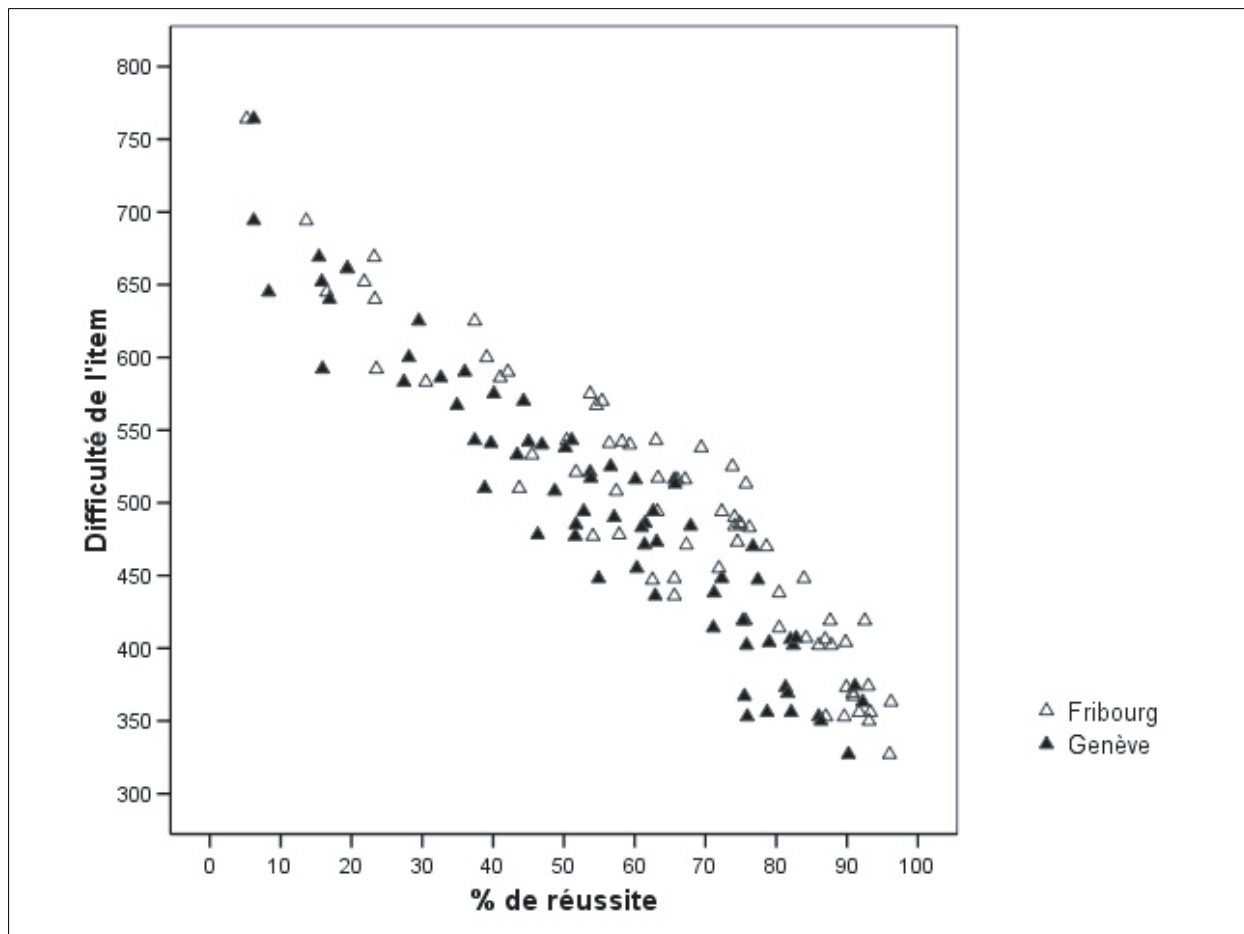
On peut donc s'interroger sur certaines conséquences liées au choix du modèle. Tout d'abord, la difficulté d'une question telle qu'elle est définie par le modèle sur le plan international est-elle cohérente avec les différentes réalités locales ? En particulier des questions de difficulté équivalente ont-elles des réussites identiques au niveau local, la Suisse

romande dans notre exemple ? En outre, la difficulté d'un item, telle qu'elle transparait sur le plan statistique, est-elle compatible avec sa définition critérielle par rapport au contenu de chaque tâche ?

Rappelons en effet, que différentes analyses statistiques de l'épreuve ont été mises en œuvre pour s'assurer des propriétés psychométriques des items dans chaque pays, en particulier : leur discrimination, leur difficulté et leur adéquation avec le modèle. Ces propriétés ne doivent pas trop différer entre les pays. Un item peut en effet être supprimé si ses caractéristiques sont insuffisantes.

Malgré ces précautions méthodologiques, on remarque que la calibration des items n'assure pas l'homogénéité des niveaux de difficulté des items dans un même pays. Si, par exemple, on compare le pourcentage de réussite de chaque item pour deux cantons suisses, Fribourg et Genève (items de lecture pour PISA 2000), avec sa difficulté telle qu'elle est évaluée par le modèle, on constate (voir figure page suivante) que certains items de niveau, de difficulté identique selon le modèle statistique, obtiennent néanmoins des pourcentages de réussite très différents dans une même population scolaire. Le calibrage du modèle au niveau international n'a donc pas nécessairement une pertinence pour une évaluation locale des compétences. Cette distorsion introduite par le calibrage international entre les évaluations des difficultés des items peut différer d'une population à l'autre. La comparaison des deux nuages de points montre qu'il ne s'agit pas d'une simple translation des nuages.

Graphique 1. – Difficultés et pourcentages de réussite des items de lecture (PISA 2000)



NB : pour des raisons de lisibilité, seuls 50 % des items (choisis aléatoirement) sont représentés dans le graphique ci-dessus.

Utilisation du modèle de Rasch pour l'évaluation des élèves

L'enquête PISA n'a pas pour objectif l'évaluation des performances de chaque élève, mais plutôt la comparaison de systèmes éducatifs. Pourtant la tentation est grande d'utiliser sans précaution la même méthodologie pour évaluer les élèves. En effet, le modèle choisi permet à la fois d'estimer la difficulté des items et de produire des scores pour chaque élève.

Les différentes tâches proposées aux élèves sont en effet hiérarchisées en fonction de critères de difficulté croissante et devraient permettre de posi-

tionner les élèves sur une échelle de Guttman. Dans une échelle de Guttman, un élève qui répond correctement à une question doit également donner une réponse correcte à toutes les questions plus faciles. On constate cependant qu'un élève qui réussit un item de difficulté donné peut échouer à des items plus faciles. En outre, un nombre important d'élèves de niveau faible (par exemple, niveau 1) réussissent des items de difficultés élevées (niveau 4 ou 5). En fait, conséquence directe du modèle de Rasch, aucun élève, même parmi les moins performants, n'a une probabilité nulle de répondre correctement à un item, même un item difficile. Des élèves peu performants ont donc pu réussir des items de niveau élevé. On constate, par exemple (PISA 2000) qu'un nombre

important d'entre eux (plus de 50% des élèves de niveau 1) répondent correctement à au moins une question difficile (niveau 4 ou 5). On comprend que les élèves les moins performants sont des élèves qui n'ont répondu correctement qu'à peu de questions et surtout à des questions faciles.

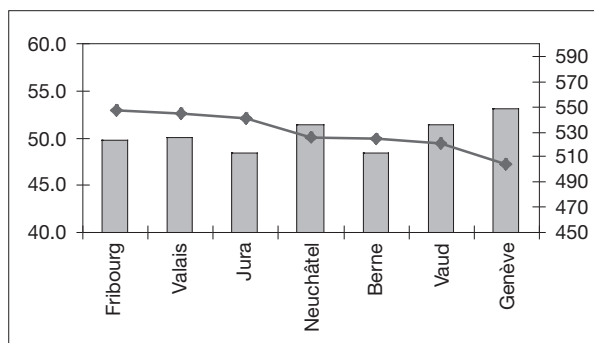
On doit également remarquer que pour un tel modèle, des élèves avec des profils de réponses différentes ont pu obtenir le même score. En effet, le score de l'élève ne dépend pour un cahier donné que du nombre d'items réussis (score brut) et non de la difficulté des items réussis. Deux élèves qui répondent correctement au même nombre de questions – du même cahier (2) – obtiennent le même score quelle que soit la difficulté des questions auxquelles ils ont répondu. Autre conséquence surprenante, un élève peut être classé dans un niveau de lecture donné sans être confronté à des items de ce niveau. Par exemple, des élèves ont pu obtenir un niveau 5 sans être confrontés à des items de niveau 5. Tout dépend du nombre d'items réussis. Certains cahiers, en effet, ne présentent aucun item de niveau 5 concernant l'interprétation du texte. Pourtant des élèves confrontés à ces cahiers ont obtenu le niveau 5.

Utilisation des indices internationaux au niveau régional : l'exemple du niveau socioéconomique

Les performances des élèves sont mises en relation avec un certain nombre d'indices comme par exemple un indice de niveau socioéconomique. De notre point de vue, cet indice pose des questions au niveau de son utilisation pour des analyses régionales. En effet, dans la plupart des études, on observe une relation entre l'indice socioéconomique et la performance des élèves. Ce phénomène est également mis en évidence dans PISA 2003 (OCDE, 2004, chapitre 4). Par contre, lorsqu'au niveau régional on met en regard le niveau socioéconomique moyen et la performance des élèves, on s'aperçoit que ce ne sont pas les cantons qui ont le niveau socioéconomique moyen le plus élevé qui obtiennent la performance moyenne la meilleure (voir graphique 2).

Comment expliquer ce phénomène ? On peut par exemple se demander si la moyenne occulte des structures différentes des niveaux socioéconomiques entre les cantons. Le graphique 3 montre que si l'on répartit les élèves en quatre groupes en fonction de leur niveau socioéconomique, on observe par exemple que Genève a plus d'élèves de niveaux 3 et 4 que des deux autres niveaux (1 et 2) comparativement aux autres cantons. Ceci nous amène à faire l'hypothèse

Graphique 2. – Moyenne en mathématiques et niveau socioéconomique (NSE) moyen

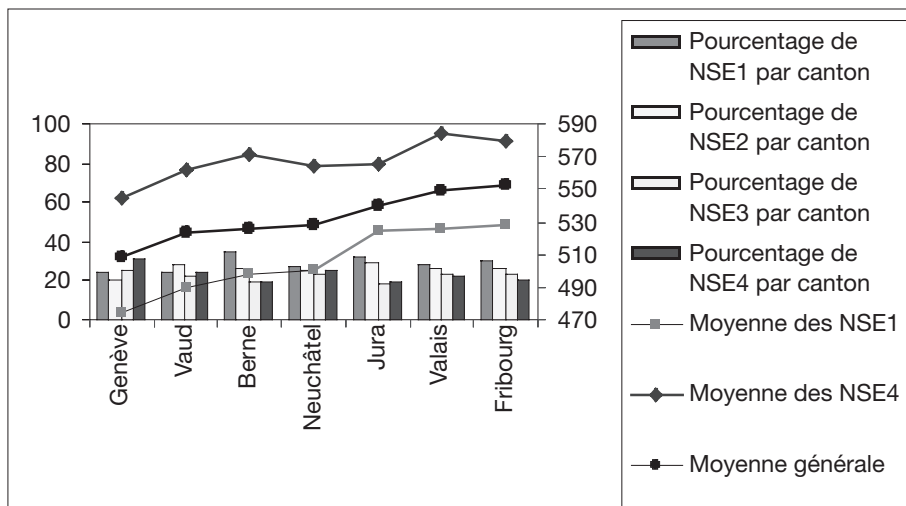


N.B. : Les histogrammes représentent le niveau socioéconomique (NSE) moyen (échelle de gauche). La courbe indique les scores moyens en mathématiques (échelle de droite).

d'une distorsion possible de cet indice international qui « favoriserait » les professions des services en leur attribuant une valeur d'indice plus élevée qu'à des professions « artisanales » qui exigent des qualifications égales ou supérieures et qui sont plus représentées dans les cantons moins urbains. Ainsi, le niveau socioéconomique moyen des cantons urbains serait en quelque sorte surévalué par une forte présence de professions donnant une valeur élevée d'indice bien que ne nécessitant pas de grandes qualifications. À l'inverse, le niveau socioéconomique moyen des cantons moins urbains (Fribourg, Valais, Jura et Berne) aurait tendance à être sous-évalué compte tenu de la proportion plus grande de professions donnant lieu à une valeur de cet indice plus faible malgré des professions à qualifications relativement élevées.

Par contre lorsque l'on met en relation la moyenne du niveau socioéconomique et les performances à l'intérieur des cantons en fonction de l'orientation des élèves, on constate bien une hiérarchie des performances en fonction de cet indice du niveau socioéconomique. En effet, la relation entre performance et niveau socioéconomique semble bien réelle mais l'indice utilisé par PISA n'est pas pertinent pour certaines comparaisons entre cantons. Ainsi lorsque l'on travaille sur les données à un niveau local, il est indispensable de se poser la question de la pertinence ou de la distance entre les instruments, conçus pour une comparaison à large échelle, et leurs exploitations locales afin d'éviter des interprétations abusives ou incohérentes des résultats observés.

Graphique 3. – **Moyenne en mathématiques et niveau socioéconomique (NSE)**



N.B. : Pour le niveau socioéconomique (NSE), les élèves de la Suisse romande ont été répartis en quatre catégories égales (de la plus basse à la plus élevée).

RÉPERCUSSIONS SUR LA POLITIQUE EN SUISSE ET MODÈLES DE COMPÉTENCES

Suite aux premiers résultats assez moyens obtenus à l'enquête PISA 2000, la Suisse par l'intermédiaire de la Conférence des directeurs de l'instruction publique (3) a décidé de mettre en place un certain nombre de mesures dont le lien avec PISA n'est pas toujours évident. Au plan cantonal, de nombreux plans d'action visant à développer et renforcer les compétences en lecture voire en littéracie ont été élaborés.

La Suisse est un système fédéral avec vingt-six cantons et autant de systèmes éducatifs et comporte trois régions linguistiques (germanophone, francophone, italophone). Ainsi, selon les cantons, qui sont responsables et souverains en matière scolaire, l'âge d'entrée à l'école n'est pas le même, le découpage entre les différents niveaux d'enseignement (primaire, secondaire I, etc.) est variable ainsi que la dotation horaire pour les différentes disciplines d'enseignement. Dans la liste des mesures proposées au niveau fédéral, on trouve un ambitieux projet d'harmonisation scolaire nationale (HarmoS) dont les principales propositions sont les suivantes :

- entrée dans l'école obligatoire à quatre ans révolus ;
- unification des finalités de l'école obligatoire ;

- description des compétences de base que doivent atteindre tous les élèves à la fin de la 2^e, 6^e et 9^e années scolaires, au moyen de standards nationaux de formation ;
- élaboration d'un plan d'étude cadre pour chaque région linguistique ;
- vérification régulière au niveau du système scolaire de l'atteinte des standards (et mesures de remédiation mises en œuvre si nécessaire).

De manière générale, le projet HarmoS a pour objectifs d'harmoniser les contenus de l'enseignement, de renforcer le pilotage coordonné de la scolarité obligatoire, d'évaluer le système de formation à l'échelle nationale et de développer la qualité de l'enseignement. Outre la définition de standards que nous développerons plus en détail, HarmoS prévoit d'autres instruments de pilotage tels que le monitoring, les plans d'étude et moyens d'enseignement ainsi que la formation des enseignants. Le monitoring vise à recueillir des données systématiques sur le système éducatif et son environnement, notamment au moyen de tests standardisés effectués régulièrement pour fournir des informations plus fiables sur la réussite, la qualité et la faiblesse du système et constitue un instrument de travail pour les responsables politiques permettant une prise de décision pour son pilotage. Concernant plans d'étude, moyens d'enseignement

et formation des enseignants, on cherche à mettre en place une coordination, notamment par région linguistique, et une professionnalisation dans le cas de la formation des enseignants afin d'améliorer la qualité de l'enseignement.

Modèles de compétences et élaboration de standards

Dans cette première phase de l'harmonisation, la définition de standards et l'élaboration de modèles de compétences concernent seulement quatre domaines de compétences : la langue I (ou d'enseignement), les langues II, les mathématiques et les sciences. S'il s'agit de domaines pour lesquels la dotation horaire est relativement importante tout au long de la scolarité, on peut aussi observer une certaine ressemblance avec les domaines mesurés dans PISA (littérature, culture mathématique et culture scientifique). Comme on peut le constater, ce projet au travers de modèles de compétences s'inscrit dans les courants actuels en éducation. Comme le relève M. Crahay (2006), on est passé de pédagogie par objectifs à une logique de la compétence.

Les responsables au niveau national ont réuni des experts de différents pays dont E. Klieme qui a fait un certain nombre de propositions en vue d'élaborer des standards sur la base de l'expérience menée en Allemagne (initiée elle aussi après les résultats de l'enquête PISA 2000).

Le projet suisse repose sur des postulats définis par E. Klieme *et al.* (2003, p. 22). La compétence est définie de la manière suivante : « En accord avec Weinert (2001, p. 27), nous entendons par compétences les capacités et aptitudes cognitives dont l'individu dispose ou qu'il peut acquérir pour résoudre des problèmes précis, ainsi que les dispositions motivationnelles, volitives et sociales qui s'y rattachent pour utiliser avec succès et responsabilité les résolutions de problèmes dans des situations variables. »

Les compétences sont donc considérées ici comme des dispositions, « c'est-à-dire comme des programmes des représentations, des mécanismes et des capacités internes qui déterminent les comportements et réalisations observables des individus » (Klieme *et al.*, 2003, p. 152).

Le concept de compétence a de nombreuses orientations, notamment dans le monde francophone. D'autres définitions existent dans le monde francophone. Ainsi, L. Allal (1999), la définit de la manière suivante : c'est « un réseau intégré et fonctionnel

formé de composantes cognitives, affectives, sociales, sensorimotrices, susceptible d'être mobilisé en actions finalisées face à une famille de situations et fondé sur l'appropriation de modes d'interaction et d'outils socioculturels. » D'autres auteurs comme X. Roegier et P. Joannaert en ont une représentation assez proche. Le second ajoute toutefois un élément intéressant : « Une compétence suppose, au-delà du traitement efficace, que ce même sujet pose un regard critique sur les résultats de ce traitement qui doit être socialement acceptable. » (2002, p. 41). Il introduit l'idée d'un regard critique du sujet qui doit non seulement mobiliser différentes ressources mais aussi les sélectionner et les coordonner.

Pour B. Rey *et al.* (2003), face à ce magma d'acceptions du terme compétences, il est nécessaire d'introduire des niveaux. Certaines sont dites de *premier degré* (procédures de base), d'autres de *deuxième degré* où il s'agit de posséder des procédures de base et de savoir sélectionner la bonne dans une situation donnée. Le troisième type, *compétence de troisième degré*, consiste à combiner plusieurs procédures de base face à une situation nouvelle et complexe.

Enfin, une autre représentation du concept vient du monde du travail. Pour G. Le Boterf (1995), une compétence est « un savoir-agir, c'est-à-dire un savoir intégrer, mobiliser et transférer un ensemble de ressources (connaissances, savoirs, aptitudes, raisonnements, etc.) dans un contexte donné pour faire face aux différents problèmes rencontrés ou pour réaliser une tâche ». Il ajoute et en cela se distingue fondamentalement de Weinert, « il n'y a de compétence que de compétence en acte [...] La compétence ne réside pas dans les ressources (connaissances, capacités, etc.) à mobiliser mais dans *la mobilisation même de ces ressources*. »

On peut observer deux types de définition, celles qui mettent en avant le contexte et la situation (compétence en acte) et celles où la compétence préexiste dans la tête du sujet. Une autre catégorisation réside dans le type de situations (déjà rencontrées ou nouvelles) : B. Rey utilise ce critère dans l'idée d'une gradation de ces compétences.

Par rapport aux standards proprement dits, on peut en recenser essentiellement trois types dans la littérature internationale, selon leur point de centration (*input* ou processus de l'apprentissage scolaire contre *output*) :

- les standards de *contenus* qui font plutôt référence aux résultats scolaires en se centrant sur les objectifs et les contenus d'apprentissages ;

- les standards de *performance*, basés eux aussi sur les résultats mais qui se situent plutôt au niveau des valeurs du test et sont marqués sur une échelle de score ;
- les standards d'*opportunity-to-learn* qui se rapportent aux *input* et processus de l'apprentissage scolaire et décrivent les programmes, le personnel et autres ressources qui doivent être mis à la disposition des établissements et des régions scolaires.

On peut aussi choisir des niveaux-cibles variés : un niveau basique à atteindre par une grande majorité d'élèves (standards *minimum*), un niveau idéal ou encore un niveau moyen d'exigences (standards *réguliers*).

Le choix qui a été fait en Suisse est de définir des standards de formation minimaux. Ces « standards de formation formulent des exigences à l'égard de l'enseignement et de l'apprentissage scolaire. Ils désignent les objectifs du travail pédagogique, en les exprimant en termes de résultats scolaires attendus. Ils concrétisent ainsi la mission éducative que doivent remplir les écoles de formation générale » (Klieme et al., 2003, p. 16). Pour Klieme, ces standards doivent présenter certaines caractéristiques :

- ils doivent donc se référer aux objectifs de l'évaluation et les transposer en exigences concrètes car les objectifs de l'éducation sont le plus souvent des définitions assez générales des contenus de savoirs, des capacités et des aptitudes visés ;
- les objectifs sont concrétisés sous forme de compétences exigées au travers de modèles de compétences qui exposent les aspects, les degrés et les progressions des compétences. Ces compétences se différencient de ce qu'on trouve dans les plans d'études (listes de matières et de contenus d'enseignement) et nécessitent que l'on identifie des dimensions fondamentales des processus d'apprentissage dans un domaine d'enseignement ou une discipline ;
- enfin, en tant que résultats des processus d'apprentissage, ils doivent être concrétisés sous forme d'exercices et de situations-problèmes ainsi que de procédés fiables permettant de saisir le niveau de compétence réellement atteint par les élèves.

En d'autres termes, ils décrivent des compétences fondées sur un modèle et des niveaux de compétences (progression) validés de manière empirique et doivent être suffisamment concrets pour être vérifiés et mesurés à partir de tests validés.

Limites et problèmes

Comme on a pu le constater, seuls quatre domaines ont été retenus, ce qui a un côté relativement réducteur. Par ailleurs, ces standards de formation doivent être mesurables pour tous les élèves ce qui constitue aussi un obstacle important (par exemple évaluer les compétences à l'oral et notamment dans des situations d'interaction). Enfin, on parle de standards minimaux ce qui peut, comme l'ont relevé K. Muller et R. Silver (2005) dans leur étude critique sur les standards, conduire à plus long terme à une baisse du niveau d'exigences.

Une autre difficulté réside dans l'aspect national de l'opération. Les standards de formation sont censés se baser sur les objectifs scolaires définis dans les *curricula*. Or la Suisse possède vingt-six systèmes scolaires différents et trois régions linguistiques. Les conceptions théoriques sous-jacentes peuvent être pas mal contrastées : les unes s'inspirant de courants francophones (France, Québec ou Belgique), les autres des courants germanophones (Allemagne principalement) et les derniers davantage de courants italiens. Deux exemples peuvent être donnés à titre d'illustration : la Suisse romande a déjà fait un travail de coordination et a élaboré un Plan d'étude-cadre romand (PECARO) sur la base de plans d'études de sept cantons défini en terme de compétences ; dans la partie germanophone, ce travail reste à faire (et le sera sur la base des standards). Dans les domaines disciplinaires, il y a des différences importantes d'approches : par exemple, en mathématiques la Suisse romande donne une place importante à la résolution de problèmes ce qui n'est pas le cas dans la partie germanophone. Par ailleurs, dans d'autres cas, certaines notions telle que la proportionnalité (notamment les fractions) sont abordées à l'école primaire dans la partie germanophone ce qui n'est pas le cas en Suisse romande, on observe l'inverse dans le cas de la géométrie et plus particulièrement de l'approche de l'espace. Dans un même ordre d'idées, en Suisse romande, un modèle didactique autour des genres textuels (Dolz & Schneuwly, 1996) a un impact important sur la conception de l'enseignement de la langue, alors que cette dimension n'est pas présente dans les autres régions. Ceci constitue un paradoxe puisqu'on doit à la fois s'inspirer des programmes scolaires tout en étant confrontés, selon les régions, à la prépondérance de courants et modèles théoriques différents, voire contradictoires. Soulignons que les concepteurs de l'enquête PISA ont pris le parti de se détacher des aspects programmatiques contrairement à d'autres enquêtes internationales menée par l'IEA (*Reading literacy* en 1991 ou TIMSS en 1995).

Un exemple : la langue I (langue d'enseignement)

Un modèle de compétences a été défini dans le domaine de la langue I valable pour les trois langues d'enseignement (allemand, français et italien). Pour ce faire, à l'intérieur de ce domaine, quatre sous-domaines (*skills*) ont été définis selon une terminologie courante sur le plan international : lire et écouter (réception) ainsi qu'écrire et parler (production).

Le modèle prévoit trois niveaux de description : 1) les aspects de l'activité langagière ; 2) la description des compétences dans les quatre sous-domaines ; 3) la description des tâches.

Pour P. Sieber, responsable du domaine Langue 1 dans le projet HarmoS (2005), les actions langagières sont le point de départ de la description des compétences. Elles sont réalisées dans différentes modalités (production ou réception) et font appel à différents médias (oral et écrit, communication directe et indirecte). Il définit l'activité langagière en cinq aspects qui ne sont pas hiérarchiques et qui interviennent de façon variable selon les situations : *situer* (reconnaître ou appliquer la fonction de l'activité langagière), *planifier* (pondérer le poids relatif des différents éléments de contenu, les structurer, les inscrire dans le genre textuel choisi, etc.), *réaliser* (il s'agit de l'activité de réception ou de production à proprement parler), *évaluer* (réfléchir et évaluer) et *réparer* (réviser un texte ou relire le texte pour vérifier et compléter sa compréhension). Ces aspects recouvrent partiellement les sous-échelles utilisées dans PISA (*retrouver une information, développer une interprétation et réfléchir sur le texte*). Il faut toutefois préciser que le modèle défini ici doit prendre en compte à la fois la réception et la production. Or, ces compétences partielles, tout à fait adéquates dans le cas de la littéracie (compréhension de l'écrit ou lecture) dans le sens défini pour l'enquête PISA, à savoir la capacité à comprendre et utiliser des textes écrits ainsi qu'à réfléchir à leur propos, n'auraient pas été adaptées pour les quatre sous-domaines et en particulier, écrire et parler. C'est pourquoi dans le modèle de compétences élaborés pour la langue I dans HarmoS on a choisi de considérer des aspects qui dépassent le cadre de la littéracie (réception) et englobent à la fois la réception et l'expression.

Des compétences partielles, dont on trouvera la description complète en annexe, ont été, entre autres, définies pour la lecture, en fonction des cinq aspects retenus. Ainsi, concernant *Situer*, on trouve entre autres, des compétences telles que s'adapter aux situations de communication ou reconnaître

la fonction du texte. Concernant *Planifier*, il s'agit notamment de définir sa propre situation par rapport à la lecture et de s'organiser (identifier l'organisation du texte – macrostructure, éléments typographiques). *Réaliser* comporte de nombreuses compétences partielles, parmi lesquelles : réaliser les correspondances phonèmes-graphèmes ; mobiliser ses compétences langagières (syntaxiques et sémantiques), verbales et non-verbales (organisateur, anaphores, temps verbaux, etc.) ; interpréter, réaliser des inférences et établir des relations causales ; ou encore retrouver des informations explicites et implicites. L'aspect *Réparer* est plus difficile à prendre en compte dans le cas de la lecture : il peut être considéré comme la compétence à contrôler sa lecture (construction du sens) et à revenir sur celle-ci. Enfin, *Évaluer* peut être considéré comme comportant les compétences à dégager les éléments les plus importants, à résumer ou à réfléchir par rapport au contenu et à la forme du texte, à analyser la lecture (plans émotionnel, cognitif et métacognitif), et à identifier l'intention de l'auteur.

On notera que certaines compétences partielles sont les mêmes que dans PISA. Ainsi, *retrouver une information* et *développer une interprétation* sont incluses dans l'aspect *réaliser*, et *réfléchir sur le texte* est une composante d'*évaluer*.

EN GUISE DE CONCLUSION

Les enquêtes internationales du type PISA contribuent à une meilleure connaissance de nos systèmes éducatifs à travers la comparaison. Elles peuvent même permettre des exploitations plus locales. Cependant, ces deux types d'exploitation nécessitent une contextualisation qui prend en compte certaines limites de ces enquêtes, notamment lors du passage du niveau international au niveau local ou réciproquement. Il faut également garder à l'esprit que ces instruments, aussi sophistiqués qu'ils soient, ne peuvent pas prendre en compte tous les aspects de l'éducation. Ils ne sont pas non plus des instruments universels à même de répondre à toutes les questions que l'on pourrait se poser dans le domaine de l'éducation. Les exemples présentés dans cet article montrent quelques-unes de ces limites.

Cependant, la tentation est forte d'utiliser le même type de démarche dans un contexte national ayant d'autres objectifs. C'est ce qui est arrivé en Suisse. En effet, les résultats de l'enquête PISA 2000 ont donné une impulsion décisive à un projet relativement

ambitieux d'harmonisation de la scolarité et notamment au volet visant la définition de standards minimaux nationaux. Ces standards nécessitent le développement d'instruments qui s'inspire largement d'une approche de type PISA. Or, PISA est une évaluation qui permet de comparer à un moment donné de la scolarité les compétences des élèves dans trois domaines spécifiques sur la base d'un référentiel de compétences attendues « universelles » communes, définies sans lien direct avec les programmes d'enseignement.

Dans le cas d'un projet du type « HarmoS », les objectifs sont différents, il ne s'agit plus de comparer des entités nationales ou locales mais de pilotage et de régulation du système éducatif ou plus exactement de plusieurs systèmes éducatifs ayant des langues et des cultures différentes par la définition de standards mesurables. Il est dès lors nécessaire de définir des compétences qui tiennent compte des programmes scolaires tout en prenant en compte leur dimension variée, tâche qui s'avère complexe et délicate. Par ailleurs, une autre difficulté réside dans l'aspect mesurable, lequel est forcément réducteur par rapport aux contenus curriculaires enseignés à l'école.

De plus, au moment où il s'agit d'élaborer des tâches, on est confronté à un autre type de difficulté : certains types de question sont plus faciles à utiliser (questions à choix multiples ou questions à réponse fermée *contra* questions ouvertes où le correcteur devra

interpréter la réponse de l'élève). Or, les résultats de PISA aussi bien en littérature qu'en mathématiques ont montré l'effet du type de questionnement. De plus, le type de test retenu sera de préférence écrit, ce qui a forcément des incidences sur les domaines mesurés : par exemple la production orale (et en particulier en interaction) ou l'expérimentation en sciences.

Pour conclure, PISA a permis des apports non seulement en termes de connaissance des systèmes éducatifs mais également du point de vue des méthodes et de leur implémentation. Cependant, il faut rester conscient des limites de ce type d'enquêtes. En effet, l'élaboration de standards nécessite le développement d'instruments plus diversifiés pour cerner les objectifs variés de formation visés par les systèmes scolaires qui dépassent largement les compétences prises en compte dans les enquêtes comparatives internationales.

Jean Moreau

jean.moreau@unil.ch

Unité de recherche pour le pilotage des systèmes
pédagogiques (URSP), Lausanne

Christian Nidegger

christian.nidegger@etat.ge.ch

Service de la recherche en éducation (SRED), Genève

Anne Soussi

anne.soussi@etat.ge.ch

Service de la recherche en éducation (SRED), Genève

NOTES

(1) Un élève de compétence x a 50 % de chance (dans PISA 2000, 62 %) de répondre correctement à l'item de difficulté x .

(2) Par exemple en 2003, l'ensemble des items a été réparti dans treize cahiers. Chaque item apparaît dans quatre cahiers différents. Les élèves ne répondent qu'à un des treize cahiers.

(3) Cette instance permanente regroupe les chefs de départements (ministres) chargés de l'instruction publique des vingt-six cantons en vue de coordonner leur action sur la base d'accords entre les cantons.

BIBLIOGRAPHIE

ALLAL L. (1999). « Acquisition et évaluation des compétences en situation scolaire ». *Raisons éducatives*, n° 2 (« L'énigme de la compétence »), p. 77-94.

BROI A.-M. ; MOREAU J. ; SOUSSI A. & WIRTHNER M. (2003). *Les compétences en littérature. Rapport thématique de l'enquête PISA 2000*. Neuchâtel : Office fédéral de la statistique (OFS) ; Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP).

CONFÉDÉRATION HELVÉTIQUE : CONFÉRENCE SUISSE DES DIRECTEURS CANTONAUX DE L'INSTRUCTION PUBLIQUE (2004). *HarmoS : finalités et conceptions du projet (juin 2004)*. Berne : CDIP. Document au format PDF disponible sur Internet

à l'adresse : <http://www.edk.ch/PDF_Downloads/Harmos/Harmos_Weissbuch_f.pdf> (consulté le 23 janvier 2007).

CRAHAY M. (2006). « Dangers, incertitudes et incomplétude de la logique de la compétence en éducation ». *Revue française de pédagogie*, n° 154, p. 97-110.

DOLZ J. & SCHNEUWLY B. (1996) « Genre et progression en expression écrite : éléments de réflexion à propos d'une expérience romande », *Enjeux*, n° 37/38, p. 49-75.

EMBRETSON S. E. & REISE S. P. (2000). *Item response theory for psychologists*. Mahwah : L. Erlbaum.

- KLIEME E. et al. (2003). *Expertise. Zur Entwicklung nationaler Bildungsstandards* Bonn : Bundesministerium für Bildung und Forschung [BMBF] : Referat Publikationen. Document au format PDF disponible sur Internet à l'adresse : <http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf> (consulté le 23 janvier 2007).
- LE BOTERF G. (1995). *De la compétence : essai sur attracteur étrange*. Paris : Éd. d'Organisation.
- JONNAERT P. (2002). *Compétences et socioconstructivisme : un cadre théorique*. Bruxelles : De Boeck.
- MÜLLER K. & SILVER R. (2005). « Standards in education. Review of US literature ». In Matthis Behrens (coord.), *Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif*. Neuchâtel : Institut de recherche et de documentation pédagogique.
- NIDEGGER C. [coord.] (2005). *PISA 2003 : Compétences des jeunes romands : Résultats de la seconde enquête PISA auprès des élèves de 9^e année*. Neuchâtel : Institut de recherche et de documentation pédagogique.
- OCDE (2003). *Cadre d'évaluation de PISA 2003. Connaissances et compétences en mathématiques, lecture, science et résolution de problèmes*. Paris : OCDE.
- OECD (2004). *Learning for tomorrow's world : First results from PISA 2003*. Paris : OECD.
- OECD (2005). *Pisa 2003 technical report*. Paris : OECD.
- REY B. ; CARETTE V. ; DEFANCE A. & KAHN S. (2003). *Les compétences à l'école*. Bruxelles : De Boeck.
- ROEGIER X. (2000). *Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement*. Bruxelles : De Boeck.
- SIEBER P. et al. (2005). *Schritte zu einem Kompetenzmodell Erstsprache*. (document provisoire).
- WEINERT F. E. (2001). « Vergleichende Leistungsmessung in Schulen : eine umrittene Selbstverständlichkeit ». In F. E. Weinert (éd.), *Leistungsmessungen in Schulen*. Weinheim & Basel : Beltz Verlage, p. 17-31.

Annexe. – Compétences partielles dans le domaine de la lecture

situer	– repérer le thème principal et le contenu central
	– reconnaître la fonction du texte (en lien avec les genres textuels)
	– s'adapter aux situations de communication
	– trouver l'intérêt de la lecture en liaison avec le texte
planifier	– définir sa propre situation par rapport à la lecture et s'organiser (identifier l'organisation du texte (macrostructure, éléments typographiques))
	– établir les particularités du texte
	– construire un projet de lecture et choisir les stratégies de lecture
réaliser	– réaliser les correspondances phonèmes-graphèmes
	– identifier et décoder au niveau des mots et des phrases
	– mobiliser ses connaissances langagières (syntaxiques et sémantiques), verbales et non-verbales (organismes, anaphores, temps verbaux, etc.)
	– décoder les représentations iconographiques
	– mobiliser les pré-requis et expériences, connaissances du monde
	– interpréter : réaliser des inférences et établir des relations causales, établir des liens entre différentes parties du texte, construire une représentation mentale, créer l'imaginaire, établir une cohérence
	– retrouver des informations implicites et explicites
	– développer une proximité émotionnelle : concernant le thème, les personnages et d'autres éléments
réparer	– identifier les éléments esthétiques, les figures de style et les structures et les utiliser pour construire le sens
	– contrôler sa lecture (construction du sens) et revenir sur sa lecture
évaluer	– dégager les éléments les plus importants, résumer
	– récapituler l'expérience en lecture et évoquer sa compréhension du texte
	– réfléchir par rapport au contenu et à la forme, analyser la lecture (plan émotionnel, cognitif et méta-cognitif), identifier l'intention de l'auteur
	– échanger avec d'autres à propos de l'expérience de la lecture et la signification du texte
	– justifier son point de vue, argumenter

Études complémentaires de PISA 2003 en Allemagne : principaux résultats et enseignements

Manfred Prenzel, Karin Zimmer

*traduit de l'allemand par Roland Boichon
avec la collaboration de la rédaction*

L'étude PISA 2003 a été menée en Allemagne enrichie de nombreux compléments nationaux. Afin de pouvoir procéder à des comparaisons entre les seize Länder, la taille de l'échantillon a été sensiblement augmentée. On a également intégré un échantillon supplémentaire d'élèves issus de l'immigration. Lors d'une deuxième journée d'épreuves, les élèves ont été soumis à des tests supplémentaires en mathématiques et en sciences. Les équipes de direction, mais aussi les enseignants ont été interrogés dans chaque établissement sur leurs charges de travail ainsi que sur leur participation à l'amélioration de la qualité de l'enseignement. On a enfin testé un an plus tard, en mathématiques et en sciences, les élèves ou les classes qui ont participé à PISA 2003, afin de mesurer leurs progrès dans l'apprentissage. Nous montrons dans notre article, d'après les résultats de notre enquête, le profit qu'on peut tirer d'un élargissement de l'étude PISA.

Descripteurs (TEE) : compétence, origine sociale, mathématiques, matière d'enseignement, résolution de problème, sciences de la vie.

INTRODUCTION

Depuis 2000, le « Programme international pour le suivi des acquis des élèves » (PISA ; *Programme for International Student Assessment*) teste tous les trois ans les compétences de base des élèves âgés de quinze ans dans les domaines suivants : mathématiques, sciences et lecture. Ces compétences de base sont conçues comme des conditions pour l'intégration culturelle ainsi que pour la formation continue à l'école, au travail et dans la vie quotidienne.

Cette étude est coordonnée par l'Organisation pour la coopération et le développement économique (OCDE). Elle offre de nombreuses possibilités d'effectuer des études complémentaires plus poussées et étendues au niveau national. Nous présenterons dans cet article comment ces possibilités ont été exploitées en Allemagne pour PISA 2003. Cependant, nous présenterons tout d'abord le cadre international de PISA 2003 pour ensuite présenter les extensions auxquelles a procédé l'Allemagne. Nous rendrons compte en particulier de constatations sur la conformité aux programmes des exercices et des résultats de PISA à partir d'une comparaison entre les *Länder*.

Le cadre international de PISA 2003

Les évaluations PISA se fondent sur des conceptions-cadres au sein desquelles les concepts analysés sont fondés et structurés (OCDE, 1999 & 2003). Nous présenterons tout d'abord brièvement les définitions des compétences analysées lors de PISA 2003. Par *compétence mathématique*, on désigne dans ce cadre « l'aptitude d'un individu à identifier et à comprendre le rôle joué par les mathématiques dans le monde, à porter des jugements fondés à leur propos, et à s'engager dans des activités mathématiques, en fonction des exigences de sa vie en tant que citoyen constructif, impliqué et réfléchi » (OCDE, 2004a, p. 39). La *compétence de lecture* est définie comme la capacité à « comprendre et à utiliser des textes écrits et à réfléchir sur leur sens pour atteindre des objectifs personnels, développer ses propres connaissances et son potentiel et participer à la vie sociale » (OCDE, 2003, p. 108). La *compétence scientifique* désigne « la capacité à d'utiliser des connaissances scientifiques pour identifier les questions auxquelles la science peut apporter une réponse et pour tirer des conclusions fondées sur des faits, en vue de comprendre le monde naturel ainsi que les changements qui y sont apportés par l'activité humaine, et de contribuer à prendre des décisions à leur propos » (OCDE, 2004a, p. 308).

Ces trois domaines de compétence sont analysés lors de tous les cycles PISA avec toutefois un changement du domaine majeur d'évaluation. Pour PISA 2003, ce sont les mathématiques qui se trouvent au centre des recherches et auxquels est consacrée la majeure partie du temps de test. Outre ces trois compétences relativement proches du domaine scolaire, on a, lors de PISA 2003, analysé la résolution de problèmes en tant que compétence transversale. La *compétence de résolution de problèmes* est définie comme la capacité « à utiliser des procédés cognitifs pour se confronter à des problèmes transversaux réels et les résoudre, alors que la voie de résolution n'apparaît pas immédiatement et que les domaines de connaissance à utiliser pour leur résolution ne proviennent pas du seul domaine des mathématiques, des sciences ou de la lecture. » (OCDE, 2003, p. 156)

On reconnaît aux compétences régulièrement évaluées par PISA une importance clé pour ce qui concerne les chances de réussite individuelles dans l'apprentissage et dans la vie. Elles apparaissent en outre comme le fondement du développement politique et économique des sociétés.

PISA évaluant les compétences d'adolescents, les tests sont conçus de manière à ce que les pré-requis pour l'apprentissage ultérieur soient à la fois d'ordre scolaire et extra-scolaire. PISA souligne qu'il est capital de pouvoir comprendre et utiliser les savoirs de manière flexible et adaptée à la situation. Les tâches proposées doivent correspondre, indépendamment des programmes scolaires, à des exigences variées que des adolescents peuvent rencontrer, qu'elles soient liées au quotidien, à la formation, ou à la profession. Ce ne sont donc pas les *performances scolaires* qui se trouvent au centre de ce test international ; il s'agit plutôt de définir quels sont les *potentiels pour l'apprentissage tout au long de la vie* dont disposent les élèves à la fin de leur scolarité obligatoire (cf. OCDE, 1999 & 2003).

Les évaluations fournissent des résultats représentatifs de la « population-cible » des élèves de quinze ans dans les pays participants. Pour cette comparaison internationale (en 2003), on a par exemple choisi et testé au hasard environ 4 500 élèves en France et en Allemagne. Afin d'utiliser PISA pour examiner des questions plus pointues, chaque pays a la possibilité d'élargir l'échantillon de manière à pouvoir entreprendre des comparaisons complémentaires. Mais il est également possible de mettre en œuvre d'autres instruments d'évaluation qui pourront par exemple traiter la seconde journée de tests nationaux soumis aux élèves. En Allemagne, nous avons utilisé sans réserves ces deux possibilités comme nous le montrerons ci-dessous (cf. Prenzel *et al.*, 2004 & 2005).

Enrichissements nationaux de PISA 2003 pour l'Allemagne

Les évaluations complémentaires en Allemagne impulsées par la Conférence permanente des ministres de l'Éducation des *Länder* étaient liées à trois objectifs.

1. PISA est envisagée par l'OCDE comme une comparaison internationale, qui permet de classer et d'évaluer internationalement un système scolaire national dans le cadre d'une analyse comparative (*benchmarking*). Il faut cependant prendre en considération le fait que cette orientation internationale peut faire fi de particularités nationales et de traditions pédagogiques – et par là même éventuellement de structures et de procédés importants – pour chaque pays. C'est pourquoi, dans le cadre des évaluations complémentaires, on a évalué dans quelle mesure les compétences testées par PISA s'accordent avec les conceptions de l'éducation et avec les programmes

d'enseignement allemands. Une telle vérification a été également jugée utile pour renforcer la recevabilité des résultats par le corps enseignant et les décideurs du système scolaire.

2. L'étude PISA permet de tirer des conclusions sur la population des élèves de quinze ans. Les qualités exceptionnelles de cet échantillon ont été utilisées pour récolter au moyen de tests et de questionnaires complémentaires – spécifiquement développés pour l'Allemagne –, des informations plus différenciées sur les élèves eux-mêmes et sur la structure des compétences évaluées. Le développement de ces outils visait aussi à approfondir les connaissances scientifiques sur l'enseignement et sur l'école et enfin, à pouvoir mieux expliquer les résultats du test international.

3. Le système scolaire allemand est organisé de manière fédérale [et décentralisée] puisque la responsabilité des établissements scolaires est du ressort des *Länder*. Ceux-ci proposent donc des systèmes scolaires différents [*unterschiedliche Konstellationen von Schularten*] et ont mis en place leurs propres programmes, leurs propres systèmes de formation continue ainsi que leur propre évaluation [*Qualitätssicherungssysteme*]. Les enseignements tirés de PISA gagnent ainsi en signification pour chacun des différents *Länder*, qui en attendent des indications sur leurs problèmes spécifiques et leurs besoins concrets. On ne peut cependant pas, sur la base de l'échantillon international, tirer des conclusions quant aux résultats de chaque *Land*. C'est pourquoi l'échantillon international a été considérablement élargi en Allemagne afin de pouvoir également situer les seize *Länder* aux niveaux national et international.

CONCEPTION DES ÉVALUATIONS NATIONALES COMPLÉMENTAIRES

En Allemagne, les élargissements ont porté sur la taille et la composition de l'échantillon, les procédures de test et les instruments d'évaluation, ainsi que sur les périodes d'évaluation.

Élargissement de l'échantillon

Afin de pouvoir mener des comparaisons complémentaires au niveau des *Länder*, la taille de l'échantillon a été considérablement élargie (cf. ci-après le tableau I). On a également augmenté la proportion

des élèves issus de l'immigration, afin de pouvoir en tirer des conclusions fiables sur les sous-groupes de pays et de périodes d'immigration différents. Dans un petit échantillon d'écoles – celles qui avaient été choisies pour la comparaison internationale – on a complété l'échantillon des élèves de quinze ans choisis au niveau international, afin de pouvoir évaluer dans ces établissements deux classes complètes de 9^e. Cet élargissement est une étape nécessaire si l'on veut examiner l'impact de l'enseignement sur le développement des compétences. Pour l'échantillon international, les élèves de quinze ans sont en effet choisis au hasard et proviennent de ce fait de classes différentes (Carstensen *et al.*, 2005 ; Prenzel, Drechsel & Carstensen, 2005).

Instruments d'évaluation supplémentaires

Dans un petit échantillon d'écoles choisies parmi l'échantillon international, les élèves ont subi une deuxième journée de tests nationaux en mathématiques, en sciences, en lecture et en maniement de l'ordinateur. Ils ont également rempli un questionnaire complémentaire, sur leur position, leur ressenti et leur motivation par rapport aux mathématiques, de même que des questions sur leurs activités extra-scolaires ainsi que leur perception du foyer familial et de l'enseignement des mathématiques. Ces tests et questionnaires avaient été spécialement élaborés par le consortium « PISA-Deutschland ». Les parents des élèves participant à ces tests complémentaires ont également reçu un questionnaire portant essentiellement sur l'existence de ressources ou de soutiens liés aux mathématiques et sur leur perception de l'enseignement des mathématiques et de l'école. Enfin, les enseignants de ces écoles ont été interrogés sur leur perception de l'école et de l'équipe de direction, sur leurs charges de travail ainsi que sur leur participation au développement de l'école et leurs coopérations professionnelles. Enfin, toutes les écoles qui ont participé à l'étude PISA ont par ailleurs reçu un questionnaire complémentaire destiné aux équipes de direction dans le but de récolter des renseignements circonstanciés sur l'établissement, des informations sur l'évolution de l'École et de sa « Qualité ». On a également développé, pour un petit groupe d'écoles, un test assisté par ordinateur lors d'une troisième journée de tests, en soumettant aux élèves des problèmes complexes et dynamiques. Il a en particulier servi à évaluer des stratégies visant à résoudre des problèmes exigeant un apprentissage auto-régulé et piloté par *feed-back* dans une situation en constante évolution (Leutner *et al.*, 2004). Tous

les processus d'évaluation utilisés au niveau national ont été rassemblés dans une documentation contenant les valeurs statistiques principales, laquelle a fait l'objet d'une publication (Ramm, 2006).

Périodes de test complémentaires

Les élèves des écoles faisant partie du groupe choisi pour la comparaison internationale n'ont pas seulement été soumis à une deuxième journée de tests en 2003. Les élèves de cet échantillon, représentant pour chaque établissement deux classes complètes de 9^e, ont à nouveau été testés un an après (et donc en dixième classe), le domaine d'évaluation portant là encore sur les mathématiques. Ceci devait fournir des renseignements sur le développement des compétences des individus ou des classes au cours d'une année scolaire (Prenzel *et al.*, 2006).

Conception de l'évaluation générale

Avec tous ces éléments, l'étude PISA-2003 en Allemagne comprend en tout trois *sous-études* représentées dans l'illustration 1.

La *première* sous-étude se rapporte à la comparaison internationale d'après les règles standard ; elle est désignée « PISA-I » (« PISA-International »). La *seconde* a enrichi la comparaison internationale d'une comparaison entre les *Länder*. Dans ce but, l'échantillon a été considérablement élargi et

complété par un sur-échantillonnage (*oversampling*) d'élèves issus de l'immigration. Les élèves concernés, âgés de quinze ans ont subi lors d'une première journée d'épreuves les tests internationaux. Cette étude partielle est appelée PISA-E (« Étude élargie »). Le but était d'aboutir à des conclusions au niveau des différents *Länder* et à des conclusions sur des adolescents d'origines socio-culturelles différentes. La *troisième* sous-étude s'est concentrée sur les classes complètes de 9^e (parmi les écoles de l'échantillon international). Les élèves ont été soumis lors d'une deuxième journée d'épreuves, à des tests de compétence nationaux complémentaires. De plus, les parents des élèves testés ainsi que leurs enseignants de mathématiques ont également été interrogés.

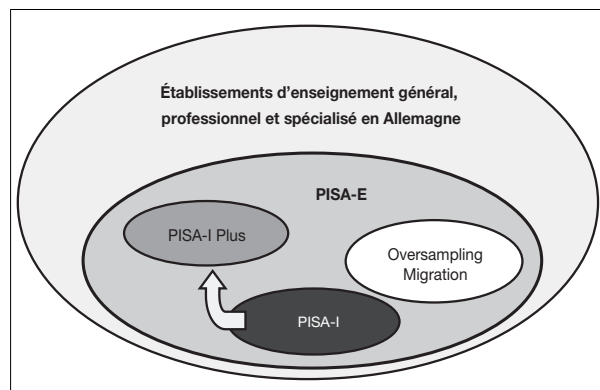
Cette étude livre des informations systématiques sur les influences du *foyer familial*, des *classes*, ainsi que sur l'équipe des enseignants de mathématiques et de *l'école* : équipe enseignante, équipe de direction. Les mêmes élèves (ou classes) ont à nouveau été testés en 2004, afin de mesurer l'évolution de leurs compétences au cours d'une année scolaire. Dans cette étude (PISA-I Plus), on a pu évaluer les influences des *individus*, des *foyers familiaux*, des *classes*, des *équipes enseignantes* et de *l'école* sur le développement des compétences de la 9^e à la 10^e classe.

RÉSULTATS

De cet ensemble de résultats obtenus grâce à ces tests élargis, nous déduisons quelques remarques montrant que ces études complémentaires menées dans le cadre de PISA sont bénéfiques – on trouvera une présentation détaillée et exhaustive de nos conclusions dans les rapports nationaux sur PISA (Prenzel *et al.*, 2004, 2005 & 2006). Nous présenterons tout d'abord l'élaboration des épreuves en mathématiques et en sciences auxquelles les jeunes ont été soumis lors de la deuxième journée de tests. Nous décrirons ensuite de quelle manière la conformité aux programmes d'enseignement des épreuves nationales et internationales a été évaluée par des experts. Nous confronterons enfin les tests nationaux et internationaux les uns aux autres et présenterons les différences de compétences mesurées entre les différents *Länder*.

L'élaboration des tests nationaux en mathématiques et en sciences poursuivait deux objectifs. Premièrement, l'évaluation internationale devait être complétée par des tests qui traitent plus spécifiquement des

Illustration 1. – Ensemble total et échantillons de PISA-I (comparaison internationale), PISA-E (comparaison des *Länder*), du « sur-échantillonnage » (*oversampling*) des élèves issus de l'immigration ainsi que l'évaluation complémentaire (PISA I-Plus)



particularités de l'enseignement des mathématiques ou des sciences en Allemagne. De ce fait, les tâches devaient être plus étroitement établies en fonction des programmes d'enseignement allemands et donc mieux correspondre aux notions mathématiques ou aux thèmes scientifiques abordés en cours. Deuxièmement, on a théoriquement tenté de définir de manière plus raffinée les sous-compétences cognitives sous-jacentes aux différents domaines. On voulait ainsi pouvoir mieux expliquer les forces et les faiblesses dans les différents domaines de compétence et en tirer des conséquences didactiques.

Au plan international, PISA a décidé de ne pas concevoir les tests en fonction des programmes, mais en fonction du concept général de *Literacy* – comprise comme une formation de base pour tous en vue de l'intégration sociale et culturelle. Étant donné les différentes « traditions curriculaires » des pays participants, il s'agit ici d'une décision judicieuse pour des comparaisons internationales justes. Cependant, le classement international du niveau de compétence atteint dans un État implique immanquablement une évaluation des systèmes de formation nationaux. Il est pour cette raison important de savoir dans quelle mesure les compétences testées par PISA dans un pays correspondent à des objectifs précis et font l'objet d'entraînements spécifiques en salle de classe. C'est par exemple le cas lorsque les exigences des épreuves correspondent aussi à celles des programmes et y sont, pour ainsi dire, « *curriculairement conformes* ». Afin d'évaluer cette conformité aux épreuves internationales et aux épreuves complémentaires nationales, on a fait appel à des experts, spécialistes en didactique des différentes matières (trois didacticiens des mathématiques et neuf des sciences). Pour chaque tâche, ils ont jugé : 1) à quel niveau de classe est enseigné le thème apparaissant pour chaque tâche selon les programmes (des pays et des différents types d'écoles) ; 2) quel est le degré de familiarité des élèves avec la formulation des tâches (sur une échelle de 1 à 4) ; 3) quelle valeur devait avoir la stimulation de la compétence nécessaire à la réalisation de la tâche à l'école (sur une échelle de 1 à 4).

Le test national en mathématiques

Caractérisation des énoncés

Le test international de PISA 2003 évalue quatre domaines en mathématiques, désignés par le terme d'*idées mathématiques transversales* : *Quantité* ; *Variations et relations* ; *Espace et formes* ; *Incerti-*

tudes. Le test international s'appuie sur un concept large de formation de base au sens d'une *mathematical literacy*, qui doit permettre l'intégration culturelle (OCDE, 2003). En revanche, le test supplémentaire développé en Allemagne et comprenant 124 tâches, s'appuie quant à lui plus largement sur les programmes des *Länder* et prend en compte trois types d'exigences qui jouent un grand rôle dans l'enseignement des mathématiques en Allemagne (Blum *et al.*, 2004). Dans les *tâches techniques*, on propose une amorce de solution, à traiter au moyen de procédés calculatoires connus. Dans les *tâches de modélisation calculatoires* (« tâche de textes »), on doit d'abord élaborer un schéma de solution puis le traduire par le calcul. Enfin, les *tâches de modélisation conceptuelle* et de *résolution de problèmes* requièrent un savoir conceptuel sur les rapports entre objets ou éléments de savoir. Les tâches du test national en mathématiques correspondent systématiquement aux domaines traités à l'école : *arithmétique*, *algèbre*, *géométrie* et *stochastique*.

La conformité curriculaire des tests nationaux et internationaux en mathématiques

En Allemagne, la plupart des élèves de quinze ans fréquente la 9^e classe. Si l'on considère la part des tâches en mathématiques traitée jusqu'à la 9^e classe (tableau I), on observe que la majeure partie des épreuves des tests nationaux et internationaux sont considérées comme conformes aux programmes (Blum *et al.*, 2004). Les tâches nationales sont un peu plus congruentes avec les exigences des programmes d'enseignement allemands : au total, 89,5 % des tâches nationales et 83,2 % des tâches internationales ont été jugées conformes aux programmes. On remarque

Tableau I. – Proportion des tâches PISA correspondantes au programme de mathématiques

Type d'école	Thèmes abordés jusqu'à la 9 ^e classe (en %)	
	Épreuves nationales	Épreuves internationales
<i>Hauptschule</i>	78,6	72,3
<i>Integrierte Gesamtschule</i>	91,4	83,9
<i>Realschule</i>	90,2	81,4
<i>Gymnasium</i>	93,1	87,4
Total	89,5	83,2

tout de suite que la conformité aux programmes des tâches destinées à la *Hauptschule* est sensiblement plus faible que pour les autres types d'école. Une raison possible en est que l'appropriation des concepts fondamentaux, par exemple du concept de variable, n'y joue qu'un rôle secondaire.

En outre, on constate, selon les estimations des experts, que les épreuves nationales sont plus familières aux élèves que les épreuves internationales. La différence correspond environ à un demi-niveau sur une échelle de quatre. Dans la mesure où test national et test international sont conformes aux programmes, cette différence ne peut s'expliquer par des écarts de contenu ou de matière enseignée. C'est donc la présentation des consignes qui semble jouer un rôle puisque dans les énoncés des tests internationaux, on s'appuie principalement sur des situations basées sur les problèmes ou la réalité. Les compétences nécessaires à la résolution des tâches mathématiques sont considérées, par les experts, comme particulièrement importantes. Cependant, pour les tâches internationales, on remarque dans tous les domaines étudiés concernant les mathématiques, un grand écart entre l'importance et la familiarité supposées avec la manière de poser les épreuves. Dans le test national, on retrouve le même type de résultats, surtout en ce qui concerne les épreuves de *géométrie* et de *stochastique*. Ces tâches sont jugées comme très importantes, selon les experts, quoiqu'elles soient encore trop peu familières aux élèves de quinze ans. Ces analyses renvoient aux traditions particulières de l'enseignement des mathématiques en Allemagne, qui pourraient donner lieu à des redéfinitions didactiques.

Au total, il faut, selon les estimations des experts, souligner que les exigences des tests internationaux et nationaux en mathématiques sont, pour une très grande part, cohérentes avec les exigences des programmes d'enseignement allemands ; eu égard à ceux-ci, le test international apparaît donc équilibré et juste.

Rapports entre les tests nationaux et internationaux en mathématiques

Les items nationaux ont été conjointement échelonnés avec les items utilisés au niveau international pour mesurer la compétence mathématique. Le test national complémentaire s'est montré un peu plus difficile que le test national (variation de 0,33 logits).

Le rapport entre le test national et le test international, situé aux alentours de $r = 0,92$, prouve de manière empirique les grandes similitudes des deux

instruments d'évaluation au niveau de leur conception. Les corrélations entre les domaines définis au niveau national et les *Idées transversales* correspondantes du test international de compétence mathématique se situe entre 0,84 (*Géométrie et Espace et formes*) et 0,89 (*Stochastique et Incertitudes*). En Allemagne, les tests internationaux se révèlent donc très significatifs et tout à fait adaptés pour traiter de problèmes pertinents au niveau national.

S'esquissent ainsi, de manière cohérente dans les résultats des test nationaux et internationaux, les forces relatives des élèves allemands dans les sous-domaines de *l'arithmétique* et de *l'algèbre*, ainsi que leurs faiblesses spécifiques pour la *stochastique*. Comme on pouvait s'y attendre, on retrouve des différences significatives de compétences entre les différents types d'école dans tous les types d'activités mathématiques – on trouvera une présentation détaillée des autres résultats dans Blum *et al.* (2004).

Le test national en sciences

Lors de la première évaluation PISA (2000), on a proposé aux élèves allemands un test en sciences qui, à côté de tests internationaux plus orientés vers le principe de la formation de base (*literacy*), prenait davantage en compte les programmes nationaux (cf. Prenzel *et al.*, 2001). Les deux tests se trouvaient dans une corrélation de l'ordre de $r = 0,84$ et attestent une correspondance entre le test international et le test national plus orienté vers les programmes des matières scientifiques. Puisqu'en 2006, les sciences seront pour la première fois le domaine majeur d'évaluation d'une étude PISA, on n'a intégré, lors des sessions PISA de 2000 et 2003, qu'un ensemble relativement restreint de tâches relevant de ce domaine. À partir de résultats des évaluations nationales de PISA 2000, on a développé, en 2003, un autre test national en sciences, théoriquement censé illustrer ces différentes facettes.

Caractérisation des énoncés

L'élargissement du test au niveau national visait à intégrer systématiquement différentes facettes de la formation scientifique de base dans l'élaboration des tâches et dans l'analyse des données (Roos *et al.*, 2004). Sept compétences partielles ont été examinées dans un test qui en comprenait au total soixante-trois :

1) travail avec des graphiques : traitement d'informations représentées sous forme d'images et transformation dans d'autres modes de représentation ;

2) modèles mentaux : utilisation de représentations de géométrie spatiale pour faire des prévisions ou expliquer des phénomènes ;

3) pensée convergente : déduction de la solution juste à partir d'une loi scientifique ;

4) pensée divergente : flux créatif d'idées et richesse d'intuition lors du traitement de problèmes scientifiques ;

5) verbalisation de contenus : donner des réponses verbales correctes, puisque la verbalisation active est indispensable pour l'appropriation et la communication de savoir ;

6) travail avec les chiffres : traitement d'informations dans le mode de représentation numérique conçu comme exigence scientifique cognitive centrale ;

7) évaluation : justification d'une décision ou d'une évaluation prévue.

En Allemagne, les sciences sont enseignées dans trois disciplines principales : biologie, physique, chimie. De ce fait, les énoncés ont été reliés à ces trois matières. Pour atteindre une représentation complète de toutes les facettes, chaque sous-compétence apparaît dans les énoncés nationaux avec la même fréquence dans les trois matières.

Conformité curriculaire des tests nationaux et internationaux en sciences

Comme pour les tests de mathématiques, la grande majorité des tâches des tests de compétence scientifique est conforme aux programmes :

Tableau II. – Proportion des tâches PISA correspondantes au programme de sciences naturelles

Type d'école	Test international	Test national
	PISA 2003	PISA 2003
<i>Hauptschule</i>	62	68
<i>Integrierte Gesamtschule</i>	73	74
<i>Realschule</i>	74	77
<i>Gymnasium</i>	75	81
Total	72	76

au total, 76 % des épreuves nationales et 72 % des épreuves internationales ont été considérées comme conformes aux programmes (Tableau II ; cf. Ross *et al.*, 2004). Comme dans le domaine de la compétence mathématique, la conformité de ces tâches pour les *Hauptschule* est légèrement inférieure que par rapport aux autres types d'école.

Le tableau III montre dans quelle mesure les experts en didactique des sciences jugent les exigences des compétences cognitives partielles en accord avec les exigences des programmes.

Toutes les compétences partielles sont estimées relativement importantes pour le développement

Tableau III. – Évaluation des épreuves nationales en sciences par rapport aux trois aspects de la conformité aux programmes d'enseignement

	Traité jusqu'à la 9 ^e classe (en %)	Familiarité	Importance pour le développement des compétences
Évaluation	60	2,35	3,20
Pensée divergente	81	3,17	3,38
Compréhension de graphiques	74	2,50	3,02
Pensée convergente	77	2,59	3,11
Modèles mentaux	79	2,60	3,15
Verbalisation de contenus	84	3,05	3,31
Rapport aux chiffres	79	2,57	2,66

Les chiffres des colonnes 3 et 4 représentent une moyenne des jugements d'experts sur une échelle de 1 à 4, par ex. : 4 « très grande familiarité », 1 : « aucune familiarité »

des compétences [générales]. Trois compétences partielles sont estimées comme les plus importantes : *évaluation*, *pensée divergente* et *verbalisation de contenus*, ces deux dernières étant par ailleurs jugées être les plus familières aux élèves. Ainsi demande-t-on aux adolescents allemands de fournir des idées et de verbaliser des contenus tout au long de leur scolarité. Ceci pourrait être une conséquence de la forme prédominante de l'enseignement des sciences en Allemagne dont la progression est basée sur le questionnement. (Seidel & Prenzel, 2004 & 2006).

Si l'on considère en revanche la compétence d'évaluation, il apparaît clairement que la proximité avec les programmes tout comme la familiarité des élèves avec cette compétence partielle sont jugées plus faibles que pour les autres compétences partielles. Cet écart indique que les processus d'évaluation ont une importance très faible dans les programmes d'enseignement nationaux.

Que ce soit pour le test de compétence mathématique ou pour le test de compétence scientifique, la part des épreuves dont le contenu a été traité à l'école se situe au-dessus de 70 %. Les tests sont donc composés pour leur plus grande part d'épreuves qui possèdent des similarités suffisantes avec les programmes d'enseignement pour en tirer des conclusions au niveau didactique. En outre, les experts interrogés reconnaissent ces tâches comme importantes pour le développement des compétences. En résumé, les enseignements tirés des tests nationaux comme internationaux peuvent-ils être considérés comme conformes aux programmes d'enseignement et par suite, congruents au système scolaire allemand.

Rapports entre les tests nationaux et internationaux en sciences

Au niveau des résultats, les tests nationaux portant sur les compétences partielles, *travail avec des graphiques*, *pensée convergente*, *modèles mentaux* et *travail avec des chiffres*, se sont avérés relativement difficiles pour les élèves. *Évaluation*, *pensée divergente* et *verbalisation de contenus* ont été en revanche beaucoup plus faciles. Dans ces exigences cognitives plus faciles, le mode de représentation verbal joue un rôle central alors que dans les exigences cognitives plus ardues dominant au contraire des modes de représentation numériques, métaphoriques ou plutôt abstraits. Ces compétences partielles sont corrélées : des élèves ayant des points forts dans le domaine de la pensée convergente s'en sortent également bien avec des formules et des informations représentées graphiquement.

Il est intéressant de constater qu'on n'a pas pu trouver d'indications concernant des compétences partielles spécifiques à une matière scolaire. Qu'une épreuve scientifique soit résolue dépend donc avant tout du type d'exigence cognitive et non du domaine disciplinaire dont est issu son contenu.

Pareillement, le « test de facette » développé de manière systématique au niveau national est en corrélation avec le test international de PISA 2003. Les corrélations entre le test international et la compétence cognitive partielle se situent dans une fourchette comprise entre $r = 0,63$ (*modèles mentaux*) et $r = 0,81$ (*rapport aux chiffres*). Le test complémentaire national aide en particulier à mieux comprendre les résultats du test international et à identifier les forces et les faiblesses relatives des élèves en Allemagne – une présentation détaillée des autres résultats peut être consultée dans Rost *et al.* (2004).

Compétence mathématique dans la comparaison nationale et internationale

Comme, par exemple, l'Australie, le Canada et la Suisse, l'Allemagne a fait usage d'une option proposée par PISA en augmentant l'échantillon nécessaire d'élèves de quinze ans pour le niveau international, afin de pouvoir caractériser le niveau de compétence des jeunes, y compris dans des unités régionales de petite taille (OCDE, 2004a). On a, dans ce but, comparé les *Länder* entre eux au niveau national (Prenzel *et al.*, 2005). Afin de pouvoir tirer des conclusions pour chaque *Land* et pour les types d'écoles qui y sont le plus fortement représentés, on a dû tester pour PISA 2003 dans les seize *Länder* environ 45 000 élèves, soit environ dix fois plus que pour la comparaison internationale.

Types d'écoles et taux de participation dans les Länder allemands

Le système éducatif des *Länder* allemands est basé sur une structure de base commune (Conférence permanente des ministres de l'enseignement des *Länder*, 2005). Il existe toutefois, en particulier dans les niveaux correspondant en France au collège (*Sekundarstufe I*) et au lycée (*Sekundarstufe II*) d'où étaient issus les élèves constituant l'échantillon, toute une série de spécificités propres à chaque *Land*.

Au niveau de l'enseignement secondaire les élèves sont repartis sur trois filières de niveau différent proposées en principe dans trois types d'établissements différents : le collège de base (*Hauptschule*) qui mène à un certificat d'études de base ; le collège

moyen (*Realschule*) qui mène à un certificat moyen ; le collège-lycée (*Gymnasium*) mène au certificat donnant accès à l'université (*Abitur*, équivalent fonctionnel du baccalauréat français). Un quatrième type d'établissement (*Gesamtschule*) correspond au collège unique. Dans quelques *Länder*, il existe aussi des établissements regroupant plusieurs filières sous un même toit sans toutefois les mélanger, lesquels possèdent des dénominations variables selon chaque *Land*. Sauf pour les *Gymnasium*, la part des élèves de quinze ans dans les différentes écoles varie de façon parfois considérable entre les différents *Länder*. Le tableau IV montre la répartition des élèves sur les différents types d'établissements par *Länder* (Prenzel *et al.*, 2005).

Comme on peut le voir dans le tableau, l'affirmation selon laquelle le système scolaire allemand serait divisé en trois grandes filières [*Hauptschule*, *Real-*

schule et *Gymnasium*] n'est pas du tout valable pour tous les *Länder*. Il existe de nombreuses variantes comprenant de deux à cinq types d'école différents, chacune étant fréquentée par au moins 5 % des élèves de quinze ans. Le tableau montre également, que le *Gymnasium* traditionnel, représenté dans tous les *Länder*, est fréquenté par un pourcentage différent d'élèves d'une même classe d'âge. Ainsi le pourcentage des élèves de quinze ans varie dans les *Gymnasium* de 25,2 % (Schleswig-Holstein) à 34,5 % (Berlin). Dans les *Hauptschule*, présentes dans onze des seize *Länder*, le pourcentage varie de 0,4 % (Mecklembourg-Poméranie occidentale) à 32,2 % (Bavière). Cette rapide caractérisation illustre les spécificités propres à chaque *Land* et justifie le fait de considérer chacun comme un objet unique. Nous allons, dans la section suivante, présenter les résultats des *Länder* en comparaison avec les autres pays à partir de l'exemple de la compétence mathématique.

Tableau IV. – Répartition des élèves de quinze ans selon le type d'école et les *Länder* (en %)

	Hauptschule	École comportant plusieurs cursus	Integrierte Gesamtschule	Realschule	Gymnasium	Berufsschule	Sonder-/Förderschule
Bade-Württemberg	27,9	–	2,0	30,3	27,8	9,0	3,0
Bavière	32,2	–	0,6	27,2	26,3	11,0	2,6
Berlin	11,2	–	28,5	21,6	34,5	1,0	3,2
Brandenbourg	–	–	50,1	15,7	30,8	–	3,4
Brême	21,7	–	16,5	26,7	30,6	1,1	3,4
Hambourg	10,6	5,1	26,5	14,9	33,4	5,9	3,6
Hesse	15,6	–	17,1	27,0	31,7	5,2	3,3
Mecklembourg-Poméranie occidentale	0,4	53,2	4,0	9,6	25,8	1,7	5,4
Basse-Saxe	28,3	0,4	4,1	33,5	26,6	3,6	3,5
Rhénanie du Nord-Westphalie	26,6	–	16,2	24,7	28,8	–	3,7
Rhénanie-Palatinat	22,9	12,8	4,7	22,2	25,8	8,5	2,9
Sarre	–	45,7	13,5	1,6	25,7	10,8	2,7
Saxe	–	61,3	–	–	32,0	2,6	4,0
Saxe-Anhalt	–	60,9	1,1	–	30,6	2,3	5,2
Schleswig-Holstein	29,3	–	7,3	31,4	25,2	3,4	3,5
Thuringe	–	57,6	1,4	–	30,5	5,6	4,9

Classement international des Länder

Le tableau V montre les valeurs moyennes et les variations standard de performances pour le test international en mathématiques, que les états membres de l'OCDE ont atteints lors de PISA 2003. Ainsi la valeur moyenne de la France (511 points) se situe-t-elle sensiblement au-dessus de la moyenne de l'OCDE (500 points) alors que la valeur moyenne pour l'Allemagne (502 points) ne se différencie pas significativement de la moyenne OCDE. Dans ce classement figurent les scores obtenus par les *Länder* (Neubrand *et al.*, 2005). Dans quatre *Länder* (Bavière, Bade-Württemberg, Saxe et Thuringe) les valeurs moyennes se situent sensiblement au-dessus de la moyenne OCDE. Dans quatre autres en revanche, les villes-États de Berlin, Hambourg et Brême ainsi que la Rhénanie du Nord-Westphalie, les résultats sont sensiblement en deçà. Il faut remarquer la différence de valeur moyenne entre le plus performant (la Bavière) et le plus faible (Brême) : elle correspond à 62 points de compétence et donc à un écart de performance de plus d'une année scolaire.

On retrouve ces différences importantes de compétence entre les *Länder*, si on analyse les performances des *Gymnasium* : alors que pour ceux de Bavière, on obtient des valeurs moyennes de 613 points, elles ne s'élèvent qu'à 562 points pour ceux de Brême. Si l'on compare un même type d'école qui rassemble dans chaque *Land* les meilleurs élèves et qui, par-delà les *Länder*, est assez similaire pour ce qui concerne l'orientation programmatique, les performances moyennes accusent également une différence de plus de 50 points. Ces différences substantielles demeurent également, si l'on contrôle les différences de répartition dans les *Gymnasium* des différents *Länder* (Neubrand *et al.*, 2005). La dispersion des compétences dans les États et les *Länder* est également intéressante. Il apparaît que, dans la comparaison internationale, de très bonnes moyennes peuvent être atteintes même si la dispersion des performances est relativement faible. Pour l'Allemagne en général, comme pour beaucoup de *Länder*, on observe des dispersions de compétences assez élevées, c'est-à-dire situées au-dessus de la moyenne OCDE. Elles sont particulièrement marquées dans les trois villes-États (Berlin, Brême et Hambourg), en Hesse et dans le Schleswig-Holstein. D'autres analyses des résultats des *Länder* selon le type d'école (Prenzel, Zimmer *et al.*, 2005) montrent par exemple que dans certains *Länder*, les élèves de *Realschule* atteignent le même niveau que les élèves de *Gymnasium* d'autres *Länder*. La proportion d'élèves se

situant à des niveaux de compétence particulièrement bas ou, au contraire, particulièrement élevés, est riche d'enseignements pour l'évaluation des systèmes éducatifs. On retrouve également des différences conséquentes entre les *Länder*, lorsqu'on observe la proportion des élèves particulièrement faibles.

Les résultats des autres domaines de compétence sont identiques à ceux obtenus en mathématiques. Au final, lorsqu'on compare les *Länder*, on remarque que les situations problématiques en Allemagne sont très différentes d'un *Land* à l'autre : alors que certains *Länder* obtiennent des moyennes qui se rapprochent de celles du groupe de tête, d'autres restent nettement en dessous de la moyenne OCDE. On peut également observer à partir de ces comparaisons, dans quel domaine se situent les performances des différents types d'école. Les résultats mettent ainsi l'accent sur des défis variés concernant le développement futur de l'enseignement et de l'école en Allemagne.

Comparaison avec la compétence de résolution de problèmes

Un résultat de la comparaison internationale de PISA 2003 qui a plus particulièrement suscité de l'intérêt en Allemagne concerne les relations entre la compétence mathématique et la compétence à résoudre des problèmes transversaux. Les tâches proposées lors du test PISA de résolution de problèmes supposent notamment une pensée analytique et représentent des exigences qui jouent également un rôle important pour les tâches mathématiques. Si l'on considère les tâches du test de résolution de problèmes, une série d'entre elles apparaissent en effet comme des tâches mathématiques « bien déguisées » (OCDE, 2004b). Les données empiriques montrent également que ces deux domaines de compétence proposent des exigences cognitives très proches. La corrélation latente entre la compétence de résolution de problèmes et la compétence mathématique atteint (au niveau individuel) $r = 0,89$ et se trouve par là même aussi élevée que la corrélation des sous-échelles mathématiques entre elles (OCDE, 2004b).

Du fait que les valeurs de compétence en mathématique et en résolution de problèmes ont au plan international les mêmes normes – la même valeur moyenne (500) et la même variation standard (100) –, on peut s'attendre à ce que les compétences dans ces deux domaines soient également situées au même niveau. Les résultats de certains États déjouent cependant de cette attente : alors que la compétence moyenne de résolution de problèmes en Allemagne

Tableau V. – Valeurs moyennes, dispersions et valeurs en pourcentage pour la compétence en mathématique (échelle internationale) : les Länder face à la comparaison internationale

États de l'OCDE et Länder allemands	valeur moyenne	(S.E.)	SD	(S.E.)	pourcentages					
					5 %	10 %	25 %	75 %	90 %	95 %
Finlande	544	(1,9)	84	(1,1)	406	438	488	603	652	680
Corée	542	(3,2)	92	(2,1)	388	423	479	606	659	690
Pays-Bas	538	(3,1)	93	(2,3)	385	415	471	608	657	683
Japon	534	(4,0)	101	(2,8)	361	402	467	605	660	690
Bavière	533	(3,7)	107	(2,4)	357	401	470	601	654	684
Canada	532	(1,8)	87	(1,0)	386	419	474	593	644	673
Belgique	529	(2,3)	110	(1,8)	334	381	456	611	664	693
Suisse	527	(3,4)	98	(2,0)	359	396	461	595	652	684
Australie	524	(2,1)	95	(1,5)	364	399	460	592	645	676
Saxe	523	(2,4)	101	(1,5)	351	396	462	593	640	669
Nouvelle-Zélande	523	(2,3)	98	(1,2)	358	394	455	593	650	682
République Tchèque	516	(3,5)	96	(1,9)	358	392	449	584	641	672
Islande	515	(1,4)	90	(1,2)	362	396	454	578	629	658
Danemark	514	(2,7)	91	(1,4)	361	396	453	578	632	662
Bade-Würtemberg	512	(2,6)	101	(1,7)	339	380	445	583	636	664
France	511	(2,5)	92	(1,8)	352	389	449	575	628	656
Thuringe	510	(2,8)	97	(1,9)	341	384	450	575	631	660
Suède	509	(2,6)	95	(1,8)	353	387	446	576	630	662
Autriche	506	(3,3)	93	(1,7)	353	384	439	571	626	658
Allemagne	503	(3,3)	103	(1,8)	324	363	432	578	632	662
Irlande	503	(2,4)	85	(1,3)	360	393	445	562	614	641
Saxe Anhalt	502	(3,1)	97	(2,3)	330	372	439	571	626	653
Moyenne OCDE	500	(0,6)	100	(0,4)	332	369	432	571	628	660
Sarre	498	(2,3)	91	(1,7)	339	377	439	562	613	642
Slovaquie	498	(3,3)	93	(2,3)	342	379	436	565	619	648
Schleswig-Holstein	497	(3,1)	102	(2,4)	322	361	425	572	626	655
Hesse	497	(3,7)	105	(2,0)	318	355	423	573	631	662
Norvège	495	(2,4)	92	(1,2)	343	376	433	560	614	645
Basse-Saxe	494	(2,7)	97	(2,1)	326	362	428	565	617	645
Mecklemburg-Poméranie occidentale	493	(2,6)	98	(2,1)	325	366	430	561	619	651
Rhénanie-Palatinat	493	(2,6)	97	(1,9)	327	365	426	562	619	649
Brandebourg	492	(3,1)	92	(2,1)	332	372	432	556	608	637
Luxembourg*	493	(1,0)	92	(1,0)	338	373	430	557	611	641
Pologne	490	(2,5)	90	(1,3)	343	376	428	553	607	640
Hongrie	490	(2,8)	94	(2,0)	335	370	426	556	611	644
Berlin	488	(2,7)	103	(1,8)	315	350	416	564	617	648
Rhénanie du Nord-Westphalie	486	(2,5)	100	(1,7)	312	350	415	559	613	642
Espagne	485	(2,4)	88	(1,3)	335	369	426	546	597	626
USA	483	(2,9)	95	(1,3)	323	356	418	550	607	638
Hambourg	481	(2,5)	102	(1,9)	313	345	407	556	613	647
Brême	471	(2,5)	101	(1,6)	305	339	399	544	605	636
Italie	466	(3,1)	96	(1,9)	307	342	400	530	589	623
Portugal	466	(3,4)	88	(1,7)	321	352	406	526	580	610
Grèce	445	(3,9)	94	(1,8)	288	324	382	508	566	598
Turquie	423	(6,7)	105	(5,3)	270	300	351	485	560	614
Mexique	385	(3,6)	85	(1,9)	247	276	327	444	497	527

■ au-dessus

■ au niveau

■ en dessous

de la moyenne OCDE

*Ce pays a été classé en dessous de la moyenne OCDE en raison d'un problème d'échantillon.

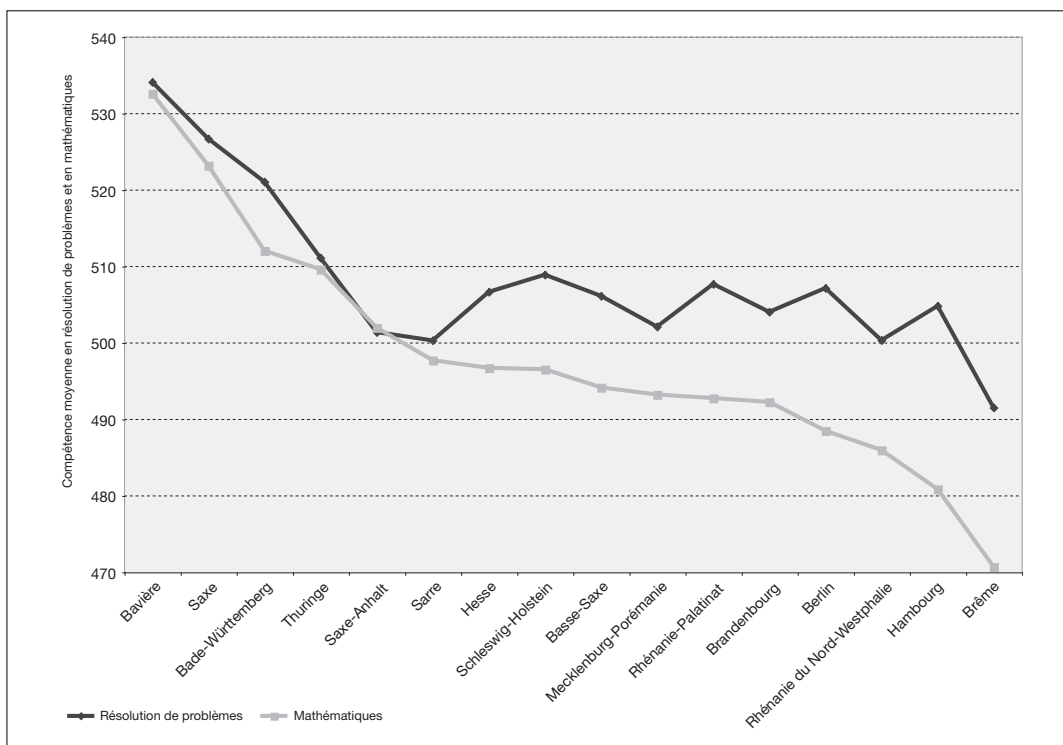
se situe au-dessus de la moyenne OCDE (autour de 513 points), on obtient seulement des valeurs médiocres pour la compétence mathématique (503 points), qui se situe dans la moyenne OCDE (Leutner *et al.*, 2004). La différence des moyennes est en Allemagne, avec ses dix points, remarquablement élevée par rapport à d'autres États : il n'y a que le Japon (treize points) et la Hongrie (onze points) qui présentent des écarts aussi élevés. La France obtient également de meilleurs résultats en résolution de problèmes (519 points) qu'en mathématiques (511 points). À l'inverse, les élèves des Pays-Bas par exemple obtiennent 520 points en résolution de problèmes mais 538 points en mathématiques.

Compte tenu des résultats allemands, on peut supposer que l'écart entre les performances en résolution de problèmes et en mathématiques dénote des potentiels cognitifs qui ne sont pas encore épuisés. Si l'on se fonde sur la propension à la pensée analytique que mesure PISA, les performances en mathématiques pourraient être meilleures.

Comme le montrent D. Leutner *et al.* (2004) pour l'échantillon allemand de la comparaison internationale, cette différence est surtout marquée pour les niveaux de performance inférieurs, au détriment de la compétence mathématique. Ces analyses indiquent que les élèves se situant dans les niveaux de performance inférieurs pourraient être bien meilleurs si l'on en juge par leurs capacités à résoudre des problèmes et à penser en mathématiques. Ils font preuve de prédispositions cognitives encore insuffisamment exploitées pour le moment.

Il faut toutefois relever les enseignements de la comparaison des *Länder* sur cette question (Leutner *et al.*, 2005). Il n'existe visiblement pas, dans certains *Länder*, d'écarts importants entre les résultats en mathématiques et en résolution de problèmes. Et il est intéressant de noter que cela se produit dans les *Länder* qui ont globalement de bons résultats en mathématiques. Comme le montre l'illustration 2, c'est surtout dans les *Länder* qui ont un niveau relativement faible en mathématiques que la différence

Illustration 2. – Différence entre la compétence de résolution de problèmes et la compétence mathématique moyenne (les *Länder* sont classés d'après leurs résultats en compétence mathématique)



entre les résultats en mathématiques et en résolution de problèmes est élevée.

Si l'on observe en outre les différences moyennes en mathématiques et en résolution de problèmes dans les niveaux de performance inférieurs, on constate qu'elles sont relativement grandes dans les trois villes-États (Berlin, Brême, Hambourg) ainsi qu'en Rhénanie du Nord-Westphalie, Schleswig-Holstein et Rhénanie-Palatinat. En Sarre, Saxe-Anhalt, Thuringe et Bavière, les potentiels cognitifs paraissent être en revanche bien exploités sur l'ensemble des performances.

Globalement, l'examen de la compétence de résolution de problèmes participe à l'interprétation des enseignements de PISA. Il apparaît pour l'Allemagne – mais on retrouve la même tendance pour la France – que les élèves disposent en moyenne de meilleures capacités cognitives que ne le montrent les tests par matière : ils ne sont aucunement « idiots » pour utiliser une formule exagérée. Ce sont précisément les élèves situés dans des niveaux de compétences inférieurs qui pourraient atteindre un meilleur niveau, par exemple en mathématiques ou en sciences. En outre, la comparaison en Allemagne montre qu'il est beaucoup plus facile, dans certaines régions que dans d'autres, de transférer [*umsetzen*] les capacités cognitives en compétences mathématiques. Et cette constatation peut tout à fait être envisagée comme un défi pédagogique.

Origine sociale et compétence mathématique

Un objectif majeur des institutions d'enseignement est que tous aient les mêmes chances de développer les compétences nécessaires à leur intégration sociale, indépendamment des conditions sociales dans lesquelles grandissent les enfants ou les adolescents. Un rapport étroit entre origine sociale et compétence scolaire peut renvoyer à une injustice sociale et montrer en même temps qu'on ne parvient à développer tous les potentiels que de manière insuffisante. Le degré de dépendance entre l'arrière-plan socio-économique et le niveau de compétences des élèves est défini dans l'étude PISA au moyen d'une régression linéaire des valeurs de compétences sur l'index de statut économique, social et culturel (ESCS ; cf. OCDE, 2004a). La variance ainsi obtenue sert d'unité de mesure pour définir le rapport entre niveau de compétence et origine sociale.

Comme le montre la comparaison internationale, l'origine sociale joue, dans tous les États, un grand rôle dans le développement des compétences. Mais

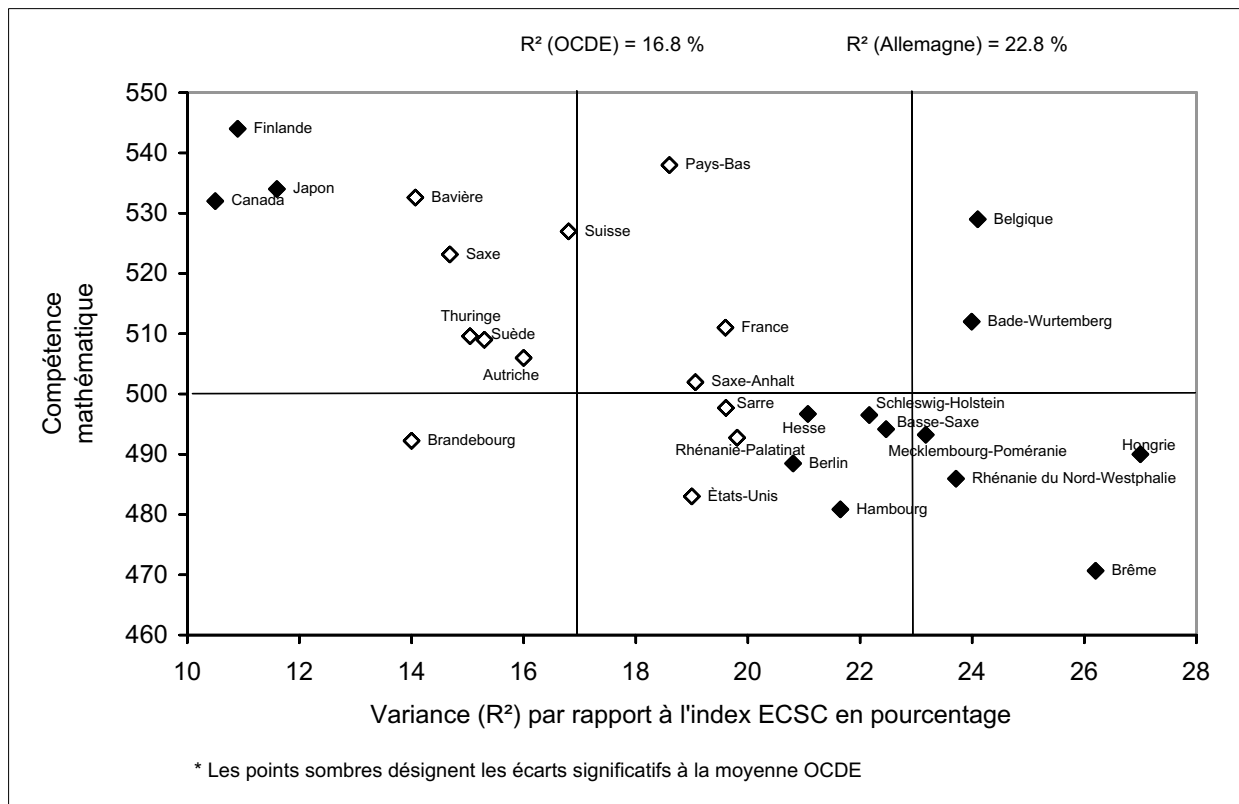
le degré de dépendance varie : des États comme la Finlande ou le Canada montrent qu'un niveau de compétence élevé peut être atteint avec un *rapport* relativement faible entre origine et compétence. Alors que dans ces États environ 10 % de la variance de performance peut s'expliquer par l'origine sociale, la dépendance est plus étroite en Allemagne (22,8 % de la variance) ou en France (16,6 % de la variance). Pour l'Allemagne, la valeur de référence du rapport se situe largement au-dessus de la moyenne OCDE (Ehmke *et al.*, 2004). Et comme le montrent des analyses plus approfondies, ce rapport est beaucoup moins marqué dans certains *Länder* que dans d'autres.

L'illustration 3 montre, d'une part, l'importance de la dépendance entre origine sociale et compétence mathématique (à l'aune de la variance), et, d'autre part, le niveau de performance moyen atteint en mathématiques d'un échantillon d'États de l'OCDE et de *Länder* allemands. Les valeurs moyennes de l'OCDE et de l'Allemagne pour le rapport entre origine sociale et compétence (respectivement 16,8 % et 22,8 % de variance) ainsi que pour le niveau de performance moyen des pays de l'OCDE (valeur 500) délimitent six champs différents. Le premier champ (en haut à gauche) correspond à une constellation favorable : pour un *rapport* relativement faible entre origine et compétence, on atteint un niveau de performance relativement élevé. Appartiennent par exemple à ce premier champ, la Finlande, le Japon et le Canada. Mais quelques *Länder* (la Bavière, la Saxe et la Thuringe) montrent une combinaison souhaitable d'un niveau de performance relativement élevé et d'un *rapport* relativement faible entre origine et compétence. Pour tous ces pays ou ces *Länder*, la proportion d'élèves faibles est également relativement minime.

En revanche, la combinaison entre faible niveau de performance et compétence faible est moins favorable. Dans ce champ (au centre et en bas à droite), on trouve plusieurs *Länder* allemands. Quatre d'entre eux (Berlin, Rhénanie du Nord-Westphalie, Hambourg et Brême) se situent par leurs performances largement en deçà de la moyenne OCDE. De plus, dans ces *Länder*, le rapport entre origine sociale et compétences mathématiques est beaucoup plus marqué que la moyenne OCDE.

Il faut donc souligner que, *in fine*, par-delà les États et les *Länder*, le niveau de compétence moyen n'augmente pas systématiquement lorsque la variance augmente. Au contraire, dans toute une série de pays, on atteint un niveau de compétence en mathématiques, supérieur à la moyenne pour un *rapport*

Illustration 3. – Compétence mathématique et variance (R^2) par rapport à l'index ECSC



relativement faible entre origine et compétences. Cela signifie également qu'un système de formation qui donne à chacun, indépendamment de son origine sociale, de bonnes chances de développer ses compétences et qui les motive, peut atteindre un niveau de compétence très élevé (dans la comparaison internationale). Les efforts pour élaborer un système de formation socialement juste sont donc tout à fait utiles pour le développement d'un niveau de compétence élevé et concurrentiel. Cet enseignement représente par là même un élément de comparaison (*benchmark*) pour plusieurs États et *Länder*.

PISA ET RECHERCHE EMPIRIQUE SUR LA FORMATION

Les études comparatives internationales comme PISA prennent les systèmes nationaux d'éducation au sérieux : elles évaluent et comparent les résultats de

la formation, dont on peut admettre qu'ils sont importants pour l'avenir des individus comme de la société. Un programme d'indicateurs de l'OCDE, conformément à la vocation de l'organisation, considère tout d'abord les résultats dans une perspective économique. Mais les compétences examinées par PISA sont également des bases du développement social et de la discussion politique. Une étude comme PISA n'est évidemment pas en mesure d'examiner de manière exhaustive tous les aspects importants de l'éducation. PISA a choisi une voie pragmatique qui est de partir d'indicateurs fiables et parlants pour analyser des compétences-clés, en incluant dans une certaine mesure des présupposés pour la formation continue. Ce n'est que de cette manière qu'il a été possible de trouver des points de comparaison acceptables par tous malgré la grande diversité internationale des programmes d'enseignement [*Lehrplänen*] et des organisations scolaires.

Dans une perspective scientifique, PISA utilise le savoir-faire développé ces dernières années dans

la recherche didactique et pédagogique ainsi qu'en psychométrie. D'un point de vue scientifique, les limites de PISA sont claires : il s'agit d'une étude en coupe, essentiellement descriptive et comparative, qui ne peut fournir de manière empirique aucune explication pour les différences ou les faits constatés. Par ailleurs, de nombreuses caractéristiques qui, d'un point de vue théorique sont des facteurs conditionnant, ne sont recueillies que par questionnaires. Au vu de l'objectif, qui est de mettre à disposition de connaissances visant à la conduite politique des systèmes d'éducation, d'autres niveaux d'action – par exemple ceux des personnels enseignants, des équipes de direction, des formateurs ou même des parents – sont restés jusqu'ici à l'arrière-plan. Si l'on veut acquérir des connaissances au niveau des personnels enseignants ou de leur formation, il faut alors choisir, lors de la définition des échantillons, d'évaluer des classes complètes.

Il est possible, au niveau national, d'apporter un correctif à toute une série de ces réserves auxquelles PISA est confrontée pour diverses raisons au niveau international. Nous avons décrit dans cet article plusieurs voies par lesquelles il est possible, au niveau national, en élargissant l'échantillon et en pratiquant des évaluations supplémentaires, d'acquérir des connaissances plus approfondies d'un point de vue scientifique.

Un premier aspect était la question de savoir dans quelle mesure la conception des tests de PISA s'accorde aux traditions et aux spécificités nationales. Nous avons pu montrer que les épreuves proposées par PISA sont porteuses d'exigences globalement en accord avec les exigences fondamentales des programmes d'enseignement allemands. Nous avons pu également montrer que les tests développés en Allemagne avec des types d'exercices plus familiers auraient finalement conduit aux mêmes résultats que les tests internationaux. À travers le développement de tests complémentaires, nous avons eu, de surcroît, la possibilité d'appréhender, de manière différenciée, certains éléments des compétences mathématiques et scientifiques et d'en expliquer la valeur. De cette façon, PISA a pu être utilisé pour acquérir des connaissances sur des modèles de compétence et des procédés de définition de compétences. Un troisième aspect de l'élargissement national a concerné l'examen des disparités régionales du fait que la responsabilité du système scolaire allemand incombe en fait aux *Länder*. Dans ce domaine, le tableau est très contrasté : il y a non seulement des différences dramatiques entre le niveau moyen des

Länder, mais également au niveau du rapport entre origine sociale et compétence. Selon le *Land* dans lequel il est scolarisé, un élève peut se retrouver avec un an de retard ou un an d'avance par rapport aux élèves d'autres *Länder*. Ces résultats remettent en cause la comparaison des performances scolaires à l'intérieur d'un État (y compris pour le *Gymnasium*). Visiblement, au sein d'un même État, on parvient dans certains *Länder* à beaucoup mieux transformer [*umsetzen*] les potentiels cognitifs en compétences dans une matière donnée. Pour interpréter ces résultats, il ne faut donc pas seulement considérer, pour chaque *Land*, les spécificités de l'organisation scolaire ou des traditions d'enseignement ni exclure les effets de la valeur de l'enseignement et de l'école, variables selon les *Länder*.

Même les comparaisons entre les *Länder*, rendues possibles par l'élargissement de l'échantillon, n'offrent, dans un premier temps, que des connaissances descriptives, car elles ne reposent que sur une seule période d'évaluation. C'est pour cette raison que, pour PISA 2003, nous avons rajouté une période d'évaluation. Dans l'échantillon d'écoles choisies pour la comparaison internationale, nous avons sélectionné, au hasard, deux classes complètes supplémentaires. Ces classes ont à nouveau été testées en mathématiques et en sciences un an après l'évaluation PISA. De cette manière, nous avons pu contrôler ce que les élèves ont appris en une année scolaire et nous avons surtout pu nous interroger sur les facteurs importants pour l'acquisition de compétences au niveau de la classe, de l'école, ou du foyer familial – on trouvera un résumé de ces résultats d'analyses que nous ne pouvons ici développer dans Prenzel *et al.*, 2006. Grâce à ces élargissements, l'étude comparative internationale est devenue un projet de recherche important, qui peut contribuer à mieux comprendre les conditions de l'enseignement et de l'environnement scolaire. Comme PISA met à disposition un excellent échantillon, une batterie d'épreuves de test ainsi que des instruments d'évaluation, la charge de travail supplémentaire est limitée lorsqu'on veut mettre en place un travail de recherche sur les différences constatées par PISA.

Manfred Prenzel

prenzelm@ipn.uni-kiel.de

Institut Leibniz des sciences de l'éducation, Kiel
(Allemagne)

Karin Zimmer

zimmer@ipn.uni-kiel.de

Institut Leibniz des sciences de l'éducation, Kiel
(Allemagne)

BIBLIOGRAPHIE

- BLUM W. ; NEUBRAND M. ; EHMKE T. ; SENKBEIL M. ; JORDAN A. ; ULFIG F. & CARSTENSEN C. H. (2004). « Mathematische Kompetenz ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland : Ergebnisse des zweiten internationalen Vergleichs*. Münster : Waxmann, p. 47-92.
- CARSTENSEN C. H. ; KNOLL St. ; SIEGLE T. ; ROST J. & PRENZEL M. (2005). « Technische Grundlagen des Ländervergleichs ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003 Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann, p. 385-401.
- EHMKE T. ; HOHENSEE F. ; HEIDEMEIER H. & PRENZEL M. (2004). « Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland : Ergebnisse des zweiten internationalen Vergleichs*. Münster : Waxmann, p. 225-254.
- EHMKE T. ; SIEGLE T. & HOHENSEE F. (2005). « Soziale Herkunft im Ländervergleich ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann, p. 235-268.
- LEUTNER D. ; KLIEME E. ; MEYER K. & WIRTH J. (2004). « Problemlösen ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland : Ergebnisse des zweiten internationalen Vergleichs*. Münster : Waxmann, p. 147-175
- LEUTNER D. ; KLIEME E. ; MEYER K. & WIRTH J. (2005). « Die Problemlösekompetenz in den Ländern der Bundesrepublik Deutschland ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann, p. 125-146.
- NEUBRAND M. ; BLUM W. ; EHMKE T. ; JORDAN A. ; SENKBEIL M. ; ULFIG F. & CARSTENSEN C. H. (2005). « Mathematische Kompetenz im Ländervergleich ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann, p. 51-84.
- OCDE (1999). *Measuring student knowledge and skills : A new framework for assessment*. Paris : OCDE.
- OCDE (2003). *The PISA 2003 Assessment Framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris : OCDE.
- OCDE (2004a). *Apprendre aujourd'hui, réussir demain : premiers résultats de PISA 2003*. Paris : OCDE.
- OCDE (2004b). *Résoudre des problèmes, un atout pour réussir : premières évaluations des compétences transdisciplinaires issues de PISA 2003*. Paris : OCDE.
- PRENZEL M. ; BAUMERT J. ; BLUM W. ; LEHMANN R. ; LEUTNER D. ; NEUBRAND M. ; PEKRUN R. ; ROLFF H.-G. ; ROST J. & SCHIEFELE U. (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland : Ergebnisse des zweiten internationalen Vergleichs*. Münster : Waxmann.
- PRENZEL M. ; BAUMERT J. ; BLUM W. ; LEHMANN R. ; LEUTNER D. ; NEUBRAND M. ; PEKRUN R. ; ROLFF H.-G. ; ROST J. & SCHIEFELE U. (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann.
- PRENZEL M. ; BAUMERT J. ; BLUM W. ; LEHMANN R. ; LEUTNER D. ; NEUBRAND M. ; PEKRUN R. ; ROLFF H.-G. ; ROST J. & SCHIEFELE U. (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster : Waxmann.
- PRENZEL M. ; DRECHSEL B. & CARSTENSEN C. H. (2005). « Einführung in den Ländervergleich PISA 2003 ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003 Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann, p. 13-50.
- PRENZEL M. ; ROST J. ; SENKBEIL M. ; HÄUSSLER P. & KLOPP A. (2001). « Naturwissenschaftliche Grundbildung : Testkonzeption und Ergebnisse ». In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen : Leske + Budrich, p. 191-248.
- PRENZEL M. ; ZIMMER K. ; DRECHSEL B. ; HEIDEMEIER H. & DRAXLER C. (2005). « Der Blick in die Länder ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003 Der zweite Vergleich der Länder in Deutschland : Was wissen und können Jugendliche ?* Münster : Waxmann, p. 169-233.
- RAMM G. ; PRENZEL M. ; BAUMERT J. ; BLUM W. ; LEHMANN R. ; LEUTNER D. ; NEUBRAND M. ; PEKRUN R. ; ROLFF H.-G. ; ROST J. & SCHIEFELE U. (Hrsg.) (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster : Waxmann.
- ROST J. ; WALTER O. ; CARSTENSEN C. H. ; SENKBEIL M. & PRENZEL M. (2004). « Naturwissenschaftliche Kompetenz ». In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland : Ergebnisse des zweiten internationalen Vergleichs*. Münster : Waxmann, p. 111-146.
- SEIDEL T. & PRENZEL M. (2004). « Muster unterrichtlicher Aktivitäten im Physikunterricht ». In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule : Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung*. Münster : Waxmann, p. 177-194.
- SEIDEL T. & PRENZEL M. (2006). « Stability of teaching patterns in physics instruction : Findings from a video study ». *Learning and Instruction*, vol. 16, p. 228-240.
- STÄNDIGE KONFERENZ DER KULTUSMINISTER DER LÄNDER DER BUNDESREPUBLIK DEUTSCHLAND [hrsg.] (2005). *Schüler, Klassen und Absolventen der Schulen 1994 bis 2003*. Bonn : KMK, Vol. 174.

Éclairages des évaluations internationales PIRLS et PISA sur les élèves français

Martine Rémond

Dans cet article, les résultats aux deux enquêtes internationales sur la reading literacy sont analysés et rapprochés de ceux des dernières évaluations-bilans françaises. Qu'ils soient en CM1 pour PIRLS ou évalués à 15 ans dans PISA, les élèves français font preuve d'un niveau assez moyen en matière de reading literacy parmi la trentaine de pays participants. Nos analyses croisées des dispositifs d'évaluation permettent de percevoir une sorte de genèse de l'hétérogénéité révélée par PISA en 2003 et de formuler des recommandations sur les contenus et sur les stratégies d'enseignement.

Descripteurs TEE : appréciation, compréhension de texte, conduite de la classe, études internationales, évaluation normative, expérience d'enseignement, métacognition.

À l'entrée dans le nouveau millénaire, deux programmes cycliques d'évaluation de la *reading literacy* ont été engagés : en 2000, PISA (*Programme for International Student Assessment*) sous l'égide de l'OCDE, en 2001, PIRLS (*Progress in International Reading Literacy Study*) sous celle de l'IEA (1). Ils entretiennent de nombreuses similitudes tant par leur conception que par leurs méthodologies.

PIRLS prend place au moment où les élèves passent de l'apprentissage de la lecture à la lecture au service des apprentissages (quatrième année de scolarisation obligatoire), alors que PISA porte sur des élèves de quinze ans, se situant donc à une phase de transition entre l'école et le monde du travail. PIRLS

cherche « à compléter PISA en adoptant une perspective centrée sur l'acquisition » (Campbell *et al.*, 2000, p. 62) (2). Il a été prévu pour recueillir des informations sur la compréhension des élèves d'une dizaine d'années en s'intéressant, entre autres, aux facteurs qui facilitent l'acquisition de la *reading literacy* aussi bien à l'école qu'à la maison. Il tente de mettre en évidence des liens entre le niveau de compréhension des élèves, le programme d'enseignement et les pratiques pédagogiques concernant la lecture.

Ces deux enquêtes reposent sur des processus d'élaboration similaires gouvernés par des cadres conceptuels de référence pour l'évaluation (*framework*), élaborés par des groupes d'experts internationaux

spécialistes du domaine (3). Chacun des cadres de référence prévoit, dans le détail, la structuration de l'enquête depuis la définition de l'objet à évaluer jusqu'aux échelles de présentation des résultats. Rendus accessibles par leur publication (Campbell *et al.*, 2000 ; OCDE, 1999), leur fonction méthodologique va au-delà de l'enquête qu'ils organisent : ils peuvent servir à nourrir la réflexion sur les contraintes d'un protocole d'évaluation, la première étant de définir l'objet qui sera évalué.

L'ORGANISATION GÉNÉRALE DE PIRLS ET DE PISA

Les définitions

Les définitions de la *reading literacy* conditionnent l'élaboration des dimensions de la compréhension. Dans PIRLS, la *reading literacy* est définie comme « l'aptitude à comprendre et à utiliser les formes du langage écrit dont la société a besoin et/ou qui sont nécessaires à l'individu. Les jeunes lecteurs sont amenés à construire le sens de textes très variés. Ils lisent pour apprendre, pour s'intégrer dans une société où la lecture joue un rôle essentiel, et pour leur plaisir. » (Campbell *et al.*, 2000, p. 3). Dans PISA, la *reading literacy* est définie comme « la capacité de comprendre, d'utiliser et d'analyser des textes écrits, mais aussi de réfléchir à leur propos afin de pouvoir réaliser ses objectifs, développer ses connaissances et son potentiel et jouer un rôle actif dans la société » (OCDE, 1999, p. 20)

Ces deux définitions, élaborées à un an d'intervalle, reflètent une conception interactive de la lecture où le lecteur fait appel à ses connaissances et à ses stratégies pour construire la signification du texte et utiliser l'écrit. Dans PIRLS, l'élève lit pour apprendre, mais comme dans PISA, il doit aussi participer à la vie collective, alors appelée « *communauté de lecteurs* », et répondre aux exigences de la société par le biais de la compréhension d'écrits variés.

La définition élaborée pour PISA permet d'envisager la dimension interprétative en mettant l'accent sur la capacité de « réflexion » ou de distance critique du lecteur, appelée *reflect on*. Elle ouvre aussi un horizon vers la « *préparation à la vie d'adulte* », les connaissances et les savoir-faire évalués étant supposés servir dans la vie d'adulte, d'où l'emploi de *reading literacy* (4).

Une fois donnée la définition de l'objet à évaluer, il faut fixer ce qui doit l'être. Nous reprendrons les terminologies employées dans chacun des programmes.

Que va-t-on évaluer ? Domaines de compétences évalués, caractéristiques des tâches

Le protocole d'évaluation (textes et items) de PIRLS est bâti sur la base du croisement de deux dimensions : les **processus de compréhension** (*processes of comprehension*) et les **objectifs de la lecture** (*purposes for reading*) qui « constituent la base de l'évaluation écrite de la compréhension de l'écrit » (Campbell *et al.*, 2000, p. 4). Des questionnaires sur les habitudes de lecture et les attitudes vis-à-vis de l'écrit sont soumis d'une part aux élèves, d'autre part à leurs parents. D'autres ciblant entre autres choses les pratiques pédagogiques et la formation sont renseignés par les maîtres des élèves évalués ainsi que par les directeurs d'école.

Trois dimensions ont été prises en considération pour élaborer le matériel d'évaluation de PISA : les types de tâches de lecture, les « types de textes » et le contexte.

Les processus de compréhension/les types de tâches de lecture

Les quatre processus que visait à évaluer PIRLS étaient les suivants (nous indiquons entre parenthèses la « dénomination courte » du processus qui sera employée par la suite) : « Prélever des informations explicites » (« Prélever »), « Faire des inférences directes » (« Inférer »), « Interpréter et assimiler idées et informations » (« Interpréter »), « Examiner et évaluer le contenu, la langue et les éléments textuels » (« Évaluer »). Chaque processus de compréhension est évalué pour chaque objectif défini ci-dessous.

Dans PISA, les « types de tâche de lecture » (*questions* ou *aspects* en anglais) définissent la nature des capacités à évaluer et sont considérées se situer à un « macro-niveau ». Ces tâches de lecture, au nombre de cinq au départ, ont été regroupées en trois catégories : « Retrouver des informations dans des matériaux variés » (« S'informer »), « Interpréter ce que l'on lit » (« Interpréter »), « Réfléchir à ce qu'on vient de lire et y réagir » (« Réfléchir »). Trois variables guidant la construction et la correction des items ont été isolées par I. Kirsch, A. Jungleblut et P. B. Mosenthal (1998) ; contrairement à PIRLS, elles sont mentionnées dans le cadre conceptuel de PISA. Ces variables qui jouent un rôle sur la difficulté de la tâche sont les suivantes : la plausibilité des distracteurs, le type d'information

recherché et le type d'appariement à réaliser. La plausibilité des distracteurs est une variable spécifique aux QCM. Ce format d'item se caractérise par plusieurs modalités de réponse : la bonne réponse et les autres qui s'appellent distracteurs. Plus les modalités de réponses partagent de traits communs avec la réponse pertinente, plus la tâche est jugée difficile. Le type d'information à rechercher pour répondre varie sur un *continuum* allant du très concret au très abstrait. Plus l'information attendue est concrète, plus la tâche est jugée facile. L'appariement entre la question et le texte est influencé par diverses conditions qui contribuent à la difficulté de la tâche : le nombre de phrases qui doit être utilisé dans la recherche, le niveau d'inférences à effectuer pour aller de la question au texte et pour identifier l'information requise, le nombre d'informations demandées par la question et leur localisation dans la phrase... L'appariement le plus simple se fait lors de tâches de localisation qui ne nécessitent que d'établir une correspondance directe, littérale ou synonyme, entre la question et le texte.

Types de supports

PIRLS retient deux **objectifs de lecture** : « Lire pour accéder aux textes littéraires » et « Lire pour acquérir et utiliser des informations », car « ils correspondent à la majorité des lectures faites par les jeunes élèves, à la fois à l'école et en dehors, et donc aux raisons de lire à cet âge » (Campbell *et al.*, 2000, p. 15).

PISA développe la dimension « **type de textes** ». Reprenant une classification présente dans des enquêtes antérieures de l'OCDE, la distinction est faite entre les textes « continus » (descriptif, narratif, informatif, argumentatif et injonctif) et les textes « non continus » (formulaire, graphiques, cartes, etc.).

En raison de la différence d'avancement des populations respectives dans leur scolarité, il est cohérent d'utiliser une plus grande variété de supports à quinze ans. Dans PIRLS, la distinction entre les deux objectifs de lecture laisse attendre des textes de nature différente (et variés si l'on se réfère à la définition). La *framework* de PISA fixe les « types de textes » et la proportion d'items dont ils seront la cible, de même que l'utilisation pour laquelle le texte a été écrit : « le contexte » ou « la situation » qui se décline en lecture à usage personnel, public, professionnel et scolaire. La lecture à usage personnel est effectuée pour satisfaire ses besoins personnels alors que la lecture à des fins publiques est pratiquée pour prendre part à des activités sociales au sens large. La lecture à des fins professionnelles peut surprendre dans le cadre

de ce travail, il s'agit de déterminer si les élèves de quinze ans sont prêts à une entrée dans le monde du travail et s'ils peuvent réussir des tâches de « lecture pour agir ». Cette dimension « contexte » est supposée refléter des situations de lecture représentatives de celles que fréquente la population de quinze ans.

Le contenu des protocoles et le traitement des données

Une méthodologie complexe conditionne la construction et l'organisation du matériel, élaboré en anglais pour former la « version source » servant de base commune aux traductions. À la base, un pool d'items est constitué (141 pour PISA 2000), chaque élève ne verra qu'une partie du matériel (textes et items afférents) selon le principe des « cahiers tournants ». Les protocoles de PIRLS reposent sur des textes longs et intégraux (au nombre de huit) sur chacun desquels portent une douzaine de questions. PISA comme d'autres évaluations de l'OCDE procède à partir d'un grand nombre d'exercices à l'unité (*unit*) qui contrairement à ceux de PIRLS sont plutôt courts, car il faut croiser de nombreux facteurs pour obtenir la déclinaison d'items correspondant aux caractéristiques définies (voir OCDE, 1999 & 2003).

Deux formats de questions sont, pour l'essentiel (il en est d'autres, au moins dans PISA, d'importance moindre) utilisés dans les deux enquêtes : des QCM à quatre modalités et des questions nécessitant de construire des réponses selon une modalité écrite.

Le codage des réponses aux questions ouvertes est fait en fonction des instructions figurant dans un guide afin d'effectuer des corrections aussi standardisées que possible, d'un pays à l'autre, d'une langue à l'autre. Les items les plus simples sont crédités d'un seul point tandis que d'autres, plus complexes, sont pondérés en fonction de la précision de la réponse.

Au final, dans un programme comme dans l'autre, on aboutit à une échelle de scores de compréhension élaborée à partir des modèles de réponse à l'item (MRI), la moyenne internationale est fixée à 500 et l'écart-type à 100 ; les pays sont classés sur une dimension commune (cf. Vriгдаud, 2006). Les résultats sont exprimés sur une échelle combinée de compréhension, et pour PISA sur trois échelles dites spécifiques rendant chacune compte d'un des trois types de tâches de lecture.

La méthodologie met l'accent sur la comparaison des performances entre les pays et dans les rapports internationaux publiés par l'OCDE ou par l'IEA,

« les scores des pays sont présentés comme sur un thermomètre (échelle de compétences), et conduisent à un palmarès... » (Rémond, 2005, p. 122). Dans les analyses qui vont suivre, nous nous appuyons essentiellement sur ces rapports (Mullis *et al.*, 2003 ; OCDE, 2001 & 2003) et sur des études complémentaires que nous avons menées à la demande de la DEP (Rémond, 2005 et soumis).

LES ÉLÈVES FRANÇAIS À L'AUNE DE LA « *READING LITERACY* »

Les deux enquêtes situent à un niveau moyen les performances en compréhension de l'écrit des élèves français. En 2000 comme en 2003, la France obtient aux épreuves de compréhension de l'écrit de PISA des scores proches de la moyenne, les scores respectifs étant de 505 et 496. Pour celles de PIRLS 2001, elle obtient un score significativement supérieur à la moyenne des pays participants (525).

Il n'y a pas de corrélation systématique entre les résultats aux deux programmes : ainsi trois pays ayant des scores moyens semblables à PISA (France, Norvège et États-Unis) obtiennent des scores moyens significativement différents les uns des autres à PIRLS.

Choix d'un sous-échantillon

Certains pays ont participé à la fois à PIRLS 2001 et à PISA 2000, mais ce n'est pas la majorité. Afin de pouvoir esquisser quelques comparaisons, nous avons sélectionné, parmi ces pays, un échantillon auquel nous limiterons notre propos. La langue anglaise y est représentée par le Royaume-Uni, les États-Unis, la Nouvelle-Zélande et le Canada (État de l'Ontario), et la langue française par la France et par le Canada, province du Québec ; l'Allemagne, l'Italie, la Norvège et la Suède font également partie de cet échantillon.

Notons que dans cette configuration, les meilleures performances sont obtenues pour PIRLS par la Suède et l'Angleterre et à PISA, par le Canada et la Nouvelle-Zélande.

Penchons-nous maintenant sur les dimensions présentées en première partie qui structurent les Programmes et tout d'abord sur les objectifs de lecture-types de supports.

La structuration des performances

Un effet des types de textes

Dans PIRLS, la France obtient de meilleurs scores aux items évaluant l'objectif « Lire pour acquérir et utiliser des informations » qu'à ceux qui relèvent de « Lire pour accéder aux textes littéraires » (5) (scores moyens respectifs : 533 *contra* 518), comme le Québec. En revanche, en Italie, en Norvège et dans tous les pays anglophones, l'objectif « Littéraire » donne lieu à de meilleures performances que l'objectif « Informatif ». Enfin, en Suède et en Allemagne les deux objectifs de lecture mènent à des performances similaires.

Dans PISA, la France réussit beaucoup mieux les items portant sur les textes « non continus » que ceux qui sont relatifs aux « textes continus », c'est d'ailleurs le pays qui présente l'écart le plus fort entre ces deux « types de textes ». L'Italie représente un cas isolé en réalisant de meilleures performances aux « textes continus » dans les deux programmes. Cet effet du « format des textes » peut probablement se rattacher à des différences culturelles et à des options pédagogiques privilégiant davantage certains écrits que d'autres.

Des capacités inférentielles et interprétatives limitées

Dans les deux Programmes, les échelles de performances exprimées en termes de scores de compréhension ont été divisées en **niveaux** qui rendent compte de la progression de la difficulté de la tâche. À chaque niveau, correspondent des tâches de lecture que les élèves sont capables d'accomplir avec une certaine réussite, montrant ainsi quelles compétences ils sont capables de mobiliser

Pour PIRLS, quatre niveaux ont été ainsi définis en fonction des quartiles (Niveaux 1 à 3, correspondant aux scores de compréhension atteints par 25, 50 ou 75 % de la population générale) et d'un repérage spécifique des 10 % ayant les meilleurs scores (niveau 4).

Dans ce programme, alors qu'elle a conditionné la construction du matériel, la variable « Processus » est à peine exploitée dans les analyses ; nous avons cherché à mieux en comprendre le poids à partir des niveaux de compétences et de leurs caractéristiques que nous avons synthétisées.

Au niveau 1, les élèves savent prélever des informations explicites et établir des inférences simples. Au niveau 2, ils maîtrisent le niveau précédent et de plus,

ils sont capables d'effectuer des inférences et des interprétations simples à partir d'informations puisées dans différentes parties du texte. Ils comprennent la structure générale du texte. Au niveau 3, ils sont également capables d'effectuer des inférences qui s'appuient sur différentes caractéristiques des personnages et des événements, et ils savent les justifier. Ils interprètent en faisant appel à leurs connaissances et expériences personnelles. Ils comprennent des procédés tels que la métaphore simple. Enfin au niveau 4, ils peuvent interpréter les intentions, les sentiments, les comportements des personnages en se basant sur le texte, ils sont capables d'intégrer des idées pour dégager le thème.

Le prélèvement d'information n'est mentionné qu'au niveau 1, mais il contribue bien évidemment à tous les niveaux. On observe que du niveau 1 au niveau 3, la nature des inférences évolue, mobilisant des processus attentionnels et langagiers de plus en plus complexes. La dimension interprétative apparaît dès le niveau 2, mais ce n'est qu'au niveau 3 que les élèves manifestent qu'ils savent aller au delà du texte, et au niveau 4, qu'ils peuvent expliciter leur démarche en se basant sur le texte.

Le tableau 1 présente les profils de résultats des pays en fonction des quatre niveaux définis ainsi que leur score global à PIRLS. L'ordre des pays figurant

dans le tableau a été commandé par des critères de langue du test en vue de faire apparaître un regroupement d'anglophones et un autre très modeste de francophones.

Le pourcentage théorique de 75 % définissant le niveau 1 est largement dépassé par la plupart des pays sur lesquels porte notre comparaison. Si l'on s'intéresse maintenant aux profils de distribution des élèves selon les niveaux et les pays, on constate qu'à une exception près, la Norvège, tous les pays retenus dépassent nettement le pourcentage théorique du niveau 2. Le niveau 3 se révèle beaucoup plus discriminant, notamment pour la France, dont le pourcentage devient similaire à celui de la population générale alors qu'aux niveaux 1 et 2, elle s'en écartait résolument. Cet effet se confirme au niveau 4. Par rapport à l'Angleterre, la Suède ou encore les États-Unis, le contraste est saisissant et montre à quel point la réalisation d'inférences et la capacité interprétative varient selon les pays pour des élèves effectuant une quatrième année d'école obligatoire.

En somme, seul le quart des élèves français de CM1 testés ici dépassent le niveau 2, et disposent de bonnes capacités inférentielles et interprétatives. Seule la Norvège obtient un pourcentage inférieur. Ce pourcentage s'élève à 45 % en Grande Bretagne et à 41 % aux États-Unis, en Ontario et en Suède.

Tableau 1. – Répartition des pourcentages d'élèves selon le score moyen et le niveau de compréhension atteint à PIRLS, en fonction du pays.

	Score moyen	Niveau 4 615 et +	Niveau 3 570 et +	Niveau 2 510 et +	Niveau 1 435 et +
Pourcentages théoriques	500	10	25	50	75
Angleterre	553	24	45	72	90
États-Unis	542	19	41	68	89
Nouvelle-Zélande	529	17	35	62	84
Canada (Ontario)	548	19	40	70	92
Canada (Québec)	537	11	31	67	94
France	525	9	26	60	90
Allemagne	539	12	34	69	93
Italie	541	14	36	69	92
Norvège	499	6	19	48	80
Suède	561	20	47	80	96

Pour PISA, les performances aux trois « types de tâches de lecture » sont exprimées par les trois échelles spécifiques : « S'informer », « Interpréter », « Réfléchir », elles-mêmes divisées en cinq niveaux. Les français retrouvent plus facilement l'information qu'ils ne développent une interprétation du texte, et ils sont significativement plus nombreux à atteindre le niveau 5 à l'échelle « S'informer » que la moyenne internationale. Leurs performances chutent pour l'échelle « Réfléchir » et passent même en dessous de la moyenne internationale. Les meilleures performances des filles sont vérifiées pour chacune de ces échelles et l'on note qu'à l'échelle « Réfléchir », les résultats des garçons diminuent davantage que ceux des filles, marquant un effet différentiel de ce type de tâche.

Chaque « type de tâches de lecture » est représenté à chacun des cinq niveaux de difficultés définis pour PISA. Les tâches regroupées sous le vocable « Interpréter » font davantage appel aux capacités inférentielles que « S'informer ». La tâche intitulée « Réfléchir » ou parfois « Réagir » évalue la capacité des élèves à faire appel à leur expérience personnelle, à leurs connaissances et à leurs idées pour les mettre au service d'un texte ; bien évidemment les capacités inférentielles et interprétatives sont convoquées, même si les caractérisations des tâches ne les mentionnent pas. Nous voyons que comme dans PIRLS, nos élèves rencontrent davantage de difficultés dès lors qu'ils doivent exercer ces capacités et que le malaise s'accroît avec « Réfléchir ». Ils manifestent, en revanche, une certaine aisance pour « Retrouver » de l'information dans PISA, mais les données disponibles ne permettent pas de le vérifier pour PIRLS.

En effet, pour PIRLS, les données concernant « les processus de compréhension », seconde dimension organisant le protocole n'ont été explorées qu'à la suite de l'insistance de pays participants désireux d'établir des comparaisons avec PISA (Mullis *et al.*, 2004). Pour des raisons méthodologiques, seules deux échelles de compréhension ont été bâties en regroupant, à chaque fois, deux catégories de « processus ; elles s'appellent « Prélever/Inférer » et « Interpréter/Apprécier ».

Un écueil perceptible dès le CM1, la « réflexion critique »

Les résultats aux échelles spécifiques de PIRLS et de PISA ont été rapprochés. Pour PISA, les pays anglophones, à l'exception de la Nouvelle-Zélande, ont des performances plus élevées à l'échelle « Réfléchir » qu'aux échelles « S'informer » et « Interpréter »,

et pour PIRLS, tous les pays anglophones ont une performance significativement meilleure à l'échelle « Interpréter-Apprécier ».

Cette constance dans la supériorité des anglophones lorsque le Processus évalué met en jeu la dimension « réflexive » (*reflect on*), mériterait d'être comprise. En nous échappant des limites de notre échantillon, signalons également un effet semblable pour cinq autres pays (Brésil, Espagne, Grèce, Mexique et Portugal), quatre d'entre eux étant hispanophones ou lusophones. L'OCDE (2003, p. 110) écrit : « certains profils de répartition semblent aller de pair avec des groupes linguistiques différents [...] Les pays où l'anglais n'est pas la langue d'enseignement présentent un tout autre profil [...] La variation de ces profils de performances peut s'expliquer par des pratiques pédagogiques différentes », avis partagé dans le cadre de PIRLS par I. V. S. Mullis *et al.* (2004) en ajoutant une perspective « liée aux programmes ». Dans une étude complémentaire sur PIRLS, M. Sainsbury *et al.* (2004) estiment que « parmi les différences internes aux pays, voire aux langues, il convient de remarquer le cas des pays anglophones... ».

L'hypothèse sous-jacente sur le poids des langues peut être émise sans encore avoir été l'objet de validation solide apportant un éclairage différent sur les performances. En revanche, quelques avancées peuvent être trouvées lorsqu'on s'intéresse aux programmes scolaires des pays participant à PIRLS. On l'a indiqué en introduction, les concepteurs de PIRLS voulaient explorer les liens potentiels entre le niveau de compréhension des élèves, le programme d'enseignement et les pratiques pédagogiques concernant la lecture. Aussi ont-ils collecté des données sur les systèmes éducatifs qu'ils ont rassemblées dans « PIRLS 2001 Encyclopaedia » (Mullis *et al.*, 2002). Le cœur du chapitre dévolu à chaque pays concerne les programmes et les politiques d'enseignement de la lecture et s'achève sur les pratiques évaluatives.

La consultation de cet ouvrage révèle que *reflect on* est inscrit dans les programmes d'enseignement de la lecture (*Reading Curriculum and Standards*) de tous les pays de langue anglaise, ainsi que de la Suède, alors qu'aucune référence n'y est faite dans ceux des autres pays de l'échantillon. Au-delà des programmes, « la réflexion critique » fait l'objet d'évaluations dans des dispositifs centralisés, dont voici deux exemples. Dans les évaluations britanniques – *National Curriculum Reading Test in England* (NCRT) –, des compétences liées aux activités « réflexives » sont évaluées tous les ans, dès l'âge de sept ans – fin du

Key-Stage (KS 1), étapes clés (6). Aux États-Unis, la « réflexion critique » est évaluée en grade 4 (vers dix ans) par la NAEP (*National Assessment of Educational Progress*).

Certains pays valorisent donc la dimension « Réfléchir » au niveau de l'enseignement scolaire, ce qui se reflète visiblement dans leurs performances aux évaluations internationales, un pôle linguistique lié à la langue maternelle anglaise étant facilement identifiable. Le pôle « Pratiques pédagogiques » a été exploré au moyen de questionnaires aux enseignants dont on rendra compte.

Les faiblesses des Français confirmées par des évaluations nationales

Les deux évaluations internationales offrent un éclairage nouveau sur les capacités de nos élèves, et les données qu'elles apportent peuvent être mises en perspective avec deux évaluations-bilans menées par la DEP en 2003 en fin de CM 2 et de classe de 3^e (cf. France, 2004a & 2004b).

Même si les deux évaluations présentent de nombreuses différences, les résultats français à PIRLS peuvent être rapprochés de ceux de l'évaluation intitulée « La maîtrise du langage et de la langue française en fin d'école primaire ». Celle-ci montre qu'un moins d'un tiers des élèves dispose des compétences attendues en fin de cycle 3 en matière de maîtrise de la langue. Ces élèves se caractérisent notamment par leurs capacités d'analyse et de synthèse, capacités sollicitées dans la réalisation d'inférences et l'interprétation des textes. PIRLS a montré que seulement 26 % des élèves de CM1 disposent de bonnes capacités inférentielles et interprétatives. Ces deux évaluations, nationale et internationale mettent donc l'accent sur la proportion limitée de nos élèves capables de mettre en oeuvre certains processus en jeu dans la compréhension de l'écrit.

Environ 35 % d'élèves participant à PISA sont scolarisés en classe de 3^e (et 50 % en classe de 2^{de}), ce qui permet de rapprocher les résultats de PISA de ceux de l'évaluation de fin de collège. Cette dernière met en évidence qu'environ la moitié des élèves encore scolarisés en classe de 3^e, ayant donc redoublé une fois au cours de leur *cursus* antérieur, n'est pas en mesure d'effectuer des inférences simples ; cette incapacité à se décentrer pour élaborer une inférence se manifeste très probablement dans PISA où une grande partie des items, à des niveaux divers, requiert ce type d'opération mentale (pour une analyse, voir Rémond, 2005).

DES COMPORTEMENTS RÉVÉLATEURS

Des relations difficiles à la production de réponses écrites

Les QCM sont nettement mieux réussis que les items demandant de rédiger une réponse, résultat assez classique et partagé par les pays participants à PIRLS ; cet effet est également constaté pour PISA mais seulement pour les pays francophones (Soussi *et al.*, 2004). Ni PIRLS, ni PISA ne visent à évaluer la production d'écrit alors que les réponses passent souvent par le canal de l'écrit, même s'il est explicitement indiqué dans PIRLS que l'écrit ne fait pas l'objet d'évaluation (Mullis *et al.*, 2003, p. 51). Plus que ceux d'autres pays, les jeunes français ne répondent pas à un certain nombre d'items pour lesquels il fallait rédiger une réponse, ce comportement s'observe dans les deux programmes.

Le pourcentage de non-réponses est spectaculairement plus élevé pour l'échelle « Réfléchir » que pour les échelles « S'informer » et « Interpréter » mais, contrairement aux autres échelles, les questions à réponse construite y sont quatre fois plus nombreuses que les QCM. Pour cette dimension, les élèves sont donc confrontés, à la fois, à un format de réponse requérant majoritairement le passage à l'écrit et à une tâche peu habituelle dans leur système scolaire.

La passation de PISA a dévoilé la réalité scolaire de la population française de quinze ans. Il s'agit d'une population dispersée depuis la classe de 4^e jusqu'à celle de 1^{re} : sans entrer dans les détails, en 2000 environ 50 % des élèves participant à l'enquête étaient scolarisés en 2^{de}, 36 % en classe de 3^e, 10 % en classe de quatrième. On remarque que la fréquence de leurs non-réponses apparaît très corrélée avec le niveau atteint dans les échelles de compétences et le retard scolaire, autrement dit, les élèves les plus en retard sont aussi ceux qui ont les performances les plus faibles et qui produisent le plus de non-réponses aux questions ouvertes.

Dans PIRLS, pour tous les processus évalués, nos élèves manifestent leurs difficultés face aux questions qui impliquent de construire une réponse consignée par écrit. Quand la tâche de rédaction est limitée et s'apparente à un prélèvement d'information, les performances deviennent supérieures à celles de la moyenne des pays, même si des opérations cognitives complexes sont en jeu. Il en est de même lorsque la question offre une « trame de réponse », en voici un exemple.

Cet item nécessite « d'interpréter et d'intégrer des idées » (texte *Le lièvre*, item H 9) :

« Les sentiments du lièvre changent pendant l'histoire.

« Complète les phrases suivantes :

- « Au début de l'histoire, le lièvre se sent...

« Parce que...

- « À la fin de l'histoire, le lièvre se sent...

« Parce que... »

Dans ce cas, nos élèves parviennent à montrer leur compréhension des relations de causalité entre deux événements alors qu'en l'absence de ce guidage, ils n'apportent pas la justification demandée dans la question et se contentent de donner une réponse partielle. Lorsque les étapes du raisonnement sont présentes dans l'ébauche de réponse, ils gèrent efficacement la tâche. Certaines tâches réputées complexes peuvent donc être accomplies avec succès si elles font l'objet d'aménagements dans leur présentation ; le coût cognitif de la mise en mots est alors réduit. Cette fluctuation des performances en fonction de la structuration des items semble indiquer leur très grande dépendance à l'égard des habitudes scolaires et un manque certain d'autonomie dans le traitement des tâches sous-jacentes, qui s'explique, entre autres, par des défauts métacognitifs.

Une gestion insuffisante des tâches

Une représentation rigide des tâches de lecture

Les élèves de CM1 ont trop souvent cherché à établir une correspondance directe entre la question et le texte, ce qui a pu les conduire dans des impasses. Certains ont persévéré dans l'application de cette conduite et recopié directement des éléments du texte alors qu'il fallait mener des opérations cognitives plus complexes et aller au-delà du texte. Leur mise en correspondance, assez directe et systématique, entre des éléments de la question et du texte amène à s'interroger sur les modèles de tâches qu'ils rencontrent dans leur vie de classe. Comme cet appariement direct constitue le mode dominant de leurs activités de lecture, certains ont construit une représentation restrictive et erronée des tâches de lecture dont ils ne s'éloignent pas, contrairement à d'autres élèves. Notre hypothèse est confirmée par les réponses des enseignants aux questionnaires de PIRLS relatées dans cet article, ainsi par les constats convergents sur la pauvreté du questionnement mené en classe (Beltrami *et al.*, 2004 ; Gaonac'h &

Fayol, 2003 ; Guernier, 1999 ; Rémond, 2003 & 2005 ; Tauveron, 2002), ou encore par l'examen de manuels récents de français pour le Cycle 3 (Nora, 2006).

Un contrôle métacognitif insuffisant

La représentation erronée de l'activité de lecture a été identifiée comme étant une caractéristique des faibles lecteurs/compreneurs ; ceux-ci disposent de moins de connaissances métacognitives que les bons lecteurs/compreneurs (Rémond, 2003). Des défaillances dans le contrôle métacognitif, se manifestant par des défauts de gestion de la tâche, ont été mises en évidence par les deux programmes.

La présomption d'un traitement partiel des consignes vaut pour les deux programmes. Certains élèves semblent s'être engagés dans la tâche sans avoir pris en considération les paramètres précisant la requête, ce qui peut les avoir conduits à choisir dans un QCM un distracteur fortement plausible faute d'avoir mené un traitement suffisant de l'information présentée.

La gestion défaillante de la consigne, et donc de la tâche, se traduit aussi d'autres manières. Ainsi dans PIRLS, où la plupart des histoires de l'objectif « Littéraire » comportaient deux personnages principaux, nous avons mis en évidence des difficultés de focalisation sur « le bon personnage », celui qui fait l'objet de la question. Des élèves continuent à raisonner sur le personnage sur lequel portait la question précédente. Cette difficulté « d'inhibition » d'un personnage se traduit aussi lors de questions dans lesquelles les deux personnages sont mentionnés, mais où il faut comprendre pour lequel les deux informations doivent être données. En voici un exemple particulièrement riche : « *Vers la fin de l'histoire, le lion essaie de rassurer le lièvre. Donne **deux** exemples de la façon dont il s'y prend* ». Certains élèves répondent en prenant un exemple pour chacun des deux personnages mentionnés dans la consigne. Il n'est pas impossible que par une stratégie de distance minimale, le pronom « il » ait été interprété comme le substitut du personnage « le lièvre ». Ce traitement défectueux d'une marque anaphorique met en lumière un problème introduit par la traduction, alors que les évaluations internationales ne portent pas sur le traitement des marques anaphoriques. La version anglaise centre sans ambiguïté la question sur le lion : « *Write **two** ways in which the lion tried to make the hare feel better at the end of the story* » Cet exemple démontre remarquablement la complexité des phénomènes que la non-équivalence de la consigne fait émerger, et l'effet du non-respect, dans la version française

de la « règle de l'ordre d'importance ». Selon cette « règle », dans la question, il faut hiérarchiser les informations présentées et indiquer en premier lieu ce qui doit être fait, ceci permettant au lecteur de mieux cibler sa tâche.

De manière répétée, au travers des réponses erronées, nous avons détecté des défauts métacognitifs. Le manque d'évaluation de la pertinence de la réponse par rapport à la question et/ou au texte est récurrent.

Un manque de flexibilité

Dans les deux programmes, les jeunes français se montrent très dépendants de stéréotypes de réponses, et donc peu préparés aux diverses exigences des rencontres avec l'écrit au sens défini par PISA. Ils s'interrogent beaucoup sur leur interprétation des tâches dès que celles-ci s'écartent de leurs habitudes scolaires (par exemple, juger un point de vue, faire abstraction de son opinion). Ils vont même jusqu'à ne pas répondre lorsqu'ils doutent de leur réponse ou qu'ils ne comprennent pas suffisamment la demande. Quand ils ne reconnaissent pas leurs objets habituels d'investissement scolaire, ils essaient, coûte que coûte, d'établir des ponts avec leur culture scolaire, en particulier, celle du cours de français. Lors d'entretiens que nous avons menés après la passation des protocoles, certains nous ont signalé des tâches auxquelles ils n'ont pas répondu en raison d'une question inhabituelle leur demandant, par exemple, d'exprimer leur opinion personnelle. Faut-il chercher un lien avec ce qui pourrait être interprété comme une peur de l'erreur se manifestant très tôt (pour une discussion, voir Rémond, 2005) ?

LES PRATIQUES PÉDAGOGIQUES AUXQUELLES SONT EXPOSÉS LES ÉLÈVES DE PIRLS

Les maîtres en charge des élèves évalués par PIRLS ont rempli un questionnaire composé d'un ensemble de quarante quatre questions dont certaines ont trait à la maîtrise de la langue. Ils ont dû indiquer la fréquence à laquelle ils font pratiquer un certain nombre d'activités visant le développement des capacités et des stratégies de compréhension. Les résultats sont rapportés en prenant l'élève comme unité, et le référent retenu comme unité d'étude est « *activité pratiquée au moins une fois par semaine* » (7).

En moyenne, « *l'identification de l'idée principale* » et « *l'explication/argumentation de ce qui est*

compris » représentent les activités les plus pratiquées : 90 % des élèves sont confrontés à ces activités au moins une fois par semaine contre 80 % en France. Deux autres activités sont assez communément proposées : « *comparer : confronter ce qui a été lu à son expérience personnelle* » (73 %) et « *effectuer des inférences et opérer des généralisations* » (71 %) ; ces deux activités ne sont pratiquées que par, respectivement, 31 % et 23 % des élèves français. Les activités « *d'anticipation de la suite du texte* » sont exercées par 61 % des élèves (contra 40 % en France), « *la comparaison de ce qui a été lu avec des lectures antérieures* » par 59 % des élèves (contra 23 % des élèves français), la « *description du style ou de la structure du texte* » par 52 % des élèves (contra 34 % de nos élèves).

Le temps qui est déclaré être dévolu à l'enseignement de la lecture est bien inférieur, en France, à la moyenne internationale et contrairement à d'autres pays, nos maîtres adaptent rarement le matériel en fonction du profil de leurs élèves. Enfin, bien qu'en France, les supports liés à d'autres disciplines soient déclarés être fréquemment des vecteurs d'activités relatives à la lecture, l'enseignement de la lecture dans toutes les disciplines reste très marginal (5 % contra 21 % à l'international).

Les maîtres français font lire à haute voix bien plus souvent que leurs collègues étrangers. Une exploration des tâches confiées aux élèves après la lecture a été menée par questionnaire. Elle montre que les élèves français répondent souvent à des « *questions orales après la lecture* », mais qu'ils reçoivent rarement et moins que la moyenne internationale des questions de compréhension nécessitant une réponse exprimée par écrit. Seuls 18 % (contra 68 %) ont l'occasion d'écrire « *un texte qui fait suite ou qui répond à ce qu'ils ont lu* », 32 % (contra 65 %) discutent « *entre eux de ce qu'ils ont lu* ». Ces trois tâches appartiennent à une liste de six tâches qui comporte aussi « *faire un dessin ou réaliser un projet artistique lié à ce qui a été lu* », « *présenter un sketch ou mettre en scène ce qui a été lu* » et « *réaliser un travail de groupe sur ce qui a été lu* ». Ces trois dernières tâches sont presque inexistantes dans nos écoles. Un indicateur a été calculé pour rendre compte de la variété des tâches, il indique la proportion d'élèves par pays qui se voient proposer, chaque semaine, au moins trois tâches parmi les six énumérées ci-dessus : seuls 17 % de nos élèves de CM1 sont dans ce cas alors que 59 % de leurs pairs participant à PIRLS réalisent au moins trois activités différentes sur les textes chaque semaine.

Les données qui viennent d'être présentées émanent de connaissances déclaratives fournies par les maîtres et non d'observations des pratiques de classes. Elles doivent bien sûr être interprétées prudemment, d'autant que l'on peut objecter des effets culturels ou pédagogiques et avancer que nos élèves rencontrent d'autres tâches au moins aussi formatives en matière de lecture. Cependant, elles contribuent à éclairer certains résultats. Ainsi, on note que la Suède, l'Angleterre, les États-Unis ou encore la Nouvelle-Zélande ont fréquemment recours à du travail écrit en lien avec la lecture ; ce sont également des pays où la « réflexion critique » est inscrite dans les programmes.

La différence d'empan dans les tâches proposées s'observe aussi dans les « types de texte » utilisés au cours des activités relatives à la lecture. La « fiction » est travaillée hebdomadairement en CM1 par quatre élèves sur cinq, nos élèves étant plus en prise avec des « livres divisés en chapitres » qu'avec des fables et des contes qui, contrairement à la moyenne des pays, demeurent largement absents des supports de lecture. Le faible usage des écrits « non-fictionnels » (« *non-fiction* » en anglais) caractérise également la situation française.

LES EFFETS INTER-LANGUES

On a déjà pu faire quelques remarques au sujet des effets inter-langues, nous allons en ajouter quelques autres en nous penchant davantage sur le matériel, car celui-ci fixe et conditionne la tâche.

On s'attendait à ce que des profils similaires de réponses aux quatre modalités d'un même QCM se dessinent, sans effet de langue ; mais de temps en temps, on observe des *patterns* de réponses différents selon les langues aux quatre modalités de réponse d'un même QCM. Ce phénomène est particulièrement facile à observer dans le cas du Canada où l'on constate pour nombre d'items un écart significatif et parfois important entre les résultats des deux provinces canadiennes, les résultats de l'Ontario étant proches de ceux des pays anglophones, ceux du Québec entretenant une forte similitude avec les résultats français. La plausibilité des distracteurs peut évoluer lors du passage à une autre langue et dans les deux programmes, on a pu observer que certains distracteurs, qui ne l'étaient pas en anglais, sont devenus parfaitement plausibles en français.

Tout ceci ramène à l'éternelle question de l'équivalence des épreuves dans un contexte international.

L'information a parfois subi une structuration bien différente de la version source en anglais, qui a souvent, mais pas toujours, complexifié la tâche de nos élèves. Les traductions ne tiennent pas forcément compte des règles d'énonciation qui devraient régir les consignes des matériels d'évaluation, en particulier de celle qui pourrait être baptisée « règle de l'ordre d'importance » évoquée précédemment (Rémond, 2001 & 2005). Enfin, les consignes de codage ne gèrent pas toujours les subtilités de la langue vers laquelle elles ont été traduites. Tout particulièrement dans PISA, la structuration de certains énoncés ne correspond pas à nos habitudes rédactionnelles. La consigne peut indiquer plus précisément, dans une des langues, la tâche à accomplir, ce qui explique alors une bien meilleure performance moyenne dans cette langue de test et, de manière corollaire, moins de réponses erronées et de non réponse. La question de la précision de l'information se pose parfois en termes d'équivalence du registre du lexique employé et donc de familiarité avec celui-ci.

Pour un certain nombre d'items de PIRLS, nous l'avons déjà souligné, la très nette réussite de la plupart des pays anglophones fait s'interroger sur la traduction en français, et parfois aussi dans d'autres langues.

Voici l'exemple d'un item (R 5) extrait de l'épreuve « Promenade le long de la rivière » dans sa version source en anglais : *How much would it cost a 10-year-old child to hire a bike for a full day ?* Cet item est devenu en français : « Un enfant de 10 ans veut louer un vélo pendant toute une journée : combien devra-t-il payer ? »

Ce item, qui prend ainsi la forme d'un énoncé de problème, est remarquablement réussi par les francophones et par la Suède (90 %), les anglophones atteignant une moyenne de 81 %, l'Italie et la Norvège environ 75 %.

Il faut donc accroître la vigilance lorsqu'on vérifie les traductions du matériel afin de ne pas changer la nature de ce qui est estimé d'une langue à l'autre, et donc ce qu'on évalue ; cette opinion est partagée par M. Sainsbury *et al.* (2004) à la suite d'un travail comparatif sur les profils de résultats dans plusieurs langues, déjà cité, et mettant en évidence la supériorité des anglophones à divers items.

Nous avons montré par exemple que les questions de PISA s'appuyant sur le « style » ne semblent pas

avoir le même sens pour un francophone et un anglophone, car manifestement, dans ces langues de référence, le mot « style » n'évoque pas le même concept. Dans PISA, évaluer « le style » ne revient pas à utiliser un savoir transmis par le cours de langue maternelle, mais à analyser, critiquer, s'approprier le texte.

Voici un exemple commenté portant sur le support de l'exercice « Grippe », un communiqué s'adressant à des salariés pour les inciter à se faire vacciner. La mise en forme, deux dessins et un slogan pour clore le communiqué donnent un caractère attractif à ce document qui se présentait à l'origine sous la forme d'un dépliant. Une question cible le « style ». La tâche requise nécessite d'évaluer la pertinence de la composition du texte et son adéquation au public visé. Repérer que le texte a été mis en page de manière attrayante et que sa tonalité est encourageante constituent une réponse correcte. L'essentiel relève de la présentation du communiqué (illustrations, typographie, mise en gras, sous-titres), ce qui ne relève pas du style, au moins, dans notre culture scolaire ; aussi le terme « style » employé dans la question ne peut que déstabiliser nos élèves (pour d'autres exemples, Rémond, 2005). Ces questions appartiennent à l'échelle « Réfléchir », si difficile pour nos élèves.

Enfin et pour prolonger la question de l'évaluation de la lecture littéraire, on a détecté des élèves qui produisent une réponse révélatrice d'une interprétation concurrente de celle qui est prévue par le codage international. Dans PIRLS, ces élèves se montrent très performants à l'ensemble de cette évaluation, faisant preuve d'un réel travail interprétatif, mais ils vont au delà de la programmation linéaire des questions, selon le déroulement chronologique de l'histoire. Le problème de la négociation de la marge d'interprétation a déjà été posé pour PISA (Rémond, 2004 & 2005). Plusieurs « bonnes réponses » à une même question pourraient être recevables, mais dans les deux programmes, des codages sélectifs ignorent des pans d'interprétation tout à fait plausibles ; ce fait se rattache à cette délicate question de l'évaluation de la lecture littéraire (Sainsbury & Watts, 2006 ; Sève, 2004).

COMMENT FAIRE PROGRESSER LES ÉLÈVES FRANÇAIS ?

Les deux évaluations internationales cibles de cet article offrent un éclairage nouveau sur les capaci-

tés de nos élèves en compréhension de l'écrit et sur leur pouvoir d'adaptation. Elles ouvrent des pistes de réflexion pour envisager de les faire évoluer, sans nécessairement chercher à les « conditionner » à ces types d'évaluation. Ainsi dans la mesure où les capacités inférentielles revêtent une grande importance dans le traitement de la langue orale et écrite, les constats que nous avons dressés, indiquent la grande vigilance que doit avoir l'école à leur égard.

S'interroger sur l'efficacité des pratiques

L'exploitation des questionnaires met en évidence un certain nombre d'éléments propres à la France en termes d'exposition à l'écrit : un temps consacré à l'enseignement de la lecture inférieur à la moyenne internationale, une adaptation peu fréquente du matériel en fonction du profil des élèves, beaucoup de questions orales après la lecture et peu de questions nécessitant une réponse exprimée par écrit, de rares échanges verbaux sur la compréhension des textes.

Les élèves français seraient moins exposés à des activités de nature à développer la compréhension que la moyenne de leurs pairs évalués par PIRLS : activités peu variées sur des supports eux-aussi peu variés. Ce « déficit » est particulièrement visible pour le travail de nature inférentielle et pour les activités de mise en perspective avec des lectures antérieures ou avec son expérience personnelle, désormais valorisées par les programmes français de 2002. Ces éléments contribuent à expliquer des performances révélées par les évaluations nationales ou internationales. Si les capacités inférentielles semblent faire défaut à nombre de nos élèves, ce serait parce que ceux-ci pratiquent très rarement des activités de nature à les développer. Il s'agit pourtant d'un champ crucial dans la construction de la signification du texte.

On observe que le traitement des aspects littéraires des textes estimé par les items de PIRLS est moins efficace chez les Français que pour la moyenne des pays. Ces items qui contribuent essentiellement aux niveaux de compétences les plus élevés font appel au point de vue, à la reconnaissance de techniques narratives, ou encore à l'évaluation de la plausibilité d'une histoire... Les maîtres déclarent travailler de manière prépondérante sur des textes de fiction, mais ils n'ont peut-être pas fait exercer ces types de traitements par leurs élèves.

Il faut continuer à faire porter les efforts sur la maîtrise de la langue en classe bien sûr, mais aussi dans

la formation des enseignants. Selon les questionnaires de PIRLS, certains secteurs sont beaucoup moins investis dans la formation initiale ou continue des enseignants français que dans la moyenne des pays participant à PIRLS : il s'agit, entre autres, de l'enseignement et des théories de la lecture, du développement du langage, de la psychologie.

Beaucoup de choses à l'école restent dans l'implicite alors que des travaux récents de psychologie « invitent à penser l'articulation entre apprentissage implicite et enseignement scolaire davantage en termes de complémentarité qu'en termes de substitution » (Perruchet & Pacton, 2004, p. 135).

Apprendre les élèves à comprendre

Dans certains programmes d'enseignement de la compréhension, il est conseillé de « rendre les processus transparents » au lecteur (Gaonac'h & Fayol, 2003), au lieu de laisser jouer l'implicite.

Pour que les élèves acquièrent, de manière graduelle, une certaine expertise et une efficacité face aux tâches de lecture, ils doivent apprendre à gérer leur compréhension et tout d'abord avoir conscience de l'attention conjointe qu'il faut apporter au texte et aux questions, par exemple dans les protocoles d'évaluation. La compréhension du texte est testée au travers de consignes qui n'ont pas toutes le même poids. Il faut donc leur apprendre à les lire et à les analyser, à se représenter la nature de l'information requise et à la situer par rapport au contenu du texte, qui devra nécessairement être lu. Ils doivent recevoir des questions variées pour découvrir les degrés de complexité du traitement à mener et dépasser le simple prélèvement d'informations ; ceci les amène à envisager qu'il faut produire des inférences pour relier les informations de la question et du texte. Leur conception tronçonnée des tâches les empêche de mobiliser des connaissances variées qu'ils doivent apprendre à activer (Beltrami *et al.*, 2004). Enfin, ils doivent évaluer la qualité et la pertinence de leur réponse, et éventuellement la modifier. En résumé, les élèves doivent se représenter la tâche à accomplir comme un fonctionnement en boucle qui oblige à examiner les relations réciproques entre les sommets du triangle : « question-texte-réponse ». Cette perspective est hautement fondée sur la métacognition et donc sur la régulation des apprentissages (Rémond, 2003). Ce raisonnement sur l'écrit doit être exercé le plus tôt possible pour que les élèves puissent construire la représentation mentale du contenu du texte la plus exacte possible. Les moda-

lités d'expression de la réponse doivent être elles-aussi travaillées afin de dépasser les constats dressés sur la production parcimonieuse de réponses écrites.

Les travaux de recherches soutenant ces perspectives se sont installés dans notre paysage, les derniers programmes de l'école parus en 2002 les ont pris en compte et le PIREF pour la conférence de consensus sur l'enseignement de la lecture (2003) avait demandé à Francis Grossmann d'établir une synthèse qui s'intitule : *Comment faciliter, développer et évaluer la compréhension des textes aux différentes étapes de la scolarité primaire*. Dans cette synthèse, l'auteur adopte deux points de vue, celui du linguiste et celui du didacticien. Dans une première partie ; il tente d'expliquer pourquoi les élèves ne comprennent pas ce qu'ils lisent et qui l'amène à définir des pistes explicatives, certaines « *pouvant être mises en relation... avec le développement des capacités interprétatives* ». Il analyse aussi les spécificités de la lecture scolaire et de « la compréhension de textes » à l'école, ou encore « les contraintes sémiographiques et la compréhension de la tâche » et « la difficulté d'articuler des objectifs différents ». Sans entrer dans une énumération inutile des entrées du document, on perçoit leur proximité avec certaines rubriques traitées dans cet article, montrant le manque de nouveauté de certains constats faits lors des évaluations internationales. La seconde partie de cette synthèse offre des propositions didactiques s'articulant sur les points de sa première analyse. Parmi celles-ci, on note : « mieux inscrire le temps de lecture dans le temps scolaire », « développer les capacités interprétatives » et « définir la nature et la place d'un entraînement à la compréhension et à l'interprétation », « mieux définir les règles de la lecture scolaire et améliorer la rédaction des questions », « diversifier et mieux intégrer l'évaluation... ».

Les missions du PIREF se sont arrêtées et ce travail n'a pas suffisamment été diffusé dans notre communauté. Il n'est pas trop tard pour en tirer profit et rendre nos élèves plus à l'aise en compréhension et en production d'écrit. Ils auront ainsi une meilleure estime d'eux-mêmes et moins peur de commettre des erreurs puisqu'ils pourront contrôler leur exécution des tâches.

Martine Rémond
martine.remond@inrp.fr
IUFM de Créteil
UMR Éducation et politiques (INRP – Université Lyon 2)

NOTES

- (1) Ces dispositifs sont détaillés dans M.-T. Céard *et al.* (2003) et dans des publications de la direction de l'évaluation et de la prospective (DEP) du ministère de l'Éducation nationale qui, en France, a mis ces évaluations en œuvre. Les sites internet de l'OCDE et de l'IEA peuvent aussi être consultés : [http://www.pisa.oecd.org] et [http://www.isc.bc.edu/pirls2001.html].
- (2) Dans la première partie de cet article, nous nous référons à cette publication dont nous avons traduit de larges extraits.
- (3) Nous avons participé à ces deux dispositifs en tant que membre du groupe d'experts internationaux.
- (4) *Reading literacy* est traduit par « culture de l'écrit » ou par « compréhension de l'écrit », dans les documents PISA-OCDE.
- (5) Pour simplifier, nous parlerons désormais de l'objectif « Littéraire » et de l'objectif « Informatif ».
- (6) Pour des précisions, on peut consulter M.-T. Céard *et al.* (2003, p. 110-111) et M. Sainsbury *et al.*, (2006).
- (7) Dans cette partie, les éléments indiqués entre guillemets correspondent au libellé de modalités de réponses à renseigner. Les pourcentages en regard de ceux de la France sont ceux de l'ensemble de la population de PIRLS.

BIBLIOGRAPHIE

- BELTRAMI D. ; QUET F. ; RÉMOND M. & RUFFIER J. (2004). *Lectures pour le cycle 3. Enseigner la compréhension par le débat interprétatif*. Paris : Hatier.
- CAMPBELL J. *et al.* (2000). *Framework and Specifications for PIRLS Assessment 2001 : Progress in International Reading Literacy Study*. Chestnut Hill : Boston College : IEA.
- CÉARD M.-T. ; RÉMOND M. & VARIER M. (2003). *L'appréciation des compétences des élèves et des jeunes en lecture et écriture et l'évolution de ces compétences dans le temps*. Rapport établi à la demande du Haut Conseil de l'évaluation de l'école, n° 11, 147 p. Disponible au format PDF à l'adresse [http://cisad.adc.education.fr/hcee/publications-2003.html]; consulté le 6 décembre 2006.
- FRANCE : MINISTÈRE DE L'ÉDUCATION NATIONALE : DIRECTION DE L'ÉVALUATION ET DE LA PROSPECTIVE (2003). *Note d'information*, n° 03.22 : « Les élèves de CM1. Premiers résultats d'une évaluation internationale en lecture (PIRLS) », rédigée par M. Colmant & A. Mulliez.
- FRANCE : MINISTÈRE DE L'ÉDUCATION NATIONALE : DIRECTION DE L'ÉVALUATION ET DE LA PROSPECTIVE (2004). *Note d'information*, n° 04.09, « Les compétences générales des élèves en fin de collège », rédigée par L. Dauphin et B. Trosseille.
- FRANCE : MINISTÈRE DE L'ÉDUCATION NATIONALE : DIRECTION DE L'ÉVALUATION ET DE LA PROSPECTIVE (2004). *Note d'information*, n° 04.10, « La maîtrise de la langue en fin d'école primaire », rédigée par F. Gilbert, J.-M. Levasseur et J.-M. Pastor.
- GAONACH'H D. & FAYOL M. (éd.), (2003). *Aider les élèves à comprendre du texte au multimédia*. Paris : Hachette.
- GUERNIER M. C. (1999). « Interactions verbales et construction de sens. Lectures d'un texte narratif ». *Repères : recherches sur le français langue maternelle*, n° 19, p. 167-182.
- GROSSMANN F. (2003). *Comment faciliter, développer et évaluer la compréhension des textes aux différentes étapes de la scolarité primaire ?* Synthèse rédigée pour la Conférence de consensus sur l'enseignement de la lecture à l'école primaire. Paris : Programme incitatif de recherche sur l'éducation et la formation (PIREF) : Paris, 4-5 décembre.
- KIRSCH I. ; JUNGLEBLUT A. & MOSENTHAL P. B. (1998). « The Measurement of Adult Literacy ». In T. S. Murray, I. S. Kirsch & L. Jenkins, *Adult Literacy in OECD Countries : Technical Report on the First International Adult Literacy Survey*. Washington [D. C.] : US Department of Education : National Center for Education Statistics.
- MULLIS I. V. S. ; MARTIN M. O. ; GONZALES E. & KENNEDY A. (2003). *PIRLS 2001 International Report*. Boston : International Study Center.
- MULLIS I. V. S. ; MARTIN M. O. & GONZALES E. (2004). *International Achievement in the Processes of Reading Comprehension. Results from PIRLS 2001 in 35 Countries*. Chestnut Hill : Boston College.
- MULLIS I. V. S. ; MARTIN M. O. ; KENNEDY A. & FLAHERTY C. L. (2002). *PIRLS 2001 Encyclopedia. A Reference Guide to Reading Education in the Countries Participating in IEA's PIRLS*. Boston : International Study Center.
- NORA E. (2006). « Le manuel scolaire au cycle 3, un outil d'apprentissage de la compréhension ». Communication au colloque international *Les manuels scolaires* : Montréal.
- OCDE (1999). *Measuring Student Knowledge and Skills : A New Framework for Assessment*. Paris : OCDE.
- OCDE (2003). *La lecture, moteur de changement : performances et engagement d'un pays à l'autre : résultats de PISA 2000*. Paris : OCDE.
- PERRUCHET P. & PACTON S. (2004). « Qu'apportent à la pédagogie les travaux de laboratoire sur l'apprentissage implicite ? » *L'Année psychologique*, vol. 104, p. 121-146.
- RÉMOND M. (2001). « Adapter n'est pas traduire : adaptation dans différents contextes culturels d'épreuves d'évaluation de la littéracie ». In C. Sabatier & P. Dasen (éd.), *Cultures, développement et éducation : autres enfants, autres écoles*. Paris : L'Harmattan, p. 171-184.
- RÉMOND M. (2003). « Enseigner à comprendre : les entraînements métacognitifs ». In D. Gaonach'h & M. Fayol (éd.), *Aider les élèves à comprendre du texte au multimédia*. Paris : Hachette, p. 205-232.
- RÉMOND M. (2005). « Regards croisés sur les évaluations institutionnelles », *Repères : recherches en didactique du français langue maternelle*, n° 32, p. 113-140.
- RÉMOND M. (soumis). « Que nous apprend PIRLS sur la compréhension des élèves français de 10 ans ? ». *Repères : recherches sur la didactique du français langue maternelle*.
- SAINSBURY M. ; SCHAGEN I. & HAMMOND P. (2004). *What did PIRLS Tell us About Reading Skills ?* Conférence de

- l'International Association for Educational Assessment* : Philadelphia, June.
- SAINSBURY M. & WATTS A. (2006). « Validity Challenges in a High-Stakes Context : National Curriculum Tests in England ». In M. Sainsbury, A. Watts, & C. Harrison (éd.), *Assessing Reading-Defining the Construct : From Theory to Classrooms*. The Mere : NFER.
- SÈVE P. (2004). « Peut-on évaluer la lecture littéraire ? » In *La lecture et la culture littéraire au cycle des approfondissements*. Versailles : CRDP-Scéren de l'académie de Versailles, p. 51-66.
- SOUSSI A. ; BROI A.-M. ; MOREAU J. & WIRTHNER M. (2004). *La littéracie dans quatre pays francophones : les résultats des jeunes de 15 ans en compréhension de l'écrit*. Neuchâtel : IRDP.
- TAUVERON C. [éd.] (2002). *Lire la littérature à l'école : pourquoi et comment conduire cet apprentissage spécifique ? De la GS au CM*. Paris : Hatier.
- VRIGNAUD P. (2006). « La mesure de la littéracie dans PISA : la méthodologie est la réponse, mais quelle était la question ? », *Revue française de pédagogie*, n° 157, p. 27-41.

Performances en littéracie, modes de faire et univers mobilisés par les élèves : analyses secondaires de l'enquête PISA 2000

Élisabeth Bautier, Jacques Crinon,
Patrick Rayou, Jean-Yves Rochex

Cet article rend compte d'un travail d'analyses secondaire et complémentaire des épreuves de l'enquête PISA 2000 portant sur la littéracie. Ont été conjugués une analyse a priori de ces épreuves, de leurs modalités de correction et de cotation – telles que prescrites par les concepteurs de PISA –, un re-traitement statistique des réponses de plus de 800 élèves français, construit à partir d'hypothèses théoriques différentes, et un recueil d'entretiens visant à mieux comprendre les modes de faire des élèves face aux épreuves considérées. L'ensemble des analyses ainsi produites conduit à interroger à la fois les conceptions de la littéracie et les présupposés méthodologiques et psychométriques sur lesquels repose l'enquête PISA 2000, et à contester la logique de « palmarès » dans laquelle ont bien souvent été publiés et commentés ses résultats.

Descripteurs (TEE) : compétence, études internationales, évaluation, littéracie, stratégie d'apprentissage.

Cet article vise à rendre compte d'un travail d'analyse secondaire de l'enquête PISA 2000, effectué en réponse à un appel d'offres de la Direction de l'évaluation et de la prospective (DEP) du ministère de l'Éducation nationale. On sait que l'enquête PISA 2000, mise en œuvre sous l'égide de l'Organisation pour la coopération et le développement économique en Europe (OCDE) dans une trentaine de pays, cherchait à évaluer les acquis des élèves de 15 ans (nés en 1984) (1) dans trois domaines distincts : la compréhension de l'écrit (plus largement, la littéracie), la culture mathématique et la culture scientifique. L'essentiel de PISA 2000 portait néanmoins sur la littéracie, les autres domaines y ayant une part

moindre ; l'inverse sera le cas pour les enquêtes ultérieures, PISA 2003 ayant été essentiellement centrée sur la culture mathématique, et PISA 2006 l'ayant été sur la culture scientifique.

LES PRÉSUPPOSÉS THÉORIQUES ET MÉTHODOLOGIQUES DE L'ENQUÊTE PISA

Dès la publication des résultats et dans la plupart des analyses secondaires qui l'ont suivie, cette enquête a donné lieu à force commentaires et

analyses axés, pour l'essentiel, sur la comparaison, voire le classement, des différents pays en termes de performance moyenne et de dispersion, d'efficacité et d'équité ; sur la mise en rapport des résultats de PISA 2000 avec les caractéristiques des différents systèmes éducatifs (cf. Duru-Bellat, Mons & Suchaut, 2004) ; sur l'existence et l'importance, malgré le soin apporté aux modalités de traduction des épreuves dans les différentes langues, de « biais culturels » entre les différents pays ou aires culturelles et linguistiques (cf. Robin, 2002 ; Rémond, 2001 & Rémond, 2006) ; plus rarement, à notre connaissance, sur les modalités mêmes de conception, théorique et méthodologique, de l'enquête ou sur les modes de faire des élèves enquêtés. Différents auteurs (Bain, 2003 ; Rocher, 2003 ; Bottani & Vrignaud, 2005) ont néanmoins émis l'hypothèse selon laquelle l'approche psychométrique mise en œuvre et la sophistication méthodologique qui en découle – avec pour visées principales de comparer ou différencier des élèves ou groupes d'élèves selon des niveaux de compétences –, avaient pu prendre le pas sur l'approche conceptuelle de la littéracie ou de la compréhension de l'écrit que tente d'évaluer PISA (2). La discussion porte à la fois sur le présupposé de l'existence et du caractère universel, ou du moins indépendant des contextes culturels et linguistiques, de l'objet que l'on cherche à mesurer, ici la ou les compétence(s) en compréhension de l'écrit, censée(s) s'appliquer à toutes les situations et tous les contextes, et sur celui (qui en découle) de la possibilité d'évaluer cette ou ces compétence(s) à l'aide d'une échelle unidimensionnelle (3). Ces présupposés conduisent à considérer la réussite aux différentes épreuves de cette échelle, évaluée à l'aide du modèle psychométrique dit « de réponse à l'item », comme relevant du seul facteur qu'est le niveau de compétence de l'élève, et les conditions de comparabilité des résultats obtenus dans les différents pays comme relevant dès lors, et pour l'essentiel, du soin accordé au choix et à la traduction des différentes épreuves de l'instrument de mesure (cf. Rocher, 2003). Si ces présupposés étaient vérifiés, on ne devrait pas constater d'écarts de réussite à un même item, à niveau de compétences équivalent, entre les élèves de différents pays ou différentes aires linguistiques. Surtout, les différents pays devraient avoir le même profil de réussite, c'est-à-dire la même hiérarchie des items ordonnés selon leur taux de réussite, ce qui n'est pas le cas : les résultats de PISA 2000 montrent non seulement qu'il existe des profils de réussite différents, mais aussi que « la proximité des profils de réussite des pays

coïncide avec leur proximité géographique, culturelle ou linguistique » (*Ibid.*). D'où l'intérêt et la nécessité de travailler à mieux comprendre ces différences et la spécificité de chaque pays ou aire culturelle ou linguistique (tout en conservant une partie comparative commune, mais plus restreinte), plutôt que de chercher à les gommer en les considérant essentiellement comme scories résultant de la seule imperfection de l'outil de mesure (pour une analyse plus approfondie, portant plus spécifiquement sur les modalités d'évaluation de la culture écrite, des présupposés, théoriques et méthodologiques, de l'enquête PISA et de ses usages sociaux, cf. Rochex, sous presse).

L'étude et l'analyse de ces différences interculturelles entre pays ou groupes de pays étant hors de portée d'une seule équipe, nous avons tenté, plus modestement, dans l'étude dont cet article tente de rendre compte, de mieux décrire et mieux comprendre les différences socioculturelles observées entre élèves scolarisés en France, en ne nous intéressant pas seulement à leurs réussites ou leurs échecs, tels que définis par les concepteurs de PISA, mais à leurs modes de faire pour répondre aux épreuves et à ce qu'ils mobilisaient pour cela.

CADRE THÉORIQUE ET MÉTHODOLOGIES DE RECHERCHE

Nous avons également procédé à une analyse *a priori* de certaines des épreuves de PISA visant à évaluer les compétences en littéracie, ainsi que de la manière dont les modalités d'interrogation et les consignes de correction et de cotation pouvaient méconnaître et réduire la diversité, voire l'inégale « qualité » des manières de faire et des réponses des élèves. Notre travail a donc procédé selon trois étapes, que nous pouvons qualifier d'analyses exploratoire, secondaire et complémentaire des données recueillies par l'enquête PISA 2000 concernant les compétences en littéracie. Deux hypothèses ou constats ont présidé à la problématique et aux méthodologies de recherche que nous avons mises en œuvre pour cela, qui interrogent l'une et l'autre la conception et le modèle unidimensionnel de la compétence en littéracie qui sont ceux de PISA.

La première étape de ce travail, d'« analyse exploratoire », a consisté en la lecture et l'examen approfondi de plus d'une centaine de cahiers renseignés par des élèves ayant subi les épreuves de PISA 2000.

Cet examen nous a permis de constater que, pour bon nombre d'élèves, il était très difficile de prédire, à la lecture de leur réponse (vraie ou fausse) à tel item ou groupe d'items, de quelle nature seraient leurs réponses aux suivants ; de constater donc une importante hétérogénéité, de réponses et de modes de faire, non seulement entre élèves mais entre réponses d'un même élève aux différents items, hétérogénéité qui atteste que les compétences que PISA 2000 visait à évaluer ne sont ni univoques ni stabilisées pour beaucoup d'élèves testés, contrairement aux présupposés des conceptions de la littéracie et du modèle psychométrique utilisés. À un tel constat, s'est rapidement ajoutée notre perplexité devant le fait qu'étaient pareillement évaluées par les consignes de correction de PISA des réponses nous semblant relever de modes de travail différents, voire intellectuellement inégaux, soit parce qu'il était possible, pour certains items, de donner la réponse attendue à partir de modes de travail minimalistes, voire peu pertinents, ou au contraire plus élaborés, soit parce qu'étaient considérées comme non valides, et donc ne permettant d'obtenir aucun crédit, aussi bien des réponses manifestement incohérentes et liées à une non compréhension du texte ou de la question, que des réponses plausibles témoignant que ceux-ci étaient réellement compris, quoique ne correspondant pas à la réponse attendue (cf. *infra* l'analyse *a priori* d'une des épreuves sur lesquels nous avons plus particulièrement travaillé).

Ces premiers constats et analyses ont renforcé notre hypothèse de travail, fondée sur une conception différente de la littéracie mais aussi sur nos recherches antérieures (Bautier & Rochex, 1998 ; Rayou, 2002 ; Bautier & Goigoux, 2004), selon laquelle le travail requis pour répondre aux épreuves de PISA, surtout à celles qui demandaient d'interpréter et de commenter les textes et documents soumis à examen, était de nature très hétérogène, empruntant à plusieurs univers de référence et sollicitant des élèves la capacité de les conjuguer et de les hiérarchiser. Parce qu'ils visent, selon les concepteurs de PISA, à évaluer « l'aptitude à réaliser des tâches qui s'inscrivent dans des situations de la vie réelle » plutôt que l'acquisition de connaissances spécifiques, et parce que leurs thématiques sont souvent proches des expériences et des univers sociaux et culturels des adolescents, les épreuves et les textes de PISA, plus sans doute que les épreuves scolaires plus traditionnelles dans notre système éducatif, sollicitent potentiellement des élèves qu'ils mobilisent et sachent conjuguer plusieurs registres de référence. Le travail à opérer pour répondre aux épreuves n'em-

prunte pas au seul registre du traitement des textes et questions proposés ; il peut, selon les épreuves et/ou les élèves, être facilité ou entravé, aidé ou parasité, par la mobilisation ou la mise à distance que les enquêtés parviennent ou non à opérer des cinq registres suivants :

- le registre des savoirs scolaires, des connaissances générales sur le monde et des références culturelles qu'il faut mobiliser pour comprendre tout ou partie des textes et pouvoir formuler des hypothèses à partir d'eux (sur les motifs conduisant les gens de tel village d'Amérique du Sud à se conduire comme ils le font, ou sur les raisons du taux d'intervention quasi nul d'une organisation humanitaire en Éthiopie, par exemple) ;
- le registre de l'expérience personnelle, individuelle ou collective, et plus ou moins familière, que les adolescents ont des objets et des pratiques sociales et culturelles évoqués par les textes des épreuves : la plage, le cinéma, les graffiti, les échanges linguistiques, etc. ;
- le registre des valeurs, des opinions et préjugés, des *doxa*, juvéniles et/ou sociales, dont cette expérience est tissée, des « perspectives » qui sont celles des élèves sur le monde et sur les diverses pratiques sociales, *doxa* et « perspectives » par rapport auxquels certains textes ou certaines de leurs propositions peuvent s'avérer contradictoires ou contre-intuitifs (4) ;
- le registre des habitudes et des modes d'interprétation des situations scolaires ou proches de l'école (telles que celle de la passation de PISA qui fait souvent appel à des capacités d'écrire de façon brève et synthétique), construits et routinisés dans la fréquentation des tâches et des exercices scolaires ;
- le registre des compétences et des postures cognitives liées à une acculturation propre à la littéracie et permettant aux élèves d'opérer un travail de reconfiguration et de secondarisation de leur expérience du monde et de leurs usages du langage (5), travail qui ne fait que rarement l'objet d'une pédagogie explicite. Un tel registre est requis par les nombreux items demandant aux enquêtés d'appréhender une opinion proposée non seulement du point de vue de son contenu, mais de celui de sa forme, de son style, de son mode d'argumentation.

La deuxième étape de notre recherche a consisté en un travail d'analyse secondaire qui visait à mieux décrire et comprendre les logiques et les modes

de faire sous-tendant les réponses fortement différenciées des élèves aux épreuves de PISA, modes de faire liés à des processus – eux-mêmes différenciés entre élèves et plus ou moins stables pour un même élève selon les épreuves – de mobilisation et de hiérarchisation des cinq registres évoqués ci-dessus. Pour ce faire, nous avons effectué de nouvelles analyses, quantitatives et qualitatives, à partir de deux des cahiers d'évaluation, ceux qui comportaient le plus grand nombre d'items liés à la compréhension de textes longs et de questions faisant appel à une écriture argumentative (6). Nous avons procédé à un recodage de l'ensemble de ces deux cahiers (7), à partir d'une grille élaborée pour tenter d'appréhender la diversité des registres mobilisés par les élèves, en construisant différents types d'indices (8). Indices de compétences en lecture tout d'abord : il s'agit, d'une part, de compétences cognitives à traiter les textes ; et d'autre part de compétences à considérer un texte comme un discours « adressé » porteur d'intentions dans le cadre d'une communication sociale, à le constituer comme un objet social et langagier. Pour prendre en compte les autres registres selon nous en jeu dans les modes de faire des élèves, nous avons également construit deux autres types de catégories et d'indices. Le premier renvoie aux manières de l'élève d'« être au monde », aux valeurs, croyances, *doxa* et habitudes intellectuelles qui lui permettent d'interpréter le monde et l'expérience et qui interviennent dans la lecture des textes qui lui sont proposés. Le second concerne des indicateurs « transversaux » codés dans l'ensemble du cahier et non épreuve par épreuve, et vise à appréhender la manière dont l'élève se situe par rapport aux savoirs scolaires, ainsi que des compétences relatives au code écrit dont l'absence est souvent jugée caractéristique des élèves en échec (lacunes syntaxiques et orthographiques massives, « pauvreté » du vocabulaire, etc.), ou encore la fréquence des non-réponses. L'ensemble des réponses ainsi recodées à partir de cette grille a fait l'objet de deux traitements statistiques – analyse en composantes principales, puis classification ascendante hiérarchique – visant à constituer des classes ou des profils d'élèves, en fonction de la proximité de leurs réponses. En vue de travailler la question de la variabilité ou de la stabilité des modes de faire en fonction des thèmes et des tâches spécifiques aux différentes épreuves, nous avons ensuite procédé (sur un seul cahier) à une analyse qualitative et longitudinale de l'ensemble des réponses d'une cinquantaine d'élèves, en nous intéressant plus particulièrement à ceux qui apparaissaient caractéristiques des différents profils.

Nous avons, dans une troisième étape, complémentaire, également tenté d'appréhender les modes de faire différenciés (inter- et intra-individuels) des élèves, au-delà de ce que nous permettaient d'analyser les seules réponses obtenues. Pour cela, nous avons reconstitué une sorte de cahier comportant six des épreuves de PISA 2000, choisies parce que, portant sur des textes longs et demandant aux élèves des réponses écrites élaborées, elles nous semblaient les plus à même de donner à voir la pluralité de postures, de registres et univers de référence, dont nous souhaitons étudier la mobilisation et les modes de configuration (9). Nous avons ensuite soumis les six épreuves de ce cahier à une quarantaine d'élèves, alors âgés de 15 ans comme l'étaient les élèves évalués par PISA 2000, avec lesquels nous avons conduit un entretien *ex post* immédiat. Nous avons choisi d'enquêter ainsi seulement auprès d'élèves scolarisés en classe de 3^e ou de 2^{de} d'enseignement général ou technologique, ces deux catégories regroupant plus de 79 % des élèves évalués par PISA 2000. En revanche, parce que nous intéressaient tout particulièrement les difficultés et les modes de faire des élèves de milieux populaires, nous avons essentiellement interrogé des élèves scolarisés dans des collèges et lycées recrutant une majorité d'élèves issus de milieux socialement défavorisés (dont deux collèges ZEP).

RÉSULTATS ET MODES DE FAIRE, UNE RELATION PROBLÉMATIQUE : EXEMPLE D'UNE ÉPREUVE PISA

Les rapports pour le moins problématiques entre, d'une part, les résultats des élèves, tels que les construisent les épreuves de PISA et leurs signes de cotation, et, d'autre part, les modes de faire différenciés des élèves face à ces épreuves peuvent être illustrés en prenant l'exemple d'une des épreuves sur lesquelles nous avons plus particulièrement travaillé. Celle-ci portait sur un texte dans lequel étaient présentés les avis, favorables ou défavorables, au nettoyage mécanique des plages, de diverses personnes ou catégories de personnes, et les arguments invoqués en faveur de chacun de ces avis. Ce qui a posé problème aux élèves dans cette épreuve tient moins du caractère peu familier ou inconnu de certains termes du texte (dont le contexte permettait assez facilement de deviner, même approximativement, la signification), que du caractère contre-intuitif, voire absurde, au regard de leur expérience

et de leur opinion commune, des positions et arguments de certains experts et défenseurs de l'environnement hostiles au nettoyage mécanique des plages parce qu'il accroît l'érosion des côtes et contribue à la destruction de la faune et de la flore, positions et arguments fondant le titre paradoxal donné au texte : « Sauvez une plage : gardez la sale ».

Deux questions de type QCM étaient posées à propos de ce texte. La première est massivement réussie par les élèves français (73 %), mais les entretiens que nous avons conduits (cf. *infra*) nous ont montré qu'une telle réussite pouvait résulter de modes de faire très différents et reposant soit sur une réelle compréhension du propos (permettant de surmonter la méconnaissance de certains des termes de la question), soit sur le seul repérage d'éléments très partiels du texte permettant de trancher entre les différentes réponses proposées, repérage qui s'accommodait parfois d'une compréhension toute relative, voire très problématique du passage concerné. La seconde des questions de ce type a posé plus de problèmes aux élèves français (taux de réussite : 52,6 %) : il s'agissait de dire quelle conséquence aurait le fait de laisser les déchets sur les plages, pour les adversaires de leur nettoyage mécanique. Le libellé de la question, invitant à réfléchir sur les conséquences d'une mesure préconisée par des personnes définies par leur opposition à la mesure contraire, s'est révélé piégeant pour une partie des élèves, pressés ou peu attentifs à la formulation de la question, et dès lors portés à dire les conséquences du nettoyage des plages plutôt que celles du maintien des déchets. Les élèves qui ne se sont pas laissés piéger par la formulation de la question pouvaient trouver la réponse attendue quasiment en toutes lettres dans le texte. Une part importante des élèves ayant bien compris le texte dans son ensemble, ont néanmoins choisi une autre réponse que la réponse attendue, réponse tout à fait plausible selon laquelle le fait de laisser les déchets sur les plages pouvait amener les touristes à se sentir plus responsables à l'égard de l'environnement, réinvestissant ainsi les discours sociaux ou institutionnels en appelant à la responsabilité de chacun en matière de préservation de la nature et du cadre de vie. Le recours à des cadres de référence extérieurs au texte conduit ici à opter pour une réponse certes plausible et intelligente, attestant une bonne compréhension de la signification du texte et de la question, mais néanmoins considérée comme incorrecte par les consignes de correction de PISA, alors que le choix de la réponse attendue n'atteste pas toujours, loin de là,

une bonne compréhension de la signification globale du texte ou du passage concerné.

Une autre question demandait aux élèves d'indiquer, en entourant le mot approprié, si les différentes personnes ou catégories de personnes évoquées dans le texte étaient favorables ou défavorables au nettoyage mécanique des plages. Cet item a été réussi par un peu plus de 50 % des élèves français, mais ce score, pas plus que les consignes de correction, ne semble pas faire de distinction entre le travail requis pour répondre correctement concernant telle ou telle personne ou catégorie de personnes, toutes les réponses justes (ou fausses) étant considérées comme équivalentes, alors que l'analyse *a priori* de la tâche et celle des propos des élèves commentant leurs réponses montrent qu'il n'en est rien. L'interrogation portait en effet sur l'avis de trois personnes, désignées par leur nom dans le texte et les questions, et à propos desquelles les élèves ne pouvaient mobiliser aucune connaissance ou aucun *a priori* extérieur au texte ; il n'en allait pas de même concernant les deux catégories de personnes désignées comme *patrons du tourisme* ou *riverains défenseurs de l'environnement*. De nombreux élèves nous ont ainsi dit ne pas avoir hésité concernant les premiers, et ne pas même avoir besoin du texte pour répondre que les patrons du tourisme ne pouvaient qu'être favorables au nettoyage des plages (en oubliant qu'il était question de nettoyage *mécanique*, oubli sans conséquence sur le choix de la réponse) ; ici, la connaissance et l'expérience « ordinaires » sont au service de la bonne réponse. Tel n'était pas le cas concernant les riverains défenseurs de l'environnement car le fait que des riverains d'une plage puissent être opposés à son nettoyage va à l'encontre de l'opinion et de l'expérience ordinaires, et il fallait, pour choisir la bonne réponse, savoir tenir celles-ci à distance pour prêter attention à la formulation *riverains défenseurs de l'environnement* (et non simplement *riverains*) et à l'enchaînement des propositions du texte, la bonne réponse ne pouvant être trouvée à partir du repérage d'une seule de celles-ci. Un tel repérage simple suffisait en revanche pour répondre correctement à la même question concernant deux des trois personnes nommément désignées, ce qui n'a donc guère posé problème aux élèves. Tout autre était la même question (en apparence) concernant une de ces trois personnes dont les propos cités dans le texte, peu compréhensibles pour les élèves, ne leur permettaient que très difficilement d'en déduire l'avis qui était le sien à l'égard du nettoyage mécanique des plages. Il fallait pour cela être attentif moins à la signification locale des propos cités,

qu'à l'expression *X a précisé* permettant d'inférer qu'il était du même avis que le personnage cité précédemment, dont les propos étaient sans ambiguïté, ou encore au début de la phrase suivante (« Les touristes ne sont pas d'accord... ») permettant d'inférer que son avis était opposé à celui des touristes ; être attentif donc à des indices plus textuels que sémantiques, qui n'ont été utilisés ou repérés que par très peu d'élèves, à en croire les commentaires recueillis auprès de ceux que nous avons interrogés, y compris lorsque nous les invitions explicitement à être attentifs à l'enchaînement des propositions du passage concerné. Une telle attention aux indices textuels et non seulement sémantiques n'a pu être observée (ou induite lors de l'entretien) que chez de bons élèves, tous scolarisés en classe de 2^{de}. Mais ni la prise en considération des seuls propos du personnage concerné, ni leur fréquente incompréhension n'empêchent les autres élèves de répondre, en ayant une forte probabilité de fournir la réponse attendue, puisqu'il suffisait, ici comme pour les questions concernant les autres personnes ou catégories de personnes, de cocher l'une des deux réponses proposées. L'examen de ces cinq questions, composantes d'une question plus générale de l'épreuve PISA, montre ainsi, d'une part, combien le choix de la réponse attendue peut masquer, d'un élève à l'autre mais aussi d'une question à l'autre pour un même élève, des modes de travail très différents et intellectuellement forts inégaux, d'autre part combien le format de la question (de type vrai/faux) et la sommation des réponses portent à occulter la diversité des difficultés posées par cinq questions apparemment semblables, mais en réalité fort différentes, quoique considérées sans doute un peu trop rapidement comme équivalentes.

La dernière question de l'épreuve demandait aux élèves de dire s'ils pensaient que le titre paradoxal donné au texte était « un bon résumé du débat présenté dans l'article », puis d'expliquer leur réponse. Était considérée comme bonne, sans pondération, toute réponse, positive ou négative, indiquant que l'article exposait deux points de vue opposés, alors que le titre n'en reprenait qu'un. Aucun crédit supplémentaire n'était donné aux réponses des élèves s'interrogeant ou réfléchissant, dans une posture seconde, sur la visée d'un tel titre (en prêtant, par exemple, à son auteur une intention ironique ou une volonté de surprendre pour accrocher le lecteur) ou sur ce que peut ou doit être un débat. Cet item obtient des scores de réussite très faibles : 23 % pour les élèves français, 17 % pour l'ensemble de l'enquête PISA. De fait, la grande majorité des élè-

ves ne parviennent pas à tenir à distance leur opinion sur les avis et arguments exposés dans le texte, pour en différencier la réflexion attendue sur la pertinence du titre donné à celui-ci ; dès lors, ils commentent moins le titre que le texte, ou leur attitude à l'égard de celui-ci ou de la thèse la plus contre-intuitive par rapport à leur opinion première. D'autres, plus réflexifs, se saisissent de la question pour rendre compte, moins de leur opinion que de leur évolution à la lecture du texte, sans pour autant réfléchir sur son titre. D'autres encore prêtent attention au titre, mais oublient l'idée de débat, soit pour dire que le titre est pertinent parce que reflétant l'avis qu'ils jugent majoritaire, soit pour proposer un autre titre résumant la même thèse (par exemple « Une autre solution pour une plage propre »). Toutes ces réponses, qui à l'évidence ne relèvent pas du même travail d'interprétation et de réflexion sur le texte et son titre, sur soi et son expérience du monde, sont néanmoins créditées du même score nul par les consignes de correction de PISA qui, là encore, écrasent des différences non négligeables à nos yeux.

CLASSES ET PROFILS D'ÉLÈVES : STABILITÉS ET INSTABILITÉS

Dans la mesure où les tâches proposées dans chacun des deux cahiers étaient en partie différentes, nos analyses statistiques ont porté de manière disjointe sur les *corpus* de réponses propres à chacun d'entre eux. Elles nous ont permis de créer, pour chacun des cahiers, cinq classes d'élèves, qui s'avèrent néanmoins très proches d'un cahier à l'autre. Trois catégories « stables » de lecteurs ressortent de l'analyse, que l'on retrouve aussi bien dans l'un et l'autre cahiers.

Une première catégorie est celle des élèves (16,3 % du *corpus* ; 19,5 % et 13,1 % pour chacun des deux cahiers) que nous avons appelés « abstentionnistes », parce que se caractérisant par une forte fréquence de non réponses, comme s'ils démissionnaient devant la difficulté. Ainsi Maryse ne répond à aucune des questions qui demandent de faire autre chose que de prélever une information ponctuelle. Elle répond aux QCM, mais se trompe dès qu'il s'agit de construire la signification globale du texte, car elle ne s'attache qu'à un détail du texte. La caractéristique dominante de ce groupe d'élèves ne permet guère d'aller plus loin dans l'analyse de leurs modes de faire et de mobilisation, si ce n'est pour dire qu'ils sont à

62 % des garçons, qu'ils obtiennent des scores faibles à PISA, et que près de huit sur dix sont encore collégiens.

Une seconde catégorie (20 % du *corpus* ; 18,8 % et 21,6 % pour chacun des deux cahiers) est celle des élèves « en échec généralisé », qui se caractérisent par leur incapacité à mobiliser aucune des connaissances, compétences ou stratégies nécessaires à la lecture. Pour répondre, ils tentent maladroitement de répéter des passages du texte à lire ou de la question posée ou bien au contraire répondent « à côté » du texte. Ils prennent alors volontiers appui sur des discours tout faits, empreints de relativisme ou de conformisme, qui viennent se substituer à ce que le texte dit vraiment, quand ils ne tiennent pas des propos dénués de logique. À la différence des premiers, ils tentent néanmoins de répondre aux questions. Obtenant, eux aussi, des résultats faibles (mais moins faibles que les précédents) à PISA, ces élèves sont pour 51 % d'entre eux des garçons et pour 73,5 % des collégiens. L'analyse longitudinale de l'ensemble des réponses d'élèves particulièrement représentatifs de cette classe permet d'aller au-delà de ces constats, et montre que ces élèves ne prennent pas le texte comme texte, mais comme expression de soi. Ils font essentiellement référence à eux-mêmes, à leurs sentiments et à leurs expériences pour répondre et ne peuvent pas construire un autre texte en croisant leur propre voix avec celle du texte proposé. Parce qu'ils sont sans doute prioritairement dans l'usage d'un genre premier du langage et dans la difficulté d'opérer un travail de secondarisation de leurs propres paroles, expérience ou pensée, celles-ci ne peuvent être tissées avec les voix de textes qui sont des voix secondes. Leurs réponses, majoritairement fondées par un point de vue affectif ou doxique, ne mobilisent que fort peu de savoirs ou de travail de mise à distance. Leur lecture des textes ne s'attache pas à ceux-ci dans leur matérialité. Ceux d'entre eux qui obtiennent des résultats un peu meilleurs sont ceux qui repèrent sans guère comprendre. Dans leurs bonnes réponses à de nombreux items ils ne manifestent pas une compétence textuelle prenant en compte le texte comme entité spécifique. Ils savent repérer une information dans un texte et peuvent ainsi produire de bonnes réponses, plus fréquemment que les précédents dont ils se distinguent par l'usage réitéré des citations de morceaux du document.

Une troisième catégorie (21 % du *corpus* ; 21,9 % et 20,1 % pour chacun des cahiers) s'oppose sans surprise aux deux premières. C'est celle des élèves

que nous avons qualifiés de « bons compreneurs », capables de mobiliser l'ensemble des compétences cognitives et des postures face aux textes favorables à la littéracie, pour aboutir à de bonnes réponses aux questions posées sur ces textes. Ces élèves manifestent de bonnes compétences orthographiques et syntaxiques (10). Ils ne mobilisent pas de points de vue conformistes et lorsqu'ils font référence à une *doxa* relativiste, cela ne gêne pas leur compréhension du texte. Ils savent, pour comprendre un texte, le questionner, faire des hypothèses, procéder à des inférences causales. Ils font preuve d'un talent particulier pour se décentrer de la lettre du texte, pour en saisir la signification essentielle et la reformuler avec leurs propres mots, et pour exprimer, en regard, leur propre position. Ils savent en distinguer les informations les plus importantes pour s'intéresser à l'intention dans laquelle l'auteur les présente et à la manière dont il procède pour les transmettre. Ces élèves, qui obtiennent des scores très supérieurs à la moyenne à PISA, sont pour 53 % d'entre eux des filles, et on n'est guère surpris de constater que plus de 90 % sont scolarisés au lycée. L'étude des réponses de ce type d'élèves montre qu'ils fournissent de façon quasi systématique les réponses attendues aux épreuves de PISA, et que cette performance résulte conjointement de la mobilisation de savoirs scolaires, d'une culture générale ou d'une expérience personnelle, et d'un traitement des documents qui repose sur de réelles compétences en littéracie, elles-mêmes adossées sur des compétences cognitives assurées. Mais elle résulte aussi de la mobilisation, tout le long du cahier, d'une attitude de reconfiguration des différentes sources mobilisées, leur permettant de se situer dans le registre des raisonnements et des savoirs requis par l'évaluation. Il s'agit bien ici de la mobilisation de soi, mais de soi comme sujet d'un écrit, d'une pensée, d'un travail, qui ne sont justement pas réductibles à soi. Ces élèves se caractérisent par une mobilisation d'eux-mêmes comme sujets de leur réflexion et de leur texte.

Alors que ces trois classes ou trois profils d'élèves apparaissent relativement stables, les deux autres classes constituées à partir de nos analyses statistiques pour chacun des deux cahiers, le sont notablement moins. Ces quatre classes regroupent plus de 42 % des élèves auxquels ont été soumis l'un ou l'autre des deux cahiers. Elles regroupent des élèves que nous avons qualifiés de « préleveurs d'information » et de « constructeurs de sens gênés par un être au monde conformiste », pour le premier cahier (respectivement 21 % et 18,8 % du *corpus*) ; de « lecteurs se limitant à l'explicite » et d'élèves

« minimalistes » pour le second (respectivement 20,8 % et 24,4 % du *corpus*). Les élèves appartenant à ces quatre classes obtiennent des scores de réussite moyens ou supérieurs à la moyenne à PISA ; ils sont pour 53,6 % d'entre eux des filles, et fréquentent pour près de 62 % une classe de lycée. Leurs réponses sont beaucoup moins homogènes que celles des élèves appartenant aux catégories présentées ci-dessus, attestant des compétences (ou lacunes) et des postures beaucoup moins stabilisées ; elles semblent plus liées à des modes de faire et des processus de mobilisation différents selon les textes et les questions, plus dépendantes de leurs thématiques et de leurs formats (QCM ou questions ouvertes appelant une réponse écrite relativement élaborée), et de la nature du travail qui y était requis (prélèvement d'informations explicites dans le texte, compréhension globale, commentaire et analyse des arguments ou du style). Les « préleveurs d'information » obtiennent une assez bonne réussite aux tâches de lecture les plus simples : prélever une information, la reconnaître lorsqu'elle est paraphrasée dans un QCM, donner son avis. En revanche, outre d'importantes lacunes syntaxiques et orthographiques, ils éprouvent une difficulté à raisonner à partir d'un texte, à en hiérarchiser les informations. Les textes semblent leur être transparents, et le travail consistant à les constituer comme objet de réflexion, d'analyse et de commentaire semble leur être très peu familier. Certains traits les rapprochent des « lecteurs se limitant à l'explicite », élèves dont les réponses au second cahier montrent qu'ils travaillent essentiellement avec le texte qu'ils sont capables de comprendre lorsque toutes les données nécessaires y figurent, mais qu'ils peinent à mobiliser autre chose que son contenu pour procéder aux inférences nécessaires. Prudence et rigueur ? Ou ignorance de la fréquente nécessité, pour comprendre un texte, d'en rattacher le contenu explicite à des connaissances extérieures ? Les élèves « minimalistes » ont un profil plus hétérogène encore. Des compétences, cognitives et linguistiques, sont acquises – beaucoup des questions de PISA sont d'ailleurs significativement mieux réussies que par les autres élèves – mais peu stables, peu assurées. Les postures « secondes » sont tantôt présentes, tantôt absentes. Les performances de ces élèves s'arrêtent là où ils auraient besoin d'utiliser le langage écrit comme un outil cognitif et sémiotique leur permettant de réfléchir plus avant. Enfin notre dernière catégorie regroupe des élèves engagés dans leur travail, plus présents que la moyenne dans les textes qu'ils rédigent, aptes à commenter les textes, à raisonner pour comprendre, mais dont la référence

à une *doxa* conformiste gêne souvent la réflexion. Si leurs réponses sont souvent plus longues que celles des autres élèves, elles n'en sont pas pour autant pertinentes : ils « glosent » en ce cas avec eux-mêmes en interprétant le texte. Tout se passe comme s'ils avaient envie de s'exprimer sur certains sujets ; de ce fait leurs réponses sont hétérogènes sur les différents registres travaillés par nos analyses, puisqu'elles dépendent de leur intérêt pour les thèmes des différents textes. Leurs argumentations sont principalement fondées sur des valeurs morales de justice et de vérité. Un tel engagement d'eux-mêmes peut dans certains items leur permettre de produire des bonnes réponses, pour des raisons de valeurs partagées ; il peut néanmoins se traduire par des affirmations erronées ou répondant à côté de la question posée, lorsque ce point de vue *doxique* ne permet pas de saisir la visée du texte ou de la question, ou par des non réponses lorsqu'il ne permet pas de construire un raisonnement. Les élèves de cette classe sont sans doute ceux qui présentent les performances les plus instables.

Pour conclure notre analyse quantitative, nous avons étudié systématiquement quelles étaient les relations entre les scores obtenus à PISA et les occurrences de chacun des indices de notre grille de codage ; cette étude, ainsi que celle des occurrences conjointes des indices qui caractérisent les classes d'élèves créées par l'analyse hiérarchique, permettent d'examiner quels sont les indices les plus différenciateurs entre élèves. Ce sont les indices qui renvoient à des compétences cognitives (indices du début de notre grille) qui ont le lien statistique le plus important avec les scores de réussite, parce qu'ils sont les plus fréquents. En revanche, ceux dont le nombre d'occurrences entre « bons » et « mauvais » élèves est le plus contrasté sont des indices moins fréquents (11), qui renvoient à ce que nous avons appelé processus ou travail de secondarisation des savoirs, de l'expérience et des discours. Ce sont les indices L3, L5, T3 et E1 de notre grille, attestant la capacité de prendre un recul réflexif par rapport aux textes, de constituer leurs aspects linguistiques et rhétoriques comme objet de réflexion, de produire un mouvement de « problématisation » dans son propre discours, ou encore de prendre le monde comme objet de connaissance et d'analyse. En revanche, contrairement à nos hypothèses, la mobilisation de différentes *doxas*, juvéniles ou sociales, ou celle de l'expérience personnelle, s'observent dans les différentes classes et ne permettent pas de différencier les élèves. Il semble bien qu'elle constitue une aide pour les bons élèves, qui font feu de tout bois pour

réfléchir sur les textes et les questions de PISA, et qu'elles puissent, comme nous l'avons vu, être soit une aide, soit un obstacle pour les élèves dont les compétences et les postures sont beaucoup moins assurées et stabilisées.

PERFORMANCES ET UNIVERS MOBILISÉS : QUATRE PROFILS DIFFÉRENCIÉS D'ÉLÈVES

Nos analyses portant sur les cahiers remplis par les élèves ayant réellement passé les épreuves de PISA 2000 ont été complétées par celles que nous avons pu effectuer à partir des observations et des entretiens recueillis, deux ans après l'enquête de l'OCDE, auprès d'élèves auxquels nous avons soumis le cahier de notre composition comprenant six des épreuves de l'enquête originale, et à propos desquels nous avons recueilli divers renseignements d'ordre scolaire, social ou familial. Nous présenterons ici nos analyses portant sur les modes de faire et les propos de quatre de ces élèves, tous élèves de collèges ou de lycées de Seine-Saint-Denis, choisis parce qu'ils nous semblent proches de quatre des classes que nos analyses statistiques nous ont permis de décrire.

Mehmet : faibles performances et rapport d'adhésion au texte et au monde

Ainsi Mehmet, élève de 3^e d'un collège ZEP particulièrement « difficile » est-il très représentatif des élèves que nous avons dits « en échec généralisé ». Quatrième d'une fratrie de cinq enfants, il est né en France de parents turcs. Son père est retraité et sa mère sans profession. Admis en 4^e « à l'ancienneté » (12) en fin de 5^e, il a toujours éprouvé de grandes difficultés depuis son entrée (avec un an de retard) au collège.

Mehmet semble éprouver d'importantes difficultés à comprendre le sens des différents textes sur lesquels nous l'interrogeons et lui demandons de réfléchir, ce indépendamment du degré d'intérêt qu'il dit avoir pour leur contenu ou pour les thématiques qui y sont évoquées. Mais ces difficultés ne le conduisent qu'une seule fois à ne pas répondre à la question posée. Elles le conduisent plus fréquemment à répondre soit *au pif* (13) (par exemple aux questions portant sur les différents avis, favorable ou défavorable au nettoyage mécanique des plages, auxquelles il a, ce faisant, une chance sur deux de répondre

« juste »), soit en ne prélevant et ne traitant qu'un élément de signification locale, indépendamment de la signification globale du texte dans lequel il le préleve. Ce dernier mode de faire le conduit parfois à proposer des réponses aberrantes, que l'on ne peut comprendre qu'en repérant qu'il y mêle, sans souci de cohérence ni syntaxique ni sémantique, diverses expressions prélevées dans la phrase du texte de référence qui est sémantiquement la plus proche d'un des mots dont il est fait usage dans le libellé de la question. La difficulté à comprendre le texte, maximale quand celui-ci s'oppose au sens commun, peut également le conduire à se rabattre sur ce sens commun, comme il le fait quand on l'interroge sur la pertinence du titre de l'article sur le nettoyage des plages : *j'avais du mal à comprendre [...] Y a plus de mots que je comprends pas que de mots que je comprends. Et par rapport à ça, ben... j'ai répondu que c'est mieux de la garder propre.* Il ne parvient pas à adopter une posture « secondarisée » qui lui permettrait de prendre en considération le titre d'un texte ou les argumentaires qui y sont développés sans le(s) rabattre sur le contenu (ou sur un élément très local de contenu) de ce texte et sur son opinion, voire sur son accord avec tel contenu, sur son adhésion à telle proposition donnée comme « vraie » dans ce qui relève plus d'un rapport de soumission au monde ou à autrui que d'une posture de questionnement de ceux-ci. Ainsi justifie-t-il presque toujours son accord avec tel texte ou tel auteur en disant que ce qu'ils disent est *vrai*, sans que ce statut de « vérité » ait jamais à se justifier ou à se référer à un critère quelconque.

Un tel rapport d'évidence et d'adhésion au « vrai » du monde et du langage se conjugue chez Mehmet, durant l'entretien, avec une hyper-dépendance à l'égard de l'adulte, qui le conduit fréquemment à interpréter les questions posées par celui-ci comme invalidation de ses propres réponses et, dès lors, à abandonner son point de vue pour adopter celui qu'il pense pouvoir déceler dans les propos de l'enquêteur, ou qu'il pense que celui-ci attend de lui : *si vous dites que c'est une autre réponse*, dit-il ainsi plusieurs fois alors que l'enquêteur n'a explicitement rien dit de tel. Il va même, alors que l'échange porte sur les raisons pour lesquelles il se dit en accord avec tel avis plutôt qu'avec tel autre (à propos de la conquête spatiale) jusqu'à sembler penser que l'enquêteur puisse mieux savoir que lui avec quel texte il s'accorde le mieux : *je me suis basé sur ça, mais peut-être que c'est une autre personne... enfin, le plus d'accord.* Il faut certes voir là un effet du cadre de l'entretien, manifestement perçu comme insécurisant. Mais, au-delà de cela,

c'est l'ensemble des propos de Mehmet, par exemple quand il évoque ses projets d'avenir ou l'orientation en fin de classe de 3^e, qui témoigne d'un tel rapport de soumission à des adultes dont les rôles institutionnels (principal, professeur principal, CPE, conseiller d'orientation) sont le plus fréquemment indifférenciés et subsumés sous le seul pronom personnel *ils*. Rapports aux textes et à la littéracie, et rapports au monde sont étroitement liés ; indissociable d'une position dominée au sein des rapports sociaux, ce lien empêche de considérer les performances de Mehmet aux épreuves de PISA que nous lui avons soumises (comme plus généralement aux épreuves scolaires) comme relevant de seules compétences cognitives de compréhension de l'écrit.

Hachem : quand trop de sens tue le sens

Hachem, élève de 3^e, n'a pas totalement le profil des élèves que nous avons qualifiés de « constructeurs de sens gênés par leur conformisme » : il ne glose pas comme eux, tout au moins pas à l'écrit, pour se saisir du texte comme prétexte et y affirmer sa propre présence. Comme eux, cependant, il ne peut s'empêcher, sur de nombreux items, d'oublier ses capacités de compréhension rationnelle et de regarder les lignes qui lui sont présentées comme des appels à projeter sa propre vision du monde. Il manque certes d'un vocabulaire dont la possession lui faciliterait vraisemblablement l'accès à sens des textes proposés. Il ignore par exemple ce qu'est un « préjugé » ou un « riverain », dit ne pas comprendre ce que signifie « négligeable » ou « une idée maîtresse ». Il présente aussi des lacunes par rapport à des connaissances scolaires qu'il aurait vraisemblablement dû intégrer (il confond par exemple « l'érosion côtière » avec la montée des eaux due au réchauffement de la planète). Il a également du mal à imaginer des situations qu'il n'a jamais connues, comme celle d'un échange linguistique (*avec Vigi-Pirate, y a plus de voyages*). Mais ces incompétences ne paraissent cependant pas responsables des dérapages de sens qu'il commet sur certains sujets alors qu'il se débrouille relativement bien sur d'autres qui ne semblent pas présenter de difficultés cognitives inférieures.

Paradoxalement, c'est souvent le trop de sens qui le fait se tromper parce qu'il abandonne la tiédeur émotionnelle requise par le traitement textuel alors qu'il se montre capable d'en faire preuve sur des thèmes auxquels il est plus indifférent. L'entretien post-passation réalisé avec lui donne quelques clés (sans

doute trop grossières vu sa brièveté) d'accès à ces moments où les contextes des textes le renvoient à ceux de son environnement quotidien et compliquent singulièrement l'adoption d'une posture objective. D'origine algérienne, Hachem déclare des difficultés en Français. Il ambitionne une 2^{de} générale car il voudrait travailler dans l'aéronautique comme pilote. Il hésite parfois sur le pays dans lequel il préférerait vivre, mais il pense qu'il a plus de chance que ses cousins restés en Algérie : *c'est normal, il y a des difficultés là-bas. C'est pas facile pour travailler, pour apprendre*. Sa tension entre deux pays, deux cultures, paraît forte, mais il aspire à une intégration dans son pays d'adoption. Les items qui peuvent mettre en tension des positions conformistes et d'autres plus centrées sur l'expression des personnes le mettent alors particulièrement mal à l'aise.

C'est notamment le cas d'une épreuve qui présente sous forme de deux textes d'une vingtaine de lignes chacun, deux argumentaires très contrastés à propos des graffiti. Dans le premier, une Helga affirme sa rage de voir les graffiti salir les murs des villes et stigmatise les « artistes criminels » qui dégradent l'environnement et ternissent la réputation des jeunes. Dans le second, une Sophie met en valeur la créativité des auteurs de graffiti, inventeurs de formes, et déplore que ceux qui les dénigrent acceptent sans mot dire que la publicité envahisse nos rues. Hachem entre partiellement dans l'argumentation de Helga. Prié de trouver dans le texte d'autres préjugés signalés par elle, il ne voit que celui du coût social lié à l'effaçage toujours recommencé (« Ils travaillent, ils se fatiguent pour faire ça et les autres ils viennent le gâcher »). Il n'évoque pas celui de la destruction de la couche d'ozone par les gaz propulseurs alors que la discussion avec lui fait apparaître qu'il connaît très bien ce phénomène. Hachem est apparemment trop centré sur sa propre expérience de collégien issu de l'immigration et en quête forte d'intégration. Il se félicite de travailler dans un établissement en bon état, qui n'est pas comme ceux *qu'on voit à la télé où y a tout le temps des problèmes*. Cette posture hyper-morale déclarée ne l'aide pas à entrer dans la totalité de l'argument du premier texte et lui interdit l'accès au second. Il ne comprend pas la tentative de Sophie de considérer comme des signes urbains semblables les graffiti et la publicité, d'affirmer l'influence des premiers sur la seconde, de faire de leur situation aux yeux de la loi une différence secondaire. Pour lui, *les graffiti, ça représente rien* ; pire, *c'est écrire son nom avec une écriture bizarre pour montrer que la personne qui a écrit ça, elle est forte ou elle connaît des gens*.

La discussion post-passation ne le fera pas changer d'avis (*j'ai rien compris avec la publicité de Sophie là [...] De toute façon, c'est sale*). Pour lui, le débat sur le bien commun porte essentiellement sur la légalité. On voit bien que *les grands peintres et tout, ils font des cadres, des peintures comme ça, mais ils vont pas gâcher quelque chose qu'on a mis du temps pour le faire*. D'autres questions du même cahier, fondées sur la tension entre expressivité et intégration sociale, suscitent chez lui des célicités cognitives analogues, lesquelles, associées à la posture hyper-morale qui est la sienne, semblent lui interdire ou du moins lui rendre très difficile toute labilité de jugement.

Cindy : chercher les réponses aux questions ou réfléchir sur les textes ?

Cindy, élève de 2^{de} d'un lycée de la périphérie parisienne, est assez représentative du groupe d'élèves « préleveurs d'informations », auxquels leurs compétences « scolaires » permettent de repérer des informations dans un texte et de fournir des bonnes réponses aux questions du type « s'informer », mais qui ont cependant bien des difficultés à raisonner sur l'ensemble du texte. Cindy paraît très consciente de cette faiblesse (*moi, mon défaut c'est que je lis pas le texte*). Elle dit se jeter sur les questions et rechercher ensuite les réponses dans le texte qu'elle lit à cette occasion. Alors qu'elle dit savoir qu'en toute rigueur il faut avoir compris son sens global pour être capable ensuite de répondre à des questions particulières, Cindy, apparemment obsédée par la perspective évaluative, essaie de réfléchir sur le texte à partir d'un point de vue qui n'est ni celui de l'auteur, ni le sien propre, mais celui, supposé, d'un correcteur. Elle voit bien que cela ne marche pas, mais ne peut adopter une autre perspective : *je me jette direct sur les questions, j'ai envie de répondre, j'sais pas, j'sais même pas pourquoi j'fais comme ça*.

Fille d'un père employé à la RATP et d'une mère employée de pharmacie, Cindy qui avait eu d'assez bons résultats lors de sa scolarité au collège, commence à perdre pied en 2^{de}, en Français notamment. Dès que l'exercice est moins guidé, fait appel à des ressources personnelles pour prendre à son compte les données et les problèmes et les organiser, la méthode « localiste » qui lui avait jusque-là réussi montre ses limites. C'est pourquoi elle apprécie particulièrement les items de PISA qui posent des questions précises sur des éléments nettement distingués d'un texte. Il en va ainsi de celui qui donne une liste de quatorze conseils à des élèves préparant

un échange linguistique : *Déjà, on nous donne des petites informations, on nous dit « lisez le point 13 du guide, le point 1 », donc on nous donne précisément la phrase qui correspond à la question*.

Une autre aubaine consiste à tomber sur des textes qui affirment des positions qu'elle pense partager. C'est le cas avec un des conseils du guide d'échange qui incite à rester soi-même quand on séjourne à l'étranger. Cindy l'identifie très bien : *On a le droit d'être différent et chacun est comme il est. Et heureusement qu'il y a de la différence, sinon on serait tous pareils*. Ici la conviction d'être en communion avec l'esprit de l'exercice lui permet d'accéder à une structure de compréhension qu'elle n'a pas ou ne s'autorise pas d'habitude à avoir.

Elle peine davantage en revanche sur l'épreuve portant sur le nettoyage des plages, analysée ci-dessus. D'une part à cause de la multiplicité des intervenants et de leurs prises de positions sur un sujet qu'apparemment elle connaît assez peu et qui lui interdit de constituer le texte comme une totalité non-additive. D'autre part parce qu'elle ne parvient pas à entrer dans certaines positions écologiques exprimées, qui considèrent qu'il vaut mieux, sous un certain rapport, ne pas nettoyer les plages : *je vais pas aller à la plage quand elle est sale*. Peu à l'aise dans cet ensemble d'arguments qui, pour beaucoup, lui paraissent incongrus, elle dit *souligner tous les sujets* dans l'espoir de répondre plus facilement aux questions dont elle n'a pas encore pris connaissance. Elle a donc perdu un temps considérable et s'est révélée assez peu capable de mobiliser les éléments pertinents du texte lorsqu'on lui a demandé de le faire. Ce n'est pourtant pas faute d'engagement dans la tâche, car, en raison même des scrupules qui la font sans cesse revenir à des relectures partielles du texte, Cindy n'a pu répondre dans les temps aux derniers items du cahier.

Pour des raisons vraisemblablement proches, elle attribue la colère de villageois furieux contre le cinématographe qu'ils découvrent, à l'état de tristesse dans lequel les aurait plongés un film alors que l'argumentation d'ensemble vise à montrer leur incompréhension du caractère fictionnel des images projetées. À propos d'une autre épreuve concernant un plan d'aide humanitaire, elle ne parvient pas à faire d'hypothèse sur le caractère paradoxal d'absence de secours à un des pays les plus pauvres de la planète : *L'Éthiopie ? On n'en a jamais parlé en géographie, non*. Tout se passe comme si, en l'absence de référence scolaire bien repérable, elle ne pouvait valider la double nécessité, pour comprendre un texte, de le

constituer comme un objet et de se saisir soi-même comme sujet face à lui. En s'inscrivant dans un genre évaluatif, PISA suscite vraisemblablement des craintes et des stratégies de réponse bien connues. Mais, par-delà le stress de l'épreuve, cette enquête, en s'affranchissant souvent de la manière scolaire, fait appel à une autonomie intellectuelle qui signe précisément le passage de 3^e en 2^{de} et que des élèves comme Cindy ne sont manifestement pas parvenus à construire en s'acquittant scrupuleusement de leurs tâches de collégiens.

Hawa : une réflexivité à toute épreuve

Hawa entre incontestablement dans la catégorie de bons lecteurs et bons compreneurs que nos analyses précédentes nous ont permis de construire, et manifeste constamment, tant dans son travail sur les épreuves proposées que durant l'entretien, une posture « seconde » de mobilisation cognitive et de réflexion sur les textes et les épreuves, sur le monde et sur elle-même. Excellente élève de 2^{de} dans le même lycée que Cindy, elle devait être admise en 1^{re}S, selon ses vœux, sans aucun problème en fin d'année. Aînée de trois filles, née au Congo, de parents congolais, elle est venue en France à deux ans et demi. Elle dit son père ingénieur (le dossier indique agent technique) et sa mère employée dans l'hôtellerie. Elle-même envisage de devenir ingénieur, après avoir pensé devenir médecin, mais elle dit vouloir *laisser aller le temps, se laisser mûrir, pour décider le moment venu quel métier je ferai plus tard*. Grande lectrice, elle dit avoir également une *passion* pour l'écriture et avoir ainsi rédigé *des histoires qui sortent de ma tête*, dont l'une a fini en roman d'une centaine de pages. Elle décrit le changement du collège au lycée en termes de *nouvelle indépendance, qui va avec notre capacité de s'exprimer* et nous dit participer à un projet de création d'un journal du lycée.

Quelle que soit la question posée, Hawa se saisit de pratiquement toutes les thématiques évoquées par les épreuves ou durant l'entretien pour les mettre à distance et les constituer en objets de réflexion, porteurs d'enseignements et de considérations génériques, et pour s'affirmer elle-même comme auteur de cette réflexion et de ces considérations, voire pour entrer en dialogue, en débat, avec les auteurs des textes soumis à sa réflexion. Les épreuves évoquées durant l'entretien, leurs thématiques ou les exemples suggérés par ses propres pratiques ou intérêts sont très rapidement l'objet d'une montée en géné-

ralité qui les situe et les commente au regard tant de principe généraux que d'une réflexion sur soi. Ainsi à propos de l'épreuve évoquant les graffiti : *Le texte de Sophie, j'ai compris ses arguments, mais ça suit un peu la ligne de conduite que je me suis mise à moi-même [...] On se dit bon, la publicité elle rapporte, et les graffiti ne rapportent pas. Dans cette société, ce qui est en première ligne, ce sont les choses qui rapportent. Comme si on compare ça à la cigarette, on nous dit partout que c'est mauvais pour la santé mais on continue à les vendre parce que ça a des effets à long terme, pas comme la drogue, et ça rapporte beaucoup à l'État. Il en est de même de son expérience du cinéma, évoquée à propos d'une autre épreuve : *J'aime bien les films qui se projettent dans le futur comme par exemple Matrix [...] J'aime bien aussi voir des films de la vie réelle, pour ne pas trop rêver. Parce qu'on aime rêver mais il faut pas oublier qu'on a les deux pieds sur Terre, qu'on marche sur du béton et qu'il peut y arriver n'importe quoi* (ces propos suivent immédiatement une évocation du film *Elephant*, de Gus van Sant, qu'Hawa souhaite aller voir prochainement).*

Les thématiques proposées à sa réflexion rencontrent plus ou moins son intérêt, et elle y accorde plus ou moins d'attention, selon qu'elle peut y reconnaître et y mobiliser des éléments de son expérience personnelle et juvénile. Toutefois, son manque d'intérêt et ce qu'elle nous décrit comme un survol du texte sur le nettoyage des plages, ne l'ont pas gênée pour en réussir toutes les questions sauf une, la plus difficile. À l'inverse, son intérêt affirmé pour les thématiques évoquées par d'autres épreuves, conjugué à la posture réflexive qu'elle manifeste à toute occasion, la conduit parfois à s'intéresser et à réfléchir au problème soulevé par tel texte ou telle épreuve, plus qu'au traitement des questions précises qui lui sont posées à leur sujet. Cela ne pose parfois aucun problème, quand les propos par lesquels elle commente textes ou questions viennent compléter des réponses tout à fait correctes à celles-ci. En revanche, son accord avec l'un des textes proposés portant sur la recherche spatiale la conduit à se tromper à une des épreuves demandant de dire quel était, parmi cinq textes, celui qui était le plus contradictoire avec l'un d'entre eux. Mais elle ne se contente pas de redire l'argument du texte choisi ; elle le reformule, le relativise et s'en saisit pour réfléchir elle-même, même si elle réfléchit moins sur la recherche spatiale que sur le désordre du monde : *Il est vrai que l'exploration et l'exploitation de l'espace font réfléchir. Mais je pense qu'il serait préférable de s'occuper des problèmes actuels. Et les milliards dépensés pour la recherche*

spatiale ne sont pas les seuls barrages à la résolution de ces problèmes. De plus, elle relativise ce point de vue en prenant en considération celui d'un autre auteur, puis en engageant une discussion virtuelle avec lui : *Félix, il a aussi raison quand il dit qu'il faut pas s'occuper simplement de notre génération mais aussi des générations futures. Cet argument, je le reprends. J'aimerais lui dire que si des jeunes gens n'existent pas aujourd'hui, ben y aura pas de demain. Il faut qu'il y ait un aujourd'hui pour qu'il y ait un demain [...] Donc faudrait plutôt penser à améliorer la condition de vie de maintenant avant de penser à demain.*

Cette discussion virtuelle, dans laquelle Hawa reconnaît le caractère légitime d'un argumentaire qu'elle ne partage pas, atteste une réelle capacité de labilité dans le jugement, dont elle fait également preuve à propos des graffiti : *j'apprécie seulement quand c'est joli à regarder. Mais quand c'est pour des injures ou par exemple dans le métro, quand on voit des vitres sur lesquelles ils ont marqué des choses, là c'est vrai que c'est pas très joli à voir, et c'est un non respect des gens.* Mais c'est à partir de son propre rapport au langage et de ses propres engagements qu'elle juge de la lettre la mieux écrite : « D'après moi, Sophie a écrit la meilleure lettre. Son ton, ses arguments frappent. Le fait de poser des questions aux lecteurs tel qu'elle le fait donne un certain engagement dans ses propos », réponse qu'elle commente oralement, juste avant de nous dire qu'elle-même écrit beaucoup, en des propos manifestant sans doute qu'elle se reconnaît dans cette manière d'écrire attribuée à Sophie : *c'est beaucoup plus convaincant, parce qu'on a l'impression qu'elle est devant nous et qu'elle nous les crie, ses questions.*

Dans l'ensemble des épreuves de PISA que nous lui avons soumises, Hawa se montre parfaitement à même de mobiliser des connaissances scolaires et culturelles pertinentes pour comprendre les textes et reformuler leurs thématiques, mais aussi pour se saisir de leur caractère paradoxal ou pour formuler des hypothèses explicatives à leur propos. Ainsi peut-elle parler de *pays qui sont en développement ou sous le seuil de pauvreté*, en utilisant une terminologie qui n'existe ni dans le texte ni dans les questions portant sur l'aide humanitaire, avant de formuler l'hypothèse selon laquelle une guerre en Éthiopie pourrait expliquer le faible taux d'aide accordée à ce pays. De même, explique-t-elle l'attitude de villageois sud-américains découvrant le cinéma en écrivant qu'ils accordent trop d'importance à ce qui n'est que fiction, avant de faire, dans son commentaire oral spon-

tané de ce texte, l'hypothèse selon laquelle *c'est sûrement la première fois qu'ils vont dans un cinéma, la fiction ne leur saute pas directement aux yeux.*

CONCLUSION

Les analyses qui précèdent confirment l'essentiel de nos hypothèses sur le caractère hétérogène des modes de travail et des univers de référence mobilisés par les élèves évalués par PISA, sur l'impossibilité de les réduire à de simples compétences de lecture et de compréhension de l'écrit. Elles renforcent l'hypothèse selon laquelle les épreuves de l'enquête PISA, que ses concepteurs présentent comme évaluant ce type de compétences évaluent des processus plus complexes et hétérogènes que cela. Non seulement parce qu'une part importante des items qui les constituent requiert certes des élèves une bonne compréhension des textes et questions qui leur sont soumis, mais aussi des compétences de production d'écrit fondées sur des processus de secondarisation et de reconfiguration. Mais également et surtout parce que ces épreuves requièrent d'eux, certes d'avoir construit et de pouvoir mobiliser des compétences de compréhension et de production d'écrits, mais aussi de pouvoir adopter et mettre en œuvre un travail leur permettant de « savoir » (savoir en actes bien plutôt que savoir explicite) ce qu'il est nécessaire et pertinent de mobiliser, de tenir à l'écart ou de reconfigurer de chacun des registres de référence potentiellement sollicités et de leurs modes de faire avec ces registres et avec les tâches scolaires.

L'un de nos résultats les plus tangibles réside donc dans l'hétérogénéité et la diversité ainsi mises au jour. Celles-ci sont repérables non seulement dans les comparaisons que nous avons pu faire entre élèves ou entre classes d'élèves créées par le traitement statistique, mais aussi, chez un même élève, d'une épreuve à l'autre, en fonction des thèmes concernés et de la nature des tâches attendues. Une telle constatation, qui concerne une proportion importante d'élèves, interroge non plus seulement la conception de la littéracie qui sous-tend l'enquête PISA, mais aussi le projet ou la possibilité d'évaluer des compétences de compréhension de l'écrit comme si celle-ci étaient stables et/ou mobilisées de manière systématique pour une classe de tâches considérées comme équivalentes par les concepteurs des épreuves, alors que l'étude des modes de faire des élèves montre que, pour une part importante d'entre eux, ces modes

de faire dépendent et varient en fonction des textes et contextes, des thématiques et des formats de tâches, autant voire plus qu'en fonction de leurs seules compétences de traitement des textes et de l'écrit. Ce qui rend sans doute plus difficile et complexe la tâche de prévoir les réussites ou les échecs des élèves aux enquêtes de ce type, mais aussi et surtout de penser les remédiations possibles, sur les plans institutionnels, curriculaires et pédagogiques.

Les analyses qui précèdent ne disqualifient bien évidemment pas tout projet d'enquête internationale visant à évaluer les acquis des élèves. Nous espérons seulement qu'elles puissent contribuer à faire que les enquêtes à venir, en permettant de mieux saisir ce qui, dans différents contextes, peut être une aide ou un obstacle au travail des évalués, faciliter ou gêner leur réussite des tâches de lecture, évitent ou minorent les effets de palmarès, permettent une analyse plus détaillée des similitudes et différences entre

pays, et accroissent les capacités de chaque pays à mieux diagnostiquer ses propres difficultés éducatives et à mieux y faire face.

Élisabeth Bautier
bautier@wanadoo.fr
université Paris 8
ESCOL

Jacques Crinon
jacques.crinon@creteil.iufm.fr
IUFM de Créteil
ESCOL

Patrick Rayou
rayou@noos.fr
IUFM de Créteil
ESCOL

Jean-Yves Rochex
rochexjy@wanadoo.fr
Université Paris 8
ESCOL

NOTES

- (1) En France, les élèves concernés pouvaient être scolarisés soit en collège (classes de quatrième ou de troisième), soit en lycée général ou technologique (classes de seconde ou de première), soit en lycée professionnel.
- (2) Ainsi, pour D. Bain, « L'absence d'une théorie élaborée, organisant les différentes dimensions de la compréhension de l'écrit en tant qu'activité langagière et proposant un point de vue déterminé sur les opérations de lecture, constitue [...] un premier obstacle majeur pour l'exploitation didactique des résultats [de PISA 2000] », un peu comme si l'on pouvait prêter aux concepteurs de PISA le même type de boutade autrefois attribué à Binet : « la littérature, c'est ce que mesure notre test ». De même, N. Bottani et P. Vignaud écrivent-ils, à propos des enquêtes internationales et particulièrement de l'enquête PISA, que « l'approche unidimensionnelle, reposant sur un score principal, peut être critiquée car elle réduit les comparaisons à un classement et, par ailleurs, ce caractère unidimensionnel des données ne semble pas complètement défendable. Il faut également souligner les questions liées à la signification des variables mesurées comme "les compétences pour vivre et travailler dans le monde moderne". Ces variables semblent être davantage justifiées par leurs qualités psychométriques que par des construits théoriques en relation avec les champs scientifiques pertinents (psychologie, sociologie, économie, sciences de l'éducation) ».
- (3) Cf. également sur ces questions, le vif débat entre Goldstein (1995) et Masters (1995) à propos des enquêtes réalisées par l'*International Association for the Evaluation of Educational Achievement*.
- (4) Il est par exemple difficile d'admettre que la conquête de l'espace puisse être un objectif primordial pour des jeunes très sensibilisés à la question de la faim dans le monde, ou qu'une posture de défense de l'environnement puisse conduire des riverains à s'opposer au nettoyage mécanique des plages pour des jeunes désireux de trouver la plage propre quand ils s'y rendent.
- (5) Le concept de *secondarisation* trouve son origine dans la distinction que fait Bakhtine (1984), dans le champ de la production littéraire, entre genres (de discours) premiers et genres seconds. Les genres premiers peuvent être décrits comme relevant d'une production spontanée, immédiate, liée au contexte qui la suscite et n'existant que par lui, dans « l'oubli » d'un quelconque appren-

tissage ou travail sous-jacent. Naissant de l'échange verbal spontané, ils sont fortement liés à l'expérience personnelle du sujet. Les genres seconds, fondés sur les premiers, se ressassent de ceux-ci, les travaillent et les reconfigurent dans une finalité qui s'émancipe de la conjoncturalité de leur production. Ils supposent une production discursive qui est à elle-même sa propre finalité et son propre objet, et qui signifie bien au-delà de l'interaction dans laquelle elle peut conjoncturellement se situer (sur ce point, cf. Schneuwly, 1994 ; Bautier & Rochex, 2004 ; Bautier, 2004).

- (6) L'enquête PISA 2000 a été conduite selon la technique des « cahiers tournants ». Tous les items de l'enquête ont été répartis dans neuf cahiers d'évaluation différents comportant néanmoins des blocs d'items communs d'un cahier à l'autre. Ces neuf cahiers ont été distribués aléatoirement aux élèves testés. Ce procédé a ainsi permis de tester l'équivalent de sept heures d'évaluation, tout en ne faisant travailler chaque élève que durant deux heures. Les présupposés de l'enquête conduisaient à considérer les neuf cahiers comme équivalents, ce qui n'est pas le cas au regard de nos hypothèses de recherche et des analyses qu'elles nous ont permis de conduire.
- (7) Soit un total de 842 cahiers. Dans la mesure où nos analyses ont fait apparaître que les deux cahiers étudiés ne convoquent pas exactement les mêmes façons de faire des élèves, il n'a pas été possible de les traiter conjointement. Au demeurant, la quantité de chacun des cahiers étant très proche (415 et 427), il a été aisément possible d'en comparer les résultats quantitatifs.
- (8) Ces catégories ont été testées une première fois à partir d'une analyse d'un corpus réduit de plusieurs dizaines de cahiers, ce qui a permis de les préciser et de les affiner, pour les mettre ensuite en œuvre sur l'ensemble des 842 cahiers. Voir en annexe notre grille d'analyse et les types d'indices que nous avons élaborés.
- (9) Ces six épreuves étaient les suivantes : *Graffiti* (items R081Q01, R081Q02, R081Q05, R081Q06A et R081Q06B), *Guide de conduite* (items R228Q02, R228Q03 et R228Q04), *Macondo* (items R061Q01, R061Q03, R061Q04 et R061Q05), *Plage* (R070Q07, R070Q03, R070Q04 et R070Q02), *Plan international* (R099Q02, R099Q04A, R099Q04B et R099Q03) et *Avis d'élèves* (R120Q01, R120Q03 et R120Q06). Trois de ces six épreuves

étaient communes aux deux cahiers sur lesquels ont porté nos analyses quantitatives.

- (10) La liaison entre les compétences orthographiques et les compétences rédactionnelles est attestée dans la littérature. Deux interprétations complémentaires sont en général proposées : l'automatisation de l'orthographe libère les ressources cognitives des sujets ; elle leur permet en outre de trouver plus d'intérêt aux tâches.
- (11) Leur moindre fréquence est due au fait que les épreuves de PISA, élaborées à partir d'une autre conception de la littéracie que celle qui a présidé à nos analyses secondaires, n'étaient pas conçues pour mettre en évidence les modes de faire et les

postures que ces indices visent à appréhender. Le fait qu'on les voie apparaître n'en est que plus révélateur de leur importance dans les processus de différenciation entre élèves.

- (12) Cette expression figure en toutes lettres sur le bulletin du conseil de classe du dernier trimestre de sa classe de 5^e, et les bulletins des classes suivantes ne montreront aucune amélioration.
- (13) Dans tout ce qui suit, les citations des propos tenus oralement par les élèves figureront en caractères italiques, alors que seront citées en caractères romains et entre guillemets français leurs réponses écrites (celles-ci sont citées sans modifier leur structure syntaxique, mais en rétablissant l'orthographe, pour des raisons de lisibilité).

BIBLIOGRAPHIE

- ADAMS R. & WU M. [éd.] (2002). *PISA 2000 Technical Report*. Paris : OCDE.
- BAIN D. (2003). « PISA et la lecture : un point de vue de didacticien ». *Revue suisse des sciences de l'éducation*, vol. 25, n° 1, p. 59-78.
- BAUTIER É. (2005). « Formes et activités scolaires, secondarisation, reconfiguration, différenciation sociale ». In N. Ramognino & P. Vergès (dir.), *Le français hier et aujourd'hui : politiques de la langue et apprentissages scolaires : études offertes à Viviane Isambert-Jamati*. Aix : Publications de l'université de Provence, p. 49-67.
- BAUTIER É. ; CRINON J. ; RAYOU P. & ROCHEX J.-Y. (2004). *Socialisation scolaire et non scolaire des élèves. Présupposés et mobilisés chez les jeunes évalués : analyse secondaire de l'enquête PISA 2000*. Rapport de recherche pour la Direction de l'évaluation et de la prospective du ministère de l'Éducation nationale. Saint-Denis [Seine-Saint-Denis] : université Paris 8 : équipe ESCOL.
- BAUTIER É. ; CRINON J. ; RAYOU P. & ROCHEX J.-Y. (2004). « Les performances en littéracie et l'hétérogénéité des univers mentaux mobilisés par les élèves », *Cadmo : giornale italiano di pedagogia sperimentale*, vol. XII, n° 2, p. 43-64.
- BAUTIER É. & GOIGOUX R. (2004). « Difficultés d'apprentissage, processus de secondarisation et pratiques enseignantes : une hypothèse relationnelle », *Revue française de pédagogie*, n° 148, p. 89-100.
- BAUTIER É. & ROCHEX J.-Y. (1998). *L'expérience scolaire des « nouveaux lycéens ». Démocratisation ou massification ?* Paris : A. Colin.
- BAUTIER É. & ROCHEX J.-Y. (2004). « Activité conjointe ne signifie pas signification partagée ». *Raisons éducatives*, n° 9, p. 199-220.
- BERNSTEIN B. (1975). *Langage et classes sociales*. Paris : Éd. de Minuit.
- BONNET G. (2002). « Reflexion in the critical eye : on the pitfalls of international assessment ». *Assessment in Education*, vol. 9, n° 3, p. 387-399.
- BOTTANI N. & VRIGNAUD P. (2005). *La France et les évaluations internationales*. Rapport pour le Haut conseil d'évaluation de l'École. Paris : ministère de l'Éducation nationale.
- BOURNY G. ; BRAXMEYER N. ; DUPÉ C. ; RÉMOND M. ; ROBIN I. & ROCHER T. (2002). *Les compétences des élèves français à l'épreuve d'une évaluation internationale. Premiers résultats de l'enquête PISA 2000*. Paris : ministère de l'Éducation nationale : Direction de la programmation et du développement.
- DURU-BELLAT M. ; MONS N. & SUCHAUT B. (2004). *Les Cahiers de l'IREDU*, n° 66 : « Caractéristiques des systèmes éducatifs et compétences des jeunes de 15 ans : l'éclairage des comparaisons entre pays ».
- GIPPS C. (1998). « Socio-Cultural Aspects of Assessment ». *Review of Research in Education*, vol. 24, p. 355-392.
- GOLDSTEIN H. (1995). « Interprétation des comparaisons internationales des résultats scolaires ». *Études et documents d'éducation*. Paris : UNESCO, p. 11-32.
- GOLDSTEIN H. (2004). « International comparison of students' attainment : Some issues arising from the PISA study ». *Assessment in Education*, vol. 11, n° 3, p. 319-330.
- MASTERS G. N. (1995). « Élaboration d'échelles de mesure et agrégation dans les études de l'IEA : critique du rapport du Professeur Goldstein ». *Études et documents d'éducation*. Paris : UNESCO, p. 33-39.
- MURAT F. & ROCHER T. (2004). « On the methods used for international assessments of educational competences ». In J. H. Moskowitz & M. Stephens (éd.), *Comparing Learning Outcomes : International assessment and education policy*. London : RoutledgeFalmer, p. 190-214.
- OCDE (2001). *Connaissances et compétences : des atouts pour la vie. Premiers résultats de PISA 2000*. Paris : OCDE.
- OLSON D. (1994). *The World on paper. The conceptual and cognitive implications of writing and reading*. Cambridge : Cambridge University Press.
- OLSON D. (1998). *L'univers de l'écrit*. Paris : Retz.
- RAYOU P. (2002). *La « dissert de philo » : sociologie d'une épreuve scolaire*. Rennes : Presses universitaires de Rennes.
- REMOND M. (2001). « Adapter n'est pas traduire. Adaptation dans différents contextes culturels d'épreuves d'évaluation de la littéracie ». In C. Sabatier & P. Dasen (éd.), *Cultures, développement et éducation : autres enfants, autres écoles*. Paris : L'Harmattan, p. 171-184.
- RÉMOND M. (2006). « Éclairages des évaluations internationales PIRLS et PISA sur les élèves français ». *Revue française de pédagogie*, n° 157, p. 71-84.
- ROBIN I. (2002). « L'enquête PISA sur les compétences de lecture des élèves de 15 ans : trois biais culturels en question », *VEI Enjeux*, n° 129, p. 65-91.

ROCHER T. (2003). « La méthodologie des évaluations internationales de compétences ». *Psychologie et psychométrie*, vol. 24, n° 2/3.

ROCHEX J.-Y. (sous presse). « Social, methodological and theoretical issues regarding assessment. Lessons from a secondary analysis of PISA 2000 literacy tests », *Review of Research in Education*, vol. 30, p. 163-212.

SCHNEUWLY B. (1994). « Genres et types de discours : considérations ontogénétiques et psychogénétiques ». In Y. Reuter (dir.), *Les interactions lecture-écriture*. Bern : P. Lang, p. 155-173.

VAN DIJK T. A. & KINTSCH W. (1983). *Strategies of discourse comprehension*. New York : Academic Press.

ANNEXE : LES CATÉGORIES DE L'ANALYSE DES RÉPONSES DES ÉLÈVES

Les compétences cognitives à traiter le texte

Comprendre un texte met en jeu des processus cognitifs permettant au lecteur de passer de la surface linguistique à un contenu sémantique propositionnel et à une représentation mentale de la situation décrite par le texte. Cinq indicateurs ont été retenus pour rendre compte de la compétence à traiter le texte écrit pour en construire la signification.

C1 : Mettre en relation les différents niveaux constitutifs d'un texte (hiérarchisation, local/global...) ou de l'ensemble des textes d'une épreuve

C2 : Construire les chaînes causales du texte

C3 : Inférer ce qui n'est pas explicite de ce qui est explicite dans le texte

C4 : Reasonner de manière vraisemblable ou plausible (contra non vraisemblable, incongrue par rapport aux connaissances sur le monde)

C5 : Reconnaître la permanence du sens sous des formes sémiotiques ou linguistiques différentes

La régulation de la lecture et la représentation de la situation de communication

Une autre manière d'envisager la compréhension en lecture est de considérer le texte comme un « discours » adressé, porteur d'intentions dans le cadre d'une communication sociale. L'appréhender oblige à le constituer comme objet, à construire le cadre plus largement social et discursif qui lui donne sens, à mettre en œuvre une familiarité avec la littéracie. Comprendre un texte, c'est en comprendre la visée. L'attribution d'une intention à un texte dans la situation scolaire (comme dans la situation d'évaluation PISA qui y ressemble fort), plus abstraite que dans la situation sociale de recours au même texte, fait tout particulièrement appel à cette compétence et

demande au lecteur de reconstituer un contexte. La compétence de décontextualisation qui permet justement une nouvelle contextualisation, participe de ce que nous avons nommé processus de reconfiguration ou de secondarisation.

La capacité (qui est donc aussi habitude) à saisir la visée a été décomposée en deux items d'analyse.

L1 : Saisir la dimension textuelle et la visée du texte ou de l'ensemble des textes d'une épreuve (vs. prélèvement d'un indice isolé entraînant une incompréhension).

L2 : Saisir la visée de la question

En outre, deux items nous serviront d'indicateurs de la position d'adhérence au langage ou au contraire d'un recul réflexif (prendre le langage, en se détachant des contenus référés, comme un objet).

L3 : Mettre à distance le texte, « faire quelque chose avec le langage », élaborer avec ses propres mots (par opposition à « faire simplement œuvre de reprise des mots du texte »)

L4 : Prendre les aspects linguistiques et rhétoriques du texte comme objets de réflexion

La façon d'être au monde

La troisième grande entrée de notre grille renvoie aux manières de l'élève d'« être au monde », aux valeurs, aux croyances et aux habitudes intellectuelles qui lui permettent d'interpréter le monde et l'expérience et qui s'interposent potentiellement dans la lecture des textes qui leur sont proposés. Dans une démarche inductive, nous en avons dénombré six dans les réponses aux épreuves PISA, qui situent les élèves tantôt du côté de la construction active d'une signification et de l'élaboration d'un point de vue personnel tenant compte du texte et/ou se confrontant

avec celui-ci, tantôt de la juxtaposition de fragments du texte lu et d'énoncés d'opinion en provenance de l'environnement social ou scolaire, voire de l'occultation du texte par des énoncés qui lui sont étrangers.

E1 : Prendre le monde comme objet de connaissance et d'analyse (secondarisation du monde)

E2 : Mobiliser un point de vue relativiste sur le monde

E3 : Mobiliser le point de vue de l'expérience personnelle

E4 : Mobiliser un point de vue conformiste ou affectif

E5 : Se référer essentiellement au texte

E6 : Mobiliser des savoirs non scolaires, culturels, nécessaires à la compréhension du texte

Éléments transversaux

Enfin une série d'indicateurs ont été pris en compte de manière « transversale », c'est-à-dire qu'ils ont été codés dans l'ensemble du cahier et non pas décomptés épreuve par épreuve comme les autres catégories.

Ils concernent la manière de se situer par rapport aux savoirs scolaires (*T1 : Mobiliser des savoirs, des attitudes et des méthodes scolaires (savoirs formels, pratiques ou de contenu)*) ; la présence du sujet dans ses réponses (ou son absence), qu'elle se manifeste par la présence de marques de modalisation (*T2 : la réponse est accompagnée d'un avis personnel et/ou de modalités manifestant l'élève comme auteur du texte*), par une activité intellectuelle de problématisation (*T3 : Problématiser (mouvement au sein du discours produit entre singulier et général, personnel et impersonnel, potentiel et réel, fait et opinion...)*), par la formulation d'hypothèses (*T4*), ou par l'engagement de soi et la prise de position dans la réponse (*T7*).

Ils concernent aussi des compétences relatives au code écrit dont l'absence est considérée par les enseignants comme caractéristique des élèves en échec : les lacunes syntaxiques et orthographiques massives (*T5*) et la « pauvreté » du vocabulaire (*T6 : Ne pas comprendre des mots et des notions*).

Dans ces indices figure également le fait de répondre ou de ne pas répondre.

Procédés discursifs et rôles sociaux d'élèves en groupes de discussion sur une controverse socio-scientifique

Virginie Albe

L'étude de questions socio-scientifiques en classe a été proposée afin de contribuer à la démocratisation des sciences. À cette fin, des situations favorisant les pratiques discursives entre élèves ont été développées. Nous avons étudié comment des élèves de première appréhendent une controverse socio-scientifique actuelle en groupes de discussion. Dans une approche micro-ethnographique, nous avons identifié différents procédés de discussion collective dans les groupes. Les résultats montrent que des facteurs épistémologiques, contextuels, et sociaux influencent l'élaboration de connaissances dans ces situations.

Descripteurs (TEE) : apprentissage, argumentation, conditions d'apprentissage, éducation scientifique, travail par équipe.

INTRODUCTION

L'enseignement scientifique est actuellement considéré en crise par de nombreux acteurs : politiques, enseignants, scientifiques, journalistes, etc. Face à ce que certains nomment la « désaffection des jeunes pour les sciences », beaucoup s'interrogent sur les causes de cette « crise » et tentent de formuler des propositions (Royal Society [GB], 1985 ; Union européenne, 1995 ; OECD, 1998). L'adéquation de la formation des jeunes à l'emploi, les évolutions sociales, la mondialisation de l'économie, la complexité du monde, le déficit du sens dans les sociétés technocratiques, l'image des sciences et des scientifiques

dans la société, la place et le rôle des médias sont questionnés (1).

Le système de formation et les pratiques pédagogiques sont également interrogés. Dans le champ de l'éducation aux sciences, les phénomènes d'enseignement et d'apprentissage font l'objet de recherche depuis des décennies. L'étude des pratiques effectives a notamment permis d'établir le constat d'un enseignement des sciences essentiellement selon une perspective positiviste (Aikenhead, 2003 ; Driver *et al.* 2000 ; Fourez 2002). Pour Fleming (1986), la science scolaire s'intéresse peu à la science. Elle traite de ses produits, de connaissances factuelles et les élèves la perçoivent comme une accumulation de faits. Dans

ce contexte, de nombreux auteurs en appellent au changement (Désautels *et al.*, 1993 ; Hodson, 2003 ; Millar & Osborne, 1998 ; Solomon & Aikenhead, 1994), soulignant qu'il est nécessaire de reconsidérer l'éducation scientifique, couramment vue comme une formation introductive *en* sciences qui se concentre sur le développement de compétences méthodologiques pour une formation *sur* les sciences.

De nouveaux programmes ont ainsi été mis en place dans divers pays dans le but d'une alphabétisation scientifique pour tous. Si ce but semble partagé par tous, les considérations sur ce qu'une personne alphabétisée doit être en mesure de réaliser sont très disparates.

Selon une orientation pragmatique, certains considèrent – puisque les sociétés sont de plus en plus influencées par les idées et les produits des sciences, et surtout des technologies – que tous les futurs citoyens seront mieux en mesure d'affronter les problèmes s'ils ont une base de connaissances scientifiques. Il s'agit par exemple de développer une confiance rationnelle dans les experts lorsque l'on se trouve confronté à des questions incertaines et controversées (Shamos, 1995).

Pour d'autres, il s'agit d'utiliser les connaissances scientifiques dans un but de démocratisation. Dans cette orientation, on considère qu'une formation scientifique de qualité à l'école permet aux citoyens de participer significativement aux décisions que les sociétés doivent maintenant prendre à l'égard de problèmes socio-scientifiques et socio-technologiques toujours plus complexes (production et consommation d'énergie, sécurité alimentaire par exemple).

Prenant en compte les préoccupations actuelles en Amérique du Nord, Hodson (2003) propose par exemple de construire des *curricula* dans une visée d'action socio-politique autour de sept thèmes : la santé humaine, l'alimentation et l'agriculture, les ressources naturelles (eau, terre, minéraux), les ressources et la consommation d'énergie, l'industrie, les technologies de l'information et les transports, la responsabilité éthique et sociale. Considérant que les liens entre l'éducation aux sciences, la mondialisation économique, l'augmentation de la production et l'expansion sans limites menacent les libertés individuelles, le bien-être des sociétés et le futur de la planète, Hodson souhaite que ce *curriculum* constitue un engagement à poursuivre la refondation des valeurs des sociétés occidentales. Il souligne qu'une formation à la nature des sciences et aux pratiques scientifiques est un élément clé dans le dévelop-

pement d'une alphabétisation scientifique critique favorisant une participation active dans nos sociétés démocratiques.

Dans la même perspective de démocratisation des sciences, Driver *et al.* (2000) proposent de favoriser en classe la formation socio-épistémologique des élèves et soulignent l'importance de mettre l'accent sur le rôle de l'argumentation dans l'élaboration des connaissances scientifiques. Ils considèrent que pour que des individus s'engagent dans des débats, il est nécessaire qu'ils connaissent la façon dont des arguments sont élaborés en science et qu'ils soient capables d'évaluer les utilisations possibles et les limites des savoirs experts pour appuyer des points de vue.

Leur approche s'inscrit dans l'esprit de l'épistémologie contemporaine et s'appuie sur les résultats de la sociologie des sciences considérant que les faits scientifiques sont élaborés au sein d'une communauté par des pratiques socio-discursives et la confrontation à la « récalcitrance » du réel.

Des questions telles que les méthodes de recherche (épidémiologie par exemple), les aspects institutionnels des sciences (sociologie) devraient ainsi faire l'objet d'enseignements et Driver *et al.* (2000) proposent que les élèves pratiquent eux-mêmes le débat et l'argumentation lors d'activités portant sur la science en train de se faire (situations historiques ou contemporaines) alors qu'aujourd'hui, la parole en classe leur est peu donnée.

Ainsi, viser la démocratisation des sciences à l'école nécessite de proposer des situations destinées à favoriser le développement d'une épistémologie renouvelée et le développement des capacités argumentatives des élèves.

Il est dans ce cadre nécessaire de recentrer les activités pédagogiques et de les articuler à des questions vives sur le plan social et significatives pour les élèves.

Une des voies proposées concerne l'introduction dans les programmes de questions socio-scientifiques controversées (réchauffement du climat, modes de production d'énergie, clonage, etc.). Dans ces situations, une participation active des élèves est recommandée (France, 1999 & 2000) (2). À l'appui de travaux qui soulignent que les apprentissages scientifiques et le langage sont inextricablement liés, il est alors proposé de mettre en place des situations où les interactions langagières entre élèves sont favorisées (travaux en petits groupes, débats, cours dialogués, exposés, etc.). Au lieu d'être considérées comme

des alternatives non problématiques aux modèles d'enseignement transmissifs, ces activités font l'objet de recherches depuis plusieurs années (Barnes & Todd, 1977 ; Kelly & Crawford, 1996 ; Gilly, 1988 ; Schubauer-Leoni & Perret-Clermont, 1980). En particulier, des travaux en didactique des sciences et des langues ont porté sur l'impact des interactions entre élèves dans la construction de savoirs scolaires en sciences (Amigues, 1988 ; Fillon & Peterfalvi, 2004 ; Garcia-Debanc, 1995 ; Jaubert & Rebière, 2001).

Dans cet article, nous présentons une étude de l'impact de situations interactives entre élèves sur leur appréhension d'une controverse socio-scientifique. En particulier, nous avons tenté de cerner comment les interactions sociales, articulées aux confrontations avec des discours scientifiques controversés, interviennent dans les productions discursives des élèves lors de petits groupes de discussion.

LE DISCOURS EN CLASSE DE SCIENCES, L'ÉTUDE DE QUESTIONS SCIENTIFIQUES SOCIALEMENT VIVES

Des études ont montré l'importance des échanges oraux entre élèves pour leur permettre de développer leur compréhension conceptuelle (Barnes, 1977 ; Barnes & Todd, 1977 ; Lemke, 1990 ; Sutton, 1995). Ceci met en avant les « relations étroites qui apparaissent, de plus en plus clairement, entre savoirs et productions langagières » (Jacobi & Peterfalvi, 2004, p. 4). Garcia-Debanc (1995, p. 102) souligne que « les interactions verbales apparaissent comme une composante essentielle de la construction des apprentissages dans une démarche scientifique ». De même Jaubert et Rebière (2001) considèrent qu'il ne peut y avoir construction de savoirs sans la construction simultanée des discours qui permettent de dire ces savoirs.

L'effet des pratiques de travail en petits groupes sur les apprentissages a ainsi fait l'objet de recherches dans le domaine de l'éducation aux sciences. Des auteurs ont montré que les relations entre les élèves, leur statut académique, leur genre influencent et structurent les interactions dans les groupes (Bianchini, 1997 ; Goffard & Goffard, 2003 ; Moje & Shepardson, 1998). Les rôles sociaux adoptés par les élèves au cours des discussions de groupe ont également été étudiés et cela a permis de montrer un effet de « *leader* » sur les connaissances élaborées au sein des groupes (Richmond & Striley, 1996). Par ailleurs,

l'analyse de l'influence des contextes d'enseignement a permis de montrer que les élèves ont souvent des difficultés à travailler collectivement (Alexopoulou & Driver, 1996 ; Bianchini, 1997 ; Kittleson & Southerland, 2004). D'autres recherches ont porté sur les procédés d'élaboration collective des savoirs entre élèves en groupes de discussion (Kelly & Crawford, 1997). Différents modes de coopération ont été identifiés (Gilly, 1988 ; Goffard & Goffard, 2003) et les dynamiques des échanges dans les groupes de discussion ont fait l'objet d'analyses (Schneeberger & Ponce, 2003). Si le rôle important des situations de collaboration entre pairs dans la marche vers l'élaboration des connaissances semble faire l'unanimité, la plupart des auteurs signalent que ce n'est pas nécessairement un processus productif.

Fillon et Peterfalvi (2004) s'interrogent par exemple sur l'impact des discussions de groupe sur les apprentissages des élèves. En référence à Meirieu qui « dès 1984, proposait une interrogation critique sur l'effet des pratiques de travail en groupe sur les apprentissages, ils indiquent que l'engagement cognitif des élèves peut s'avérer très inégal selon leur position dans le groupe ou par rapport à l'enseignant » (*ibid.*, p 176). Pour Dawes (2004) si les apprentissages scientifiques et le langage sont inextricablement liés, les groupes de discussion ne conduisent pas nécessairement à une collaboration des élèves en vue de construire des connaissances dans la mesure où la demande sociale de la situation est souvent trop forte.

Par ailleurs, les travaux qui ont porté sur l'étude par les élèves de controverses socio-scientifiques montrent qu'il s'agit de phénomènes complexes (Bader, 2003 ; Sadler, Chambers & Zeidler, 2004 ; Schweizer & Kelly, 2001 à propos du réchauffement climatique ; Albe, 2005 sur les technologies de la téléphonie mobile ; Simonneaux, 2001 & 2003 sur les biotechnologies). Plusieurs tendances peuvent toutefois être identifiées : l'influence significative de l'expérience personnelle des élèves et de leurs émotions, la prise en compte de considérations sociales et l'importance de considérations épistémologiques, et, dans certains cas, la primauté de considérations morales (Sadler & Zeidler, 2005). La mobilisation de connaissances scientifiques et technologiques par les élèves dans l'étude de ces controverses a également été questionnée.

L'expérience personnelle apparaît comme un moyen pour interpréter et problématiser la controverse, une référence pour argumenter et justifier une position : étude par des lycéens de questions environnemen-

tales (Patronis *et al.*, 1999), liées à la thérapie génique (Sadler & Zeidler, 2004) et à l'énergie nucléaire (Fleming, 1986). Pour certains auteurs, les émotions ont un effet facilitateur pour engager les élèves dans l'étude de controverses socio-scientifiques (Sadler *et al.*, 2004). L'empathie ou la sollicitude peuvent en particulier permettre de prendre en considération les points de vue d'autres individus ou de cerner l'impact individuel et social des décisions prises. Ainsi, dans la recherche de Sadler & Zeidler (2005) les élèves se sont identifiés aux personnages mis en scène dans les scénarios socio-scientifiques proposés sur le clonage et ont adopté une perspective plurielle pour appréhender des points de vue opposés et controversés (Sadler & Zeidler, 2005).

Par ailleurs, des recherches ont montré qu'un engagement émotionnel important accompagné de la capacité de réguler et contrôler cet engagement par des mécanismes communicationnels et sociaux ont conduit à développer les points de vue empathiques, les comportements en faveur d'action sociale et les comportements visant à faire face aux problèmes de façon constructive (Eisenberg, 2000).

Si ces résultats plaident en faveur de la prise en compte du rôle des émotions dans l'éducation aux sciences alors que traditionnellement seul un raisonnement rationaliste est [positivement] considéré [ou uniquement pris en considération], il nous semble que dans ce domaine également une attitude de prudence soit nécessaire.

De même que mettre des élèves en petits groupes ne conduit pas nécessairement à une co-construction de connaissances, proposer aux élèves des situations portant sur des controverses socio-scientifiques ayant une résonance émotionnelle forte n'assure pas leur engagement dans leur étude. Si des questions telles que l'usage de téléphones cellulaires, la production et la consommation d'OGM, le réchauffement climatique, etc. permettent de rejoindre des préoccupations personnelles des élèves, ce qui est peu le cas dans les cours de sciences traditionnels, il se peut également qu'une dimension affective forte dans ces controverses ait un effet inhibiteur, des élèves pouvant être amenés à taire leurs points de vue dans des discussions qui les impliqueraient trop personnellement. Par ailleurs, la confrontation des élèves à des controverses où la science est incertaine et ne peut fournir de certitudes, de faits incontestables qui font autorité comme lors des enseignements traditionnels peut être déstabilisante pour les élèves et les enseignants (Hind, Leach & Ryder, 2001) et générer anxiété, vulnérabilité et insécurité.

Dans la mesure où les controverses socio-scientifiques peuvent avoir une dimension affective forte, pour Dewhurst (1992) adopter un autre point de vue que le sien peut aider à mieux comprendre mais il signale que ce n'est pas un processus évident. Il nous semble également qu'un recours aux émotions pour engager des élèves dans l'étude de controverses pourrait conduire à une polarisation des points de vue, ce qui peut être un obstacle à la compréhension de questions complexes.

Pour d'autres auteurs, la prise en compte par les élèves de considérations sociales jouent un rôle majeur dans leur appréhension de controverses socio-scientifiques. Fleming (1986) a montré que les étudiants raisonnent principalement la question d'accepter un emploi dans une centrale nucléaire dans le domaine social en évaluant notamment la possibilité d'effets sur d'autres individus ou dans le domaine personnel en recherchant un intérêt individuel sur le plan économique. Les élèves utilisent peu de connaissances scientifiques et ne considèrent pas les connaissances scientifiques scolaires comme pertinentes. Des étudiants précisent par exemple qu'il leur faudrait plus que la science scolaire sur les radiations comme par exemple connaître les effets sur la santé des personnes exposées. Tous les élèves interrogés seraient prêts à s'engager dans des études en classe sur ces aspects scientifiques.

Si les connaissances scientifiques sont réduites à l'emploi de quelques terminologies face à ces questions socio-scientifiques, en revanche, les connaissances sur les scientifiques et la science comme une institution constituent les connaissances les plus couramment utilisées par les étudiants (Fleming, 1986). La science pour les étudiants est une accumulation de faits (c'est ce qu'ils connaissent de la science scolaire), la technologie utilise ces faits afin de conduire au progrès pour faire avancer « notre civilisation ». Ainsi, ils considèrent qu'avec plus de faits, on résoudrait la question socio-scientifique.

Par ailleurs, pour Yang et Anderson (2003), les considérations sociales sont le plus souvent articulées à des considérations scientifiques. De même pour Sadler *et al.* (2004), les étudiants considéraient les données scientifiques associées à différentes positions ainsi que les conséquences sociales de ces positions.

Peu de travaux se sont centrés sur les pratiques discursives des élèves lors de groupes de discussion portant sur des controverses socio-scientifiques.

Leach & Lewis (2002) ont décrit les manières de parler au sein de groupes discutant du diagnostic prénatal de la fibrose kystique, en utilisant les catégories de Mercer (1996). Ils ont observé quelques groupes, composés de filles, avec un style de discours « d'exploration ». Dans les autres groupes, les élèves répètent leurs avis personnels, ne tentent pas de les justifier ou d'argumenter pour s'opposer aux affirmations opposées.

Dans la mesure où les questions soulevées par ces controverses sont porteuses d'affect, où les connaissances scientifiques élaborées dans la cité savante sont incertaines et controversées, et par conséquent une maîtrise des contenus académiques de la science scolaire n'offre que peu de recours, comment les élèves appréhendent la controverse dans le cas de groupes de discussions ? Quels procédés discursifs élaborent-ils collectivement ? Quelle organisation sociale développent-ils dans la discussion ? Quels sont les éléments qui influencent les connaissances élaborées dans les groupes ? Telles sont les questions que nous avons étudiées dans la présente recherche.

ANCRAGES THÉORIQUES

Dans une approche située de la cognition, il s'agit « d'envisager les processus cognitifs et l'activité comme indissociables d'une situation, dont les éléments physiques, artefactuels autant que sociaux, offrent des ressources signifiantes pour l'action des sujets » (Grison, 2004).

La cognition de l'être humain y est vue comme incarnée et historiquement construite. Les connaissances sont un résultat de l'activité et des situations dans lesquelles elles ont été produites. La cognition située met également l'accent sur le caractère social de la construction des connaissances. L'apprentissage est réalisé par des interactions sociales collaboratives (Resnick, 1988). Autrement dit, dans l'activité, le sens est socialement construit à travers les négociations entre acteurs (Brown, Collins & Duguid, 1989). Cette perspective de la cognition défend une posture anti-mentaliste (Coulter, 1979) et rejoint la sociologie interactionniste et l'ethnométhodologie (Garfinkel).

L'interactionnisme symbolique met également l'accent sur la dynamique co-constructive du sens considérant que la signification donnée à une situation se crée pendant l'interaction. La perspective prag-

matique de l'interactionnisme rejoint ainsi les approches constructivistes de la cognition. Cette approche met aussi en avant la complexité des rôles sociaux et le caractère dynamique et négocié de l'interaction sociale. Les rôles sociaux sont le résultat de processus complexes où chaque acteur participe à la représentation de la situation dans laquelle il est engagé, où le déroulement de l'interaction n'est jamais figé, mais sans cesse ré-adapté, réinvesti, négocié et donc, bien plus le fait de conflits, de négociations, de divergences d'intérêts, de tensions affectives que la conséquence de règles strictes et fonctionnelles – souvent ignorées d'ailleurs – par les acteurs auxquels elles sont censées s'appliquer (Strauss, 1992).

Quant à l'ethnométhodologie qui s'attache à comprendre comment le langage est imbriqué dans l'organisation des activités sociales, elle ne recourt pas à des explications intentionnelles ou psychologiques pour expliquer les actions parce ce qu'elle les considère d'abord comme des constructions discursives.

Ainsi, les points de vue sur la cognition, les interactions sociales et le langage se rejoignent pour considérer que les connaissances sont co-construites à travers les interactions interpersonnelles en situation grâce à une réalité intersubjective reposant sur des symboles langagiers partagés.

Afin de cerner les procédés discursifs collectivement élaborés par les élèves en groupes de discussion lors de l'étude d'une question scientifique socialement vive, nous avons opté pour une approche focalisée sur les pratiques des locuteurs qui mobilisent des ressources en situation, et ainsi tient compte des orientations des participants dans l'action. Mettant l'accent sur la dimension située et interactive des pratiques langagières, nous aurons recours à la notion d'objet de discours. Comme le souligne Mondada (2005, p. 23), « les objets de discours sont interactivement proposés, repris, ratifiés, transformés, rejetés par les interlocuteurs. Ces objets sont ainsi constitués au fil des activités communicationnelles ; Ils construisent une « version publique du monde ». Celle-ci est élaborée en adéquation avec les relations intersubjectives et sociales des participants et avec le contexte social dans lequel elle est formulée. »

Par ailleurs, nous nous appuyons sur l'interactionnisme symbolique pour étudier l'organisation sociale développée par les élèves dans la discussion, considérant que l'ordre social au sein des groupes de discussions est un ordre négocié. Pour les interactionnistes symboliques, les structures se construisent

pendant les interactions entre acteurs et sont le résultat de ce processus. Ainsi, dans cette étude, nous considérons que les rôles sociaux sont le fruit de l'activité des membres des groupes de discussion. Ces rôles sont structurés par les élèves à travers les processus d'interaction. Ils se construisent, évoluent et se négocient dans les interactions. Ces rôles sociaux sont également structurants des discussions au sein des petits groupes.

CONSIDÉRATIONS MÉTHODOLOGIQUES

« Les téléphones cellulaires sont-ils dangereux pour la santé ? » (3) est une situation-débat (Albe & Simonneaux, 2003) inspirée d'un module de formation développé par un groupe de chercheurs et d'enseignants pour favoriser la formation épistémologique des élèves (Hind *et al.*, 2001).

Nous avons conduit cette étude de la question controversée de la dangerosité des téléphones cellulaires avec une classe de douze élèves de série technologique de l'enseignement agricole (Sciences et technologies de l'agronomie et de l'environnement) pendant une séance ordinaire d'environ deux heures. Il s'agit de préparer et de réaliser un jeu de rôle : les élèves jouent le rôle d'avocats dans le cas d'un procès où un employé poursuit son employeur pour son mauvais état de santé, qu'il estime dû à l'usage du téléphone portable et en raison duquel il a dû quitter son travail. Les élèves se répartissent en deux groupes : les avocats de la partie civile et de la défense. Le groupe A défend la victime et la thèse que les téléphones portables sont dangereux pour la santé ; le groupe B défend l'employeur et la thèse que les téléphones portables ne sont pas dangereux pour la santé.

Un dossier est distribué aux élèves. Il est constitué de sept extraits de recherche actuelles portant sur l'apparition de maladies sur des animaux, des enquêtes épidémiologiques, des tests sur la mémoire. Chaque groupe étudie les documents et prépare des arguments pour le jeu de rôle.

Toute la leçon a fait l'objet d'enregistrements audio et vidéo. La totalité des échanges ont été retranscrits (4).

Considérant que le langage est une activité sociale qui permet l'élaboration du sens et que structuration de la pensée et échanges langagiers sont liés (Billig, 1996), nous nous sommes centrée sur les

pratiques discursives des élèves et leur construction de savoirs plutôt que sur les caractéristiques linguistiques du discours lui-même (Edwards & Mercer, 1987). Considérant les connaissances élaborées par les élèves dans cette activité comme locales, contingentes et dépendantes des groupes sociaux, nous avons analysé les pratiques discursives au sein des groupes d'élèves sans séparer la cognition du social, comme le suggèrent Kelly et Crawford (1997). Pour cela, nous avons effectué une analyse de discours d'inspiration ethnométhodologique car comme l'indique Mondada (2005, p. 84) « cette perspective permet d'observer, à la fois, le fonctionnement d'un collectif et la production collective des savoirs, en veillant à ne pas les séparer l'un de l'autre ». Ainsi, nous nous sommes focalisée sur les procédés d'élaboration collective des savoirs, autrement dit les procédés par lesquels des objets de discours introduits, identifiés et reconnus comme tels par les participants à la discussion sont développés (Mondada, 2005). Nous analysons l'ensemble des échanges au sein des groupes dans la mesure où les interventions des élèves dépendent du contexte de l'ensemble de la discussion (Pontecorvo, 1993). Comme le souligne Jimenez-Aleixandre (2003), dans le modèle de la cognition située, on considère qu'apprendre les sciences c'est être apprenti de pratiques discursives de la communauté scientifique. La co-construction de connaissances par les élèves a lieu lors de dialogues interactifs. Les arguments construits en collaboration entre deux ou plusieurs personnes tiennent compte des arguments des autres, pour les appuyer ou pour s'opposer à eux. Il faut ainsi procéder à une analyse de l'ensemble des discours produits, dans la mesure où une proposition d'un locuteur tient compte des échanges précédents. Pour cela, nous suivons pas à pas les échanges verbaux des élèves.

Pour ce faire, les outils d'analyse développés n'étaient pas pré-déterminés à la collecte de données mais ont été construits lors de multiples confrontations avec les données recueillies.

Dans un premier temps, nous repérons dans l'ensemble du *corpus* les passages où les échanges entre les participants à la discussion sont centrés sur un objet de discours. Les élèves élaborent collectivement ces objets, ils ne pré-existent pas à leur discussion dans l'interaction mais peuvent par exemple faire intervenir des éléments de savoirs fournis dans les extraits de recherche à étudier, qui seront alors négociés et retenus comme plausibles par les élèves.

Nous procédons ensuite à une analyse microscopique des épisodes précédemment identifiés.

Une tendance actuelle majeure consiste à généraliser toute activité discursive à l'argumentation. Pour Plantin (1996), « du point de vue du dialogue, peut être considéré comme argumentatif tout discours produit dans un contexte de débat orienté par une question ». Cette définition nous semble particulièrement adaptée à notre situation où les élèves se trouvent confrontés à une controverse socio-scientifique. Face à des discours scientifiques contradictoires, les élèves développent des pratiques argumentatives. Pour décrire ces pratiques, nous nous sommes inspirée de travaux de didacticiens portant sur l'argumentation en classe de sciences (Jiménez-Aleixandre *et al.*, 2000 ; Simonneaux 2003 ; Weisser, Masclat & Remigny, 2003). Le modèle de Toulmin a été considéré comme peu pertinent pour analyser l'argumentation dans des débats sur des questions scientifiques socialement vives (Simonneaux, 2003) et insuffisant pour rendre compte de ce qui s'échange dans la dynamique des interactions verbales (Fillon & Peterfalvi, 2005).

Afin de saisir les pratiques argumentatives des élèves et leur dimension sociale, l'analyse des tours de parole nous permet de repérer la nature et le contenu des interventions des élèves. Il peut s'agir de questions, propositions avec ou sans justifications, oppositions, réfutations, d'expressions d'accord ou de désaccord, de répétitions, de reformulations (Simonneaux 2003 ; Weisser, Masclat & Remigny, 2003). Les élèves peuvent mettre en discussion des éléments issus des extraits de recherche, des considérations stratégiques, sociales, des connaissances scientifiques, techniques, profanes, des points de vue sur les sciences, sur le travail du groupe, etc. Nous identifions également les élèves qui sont à l'initiative et à la clôture des épisodes (Jiménez-Aleixandre *et al.*, 2000), ceux qui éventuellement gèrent les prises de parole dans le groupe, les transitions entre les différents épisodes de l'activité, qui évaluent, soutiennent ou contestent les propositions des autres, font usage d'autorité, soulèvent des critiques, etc., afin de caractériser les rôles sociaux développés dans les interactions entre élèves.

RÉSULTATS

Les élèves du groupe A mettent en œuvre plusieurs procédés collaboratifs à partir des extraits de recherche, de connaissances quotidiennes et de considérations socialement partagées ou « truismes culturels »

(Billig, 1987). L'étude de la trajectoire des objets de discours montre plusieurs reprises des objets de discours, pouvant conduire à une argumentation collaborative. Dans certains cas, les élèves ont articulé leurs désaccords lors de procédés discursifs complexes.

Dans l'extrait suivant par exemple, qui a lieu au début de la discussion du groupe, il s'agit pour les élèves de statuer sur le fait que le lien entre les résultats obtenus sur des tissus animaux et les résultats chez les humains soit valide ou non. Ceci fait l'objet de confrontations contradictoires.

7. Arnaud : De toute façon on a tout intérêt à dire qu'il y a un lien.
- [Interruption par l'enseignante]
8. Sébastien : Mais justement, le lien, il faut le démontrer.
9. Arnaud : Hé bien, tu mets que ce sont des cellules animales, et que nous aussi on est constitués de cellules animales. Pourquoi ça marcherait chez l'un et pas chez l'autre ?
10. Arnaud : On peut leur dire : « Non, non, il n'y a pas de problème, c'est pas dangereux. »
- [interruption par l'enseignante pour préciser les consignes de l'activité]
15. Sébastien : Bon, [Arnaud], pour le premier, là. Moi je te dis que ce n'est pas cohérent parce que nous on est des mammifères et eux c'est des cellules animales.
16. Katia : Oui on est des mammifères et eux c'est des cellules animales.
17. Arnaud : Oui, mais c'est quand même le règne animal.
18. Sébastien : Oui, mais il est grand, le règne animal.
19. Arnaud : Tu ne vas pas leur dire aux autres : « Oui, de toute façon vous avez raison, le portable... »
20. Sébastien : On pourra dire oui, autre part, il y en a plein...
21. Arnaud : Il faut dire oui partout.
22. William : Comme ça on est sûrs.
23. Arnaud : Eux ils vont dire non partout, alors...
24. Sébastien : Bon, donnez votre opinion, là, qu'on soit plusieurs et après on vote.

Arnaud fait une proposition stratégique, Sébastien exprime son désaccord et argumente. Katia acquiesce et par ce procédé collaboratif soutient l'argument énoncé par Sébastien. William appuie les propositions stratégiques de Arnaud, en désaccord avec Sébastien. Sébastien propose de voter et organise le recueil des points de vue des autres dans

* Nous utilisons des pseudonymes pour les prénoms des élèves.

les échanges suivants, adoptant en cela un rôle de leader. Il apparaît que les désaccords ont déstabilisé l'objet de discours jusqu'à sa fragmentation.

La discussion porte sur une question de nature ontologique relative au statut de l'homme et les élèves montrent des difficultés à élaborer collectivement un point de vue. Hill (1989) cité par Laroche (2002) a d'ailleurs montré la difficulté pour les élèves de placer l'homme du côté de l'animal dans des catégorisations. Mais il apparaît également dans cet extrait que des considérations stratégiques orientent les discussions et façonnent les objets de discours.

Plus loin dans les échanges du groupe, à l'occasion d'une discussion sur les extraits à retenir pour justifier leur thèse, les élèves reviennent sur la question du lien entre les résultats obtenus sur des tissus animaux et les effets chez les humains.

Sébastien propose que l'extrait de recherche n° 2 soit utilisé. William exprime son accord et le renforce en précisant la conclusion de cette recherche. Puis Sébastien change d'avis et suggère d'utiliser l'extrait de recherche n° 1. William exprime son désaccord et le justifie en référence à un objet de discours élaboré au début de l'activité : la validité du lien entre résultats obtenus sur des animaux et effets sur l'homme est incertaine. Il apparaît ici que les élèves tendent à exclure les résultats qui ne peuvent fournir de certitude, ce qui ne semble pas sans lien avec une certaine conception des sciences comme établissant des vérités, des preuves, des faits certains.

514. Sébastien : Bon, moi, j'ai mis la 2.
515. William : La 2 aussi, réceptive au changement, oui, oui, ça provoque des changements pour la mémoire.
516. Sébastien : Ha non, on pourrait utiliser la 1, je pense plutôt. Comme ça on peut dire...
517. William : Mais non, mais non, parce que c'est d'un lien possible. Alors là, on n'est vraiment pas sûrs.
518. Arnaud : Là, c'est les nématodes ?
519. Sébastien : On peut dire que si ça agit sur les vers, alors ça a peut-être une incidence sur nous.
520. Arnaud : Voilà...C'est quand même avant tout des animaux enfin des...
521. Sébastien : Oui, c'est vivant, c'est...

Au tour 519, Sébastien propose un argument que Arnaud évalue au tour suivant puis complète. Sa formulation permet l'établissement d'un accord. En référence aux discussions précédentes (tours de parole 15 à 18), il place « vers » et « être humains »

dans une même catégorie, « les animaux », puis nuance sa formulation avec « enfin », ce qui permet à Sébastien d'affirmer son accord et de compléter l'énoncé en train de se faire en reformulant la catégorie avec le terme « vivant ». On constate lors de cet épisode que l'élaboration collective a lieu entre deux élèves du groupe, un autre pouvant être considéré comme une aide à la discussion, les autres (les filles) n'y participant pas.

Dans l'extrait qui suit, les désaccords ne parviennent pas à être dépassés par la discussion.

194. Arnaud : Oui, mais tu peux dire qu'en dix-huit mois tu ne peux pas développer un cancer.
195. Sébastien : Oui, mais tu peux quand même avoir une tumeur qui se forme. Si ! elle grossit au fur et à mesure, la tumeur.
196. Arnaud : Oui mais ...
197. Agnès : Oui ça arrive vite aussi
[Interruption : changement de face de la cassette audio]
202. Arnaud : Moi je réponds que la durée est trop courte.
203. Agnès : Oui, oui
204. Katia : Mais non, parce que, là, tu vas dans leur sens à eux en disant ça.
205. Arnaud : Mais non !
206. William : mais non au contraire, si on l'avait laissé plus longtemps, peut-être qu'il y aurait eu des risques.
207. Arnaud : Plus tu exposes plus tu as de risques.
208. Katia : Ah oui OK
209. Sébastien : Et c'est où qu'ils... ? Ha oui, vingt minutes, quand même vingt minutes c'est à peu près le temps qu'on... Tu ne passes même pas vingt minutes au téléphone par jour.
210. Arnaud : Enfin.
211. Sébastien : Hein, vingt minutes ?
212. Arnaud : Certains, je suis certain qu'ils y passent...
213. Sébastien : tu as un forfait de combien ?
214. Arnaud : Pas forcément moi !
215. Agnès : c'est vrai, il y en a ils passent beaucoup de temps
216. Sébastien : En moyenne, en moyenne, on ne passe pas vingt minutes, non, c'est vingt heures, c'est vingt heures par jour, ouais, tu ne passes pas vingt heures par jour au téléphone.
217. William : Ça n'existe même pas les forfaits de vingt heures pour un mois alors !!
218. Sébastien : Alors sur dix-huit mois je pense que l'expérience, elle est valable.
219. William : Alors, ça ne veut rien dire.
220. Sébastien : vingt heures, ça veut dire que quatre heures de repos.
221. William : [inaudible] ...ton portable pendant cinq ans.

222. Arnaud : Peut-être qu'en vingt heures, dix-huit mois c'est quand même trop court.
223. Agnès : oui, c'est vrai, c'est court
224. Sébastien : attends, c'est court ?
225. Arnaud : C'est peut-être beaucoup par jour mais...
- [...]
229. Arnaud : C'est trop court, un an et demi, tu ne peux pas développer un cancer en un an et demi quand même.
230. Sébastien : Ha si, ha, ça dépend. Non, non, non, moi je te dis qu'elle est valable. Vingt heures quand même, vingt heures ! Tu téléphones, allez, on va dire tu téléphones cinq minutes par jour,
231. Arnaud : je m'en fous
232. Sébastien : allez, un peu plus, allez on va dire un quart d'heure, donc vingt heures...
233. Arnaud : mais je m'en fous de la quantité, moi. Moi je te dis : c'est le temps, ceux qui développent les cancers, les humains, ils n'ont pas dix ans. Ils ont quel âge ? entre quarante et soixante ans.
234. Sébastien : De toute façon, le téléphone, ça fait combien de temps que ça existe, alors ? non, dix-huit mois je pense que c'est valable.
235. William : non, c'est pas assez
236. Sébastien : On va voter, on va voter, qui est-ce qui est de mon avis ?

La discussion porte au début de cet extrait entre Arnaud et Sébastien sur la durée d'exposition des souris dans la recherche. On note les formulations « oui mais » qui initient chaque prise de parole et indiquent l'opposition entre les deux locuteurs. Leurs propos indiquent qu'ils se réfèrent à la carcinogenèse humaine, ce qui semble traduire une tendance à l'anthropomorphisme dans leurs justifications comme d'autres auteurs l'ont observé (Jimenez *et al.*, 2000 ; Grace, 2005). Arnaud réaffirme son point de vue (202) et trouve le soutien d'Agnès dans une double affirmation puis l'opposition de Katia qui la justifie par des considérations stratégiques. Arnaud et William justifient la durée insuffisante avec des formulations collaboratives (reprise des formulations du locuteur précédent, complétées ou reformulées). Puis Sébastien fait émerger un nouvel objet de discours qui a pour effet de ne pas stabiliser l'argument précédent et de déplacer le désaccord entre Sébastien et Arnaud sur ce nouvel objet de discours.

On note lors de leurs échanges que la référence est l'expérience personnelle (durées des forfaits, des communications téléphoniques, de repos journalier). Au tour 222, Arnaud réaffirme son point de vue y intégrant l'objet de discours introduit par Sébastien.

Agnès soutient à nouveau cet argument dans un procédé de co-élaboration acquiesçante (Gilly, 1988) et Sébastien s'y oppose sous une forme interrogative ce qui conduit Arnaud à développer son point de vue. Arnaud et Sébastien s'opposent ensuite et on note la référence à la carcinogenèse humaine et à des connaissances quotidiennes, socialement partagées ou truismes culturels selon l'expression de Billig (1987). Sébastien s'oppose en utilisant la formulation « de toute façon » ce qui semble indiquer qu'il ne peut que se ranger à l'argument de Arnaud, son inscription dans une « évidence sociale » ne pouvant être discutée. Puis, se plaçant également sur le plan social, Sébastien questionne la durée d'utilisation de cette technologie pour s'opposer à l'argument de Arnaud. Enfin, il réaffirme son opposition sans la justifier. William s'oppose sur le même mode : il affirme son désaccord sans le justifier ou l'argumenter. Sébastien propose alors de voter et sollicite les autres, dans un rôle de *leader*. Les désaccords ont ici pour effet une déstabilisation de l'objet de discours allant jusqu'à sa fragmentation.

Par ailleurs, les élèves discutent à plusieurs reprises des résultats obtenus sur l'hippocampe de rats dans l'extrait de recherche n° 3. Lors d'un premier épisode de discussion de cette question, il apparaît que la prise en compte du discours potentiel de l'autre groupe (portée par Nicolas) permet aux élèves de développer des justifications. On constate ensuite que par deux fois un questionnement prenant en compte les contre arguments que pourrait avancer l'autre groupe (soulevé par Katia) ne crée pas d'opportunité pour développer le discours. Dans un cas, deux autres élèves assumant respectivement les rôles de *leader* et d'aide, collaborent sur un autre objet de discours. Dans l'autre cas, un autre objet de discours est collectivement discuté par l'élève dans le rôle de « *leader* » et deux élèves assumant le rôle d'aide dont l'élève ayant soulevé le questionnement (Katia et William). L'extrait suivant constitue le dernier épisode de discussion sur cette question.

587. Arnaud : Je n'ai pas compris l'hippocampe, là,...
588. Sébastien : Les arguments, on va dire quoi ?
589. Arnaud : L'hippocampe est trop profondément enfoui dans le cerveau.
590. Katia : Non, mais ça, c'est ce qu'ils risquent de dire si on dit que ça atteint...
591. Arnaud : Oui, mais si ça a effet sur quelque chose qui est au milieu du cerveau, pourquoi ça ne pourrait pas...

592. Katia : Oui, parce que si ça atteint une partie du cerveau, après, les cellules elles vont se développer et puis atteindre le...
593. Arnaud : Et puis, il n'y a pas de raison que ça attaque que juste le milieu du cerveau et pas le reste.
594. Katia : Oui et puis, après, ça ce propage partout.
595. Sébastien : Ouais.Ouais,
596. William : Ouais, ça va se propager.
597. Katia : Oui, c'est comme la tumeur au début c'est un petit truc, et puis, après...
598. William : Ça grandit.
599. Arnaud : Tu peux dire que c'est peut-être l'endroit le plus sensible du cerveau, l'hippocampe justement, mais que...
600. Katia : mais dire que c'est un endroit précis... C'est une raison à la con.. [inaudible]
601. Arnaud : Ça ne tient pas debout.
602. Katia : Oui, c'est débile. Bon allez c'est parti. Donc recherche 2.

La question de l'hippocampe est reprise par Arnaud au tour 587. Katia répète alors que cela pourrait être un argument que leur opposerait l'autre groupe. Arnaud propose un contre-argument : des effets ayant été observés sur l'hippocampe qui serait profondément enfoui dans le cerveau, pourquoi d'autres zones du cerveau ne seraient pas atteintes ? Les autres acquiescent et prolongent cette idée. Pour Katia « si ça atteint une partie du cerveau [...] après, ça ce propage partout » (594). William acquiesce en reprenant la formulation de Katia (596) et en la reformulant (598). Par ces procédés collaboratifs, les élèves ont réalisé un accord sur cet argument. Arnaud et Katia partagent ici la construction de l'argumentation. Les autres acquiescent par divers procédés (simple accord, accord et prolongement du discours).

Dans le groupe de discussion B, il apparaît que les élèves éprouvent des difficultés pour faire émerger des objets de discours collectifs et ne parviennent pas à articuler leurs désaccords dans des processus de discussion collective. Le plus souvent, ils mettent en œuvre des procédés d'acquiescement (un élève fait une proposition qui est acceptée). Les procédés collaboratifs ont le plus souvent lieu entre deux élèves plutôt que dans le groupe de discussion comme dans l'extrait suivant où les élèves établissent la non fiabilité d'une recherche à partir de considérations sur le mode d'établissement des connaissances par les chercheurs.

396. Cécile : Hé bien, ça, ils ne sont pas sûrs...
397. Nathan : Ils se basent sur des hypothèses, donc, encore une fois.
398. Cécile : qu'ils croient
399. Nathan : Donc les chercheurs se basent sur des hypothèses qui ne sont pas forcément justes.

À plusieurs reprises, les élèves expriment leurs points de vue individuels plutôt qu'ils ne font émerger un objet de discours collectif comme l'illustre l'extrait suivant.

30. Fabien : Qu'est-ce que c'est que ça ?
31. Caroline : Là il y a marqué : « une recherche plus récente »
32. Fabien : On met oui la seconde aussi, alors. Non, je me trompe.
33. Nathan : pour le débat, après, il faut qu'on soit du même avis quand même.
34. Fabien : Tu n'as qu'à imposer ton opinion.
35. Caroline : Non.
36. Fabien : Je veux dire : il n'a qu'à parler, au lieu de faire l'idiot.
37. Nathan : Je ne veux pas faire l'idiot, mais vous êtes : toi tu mets une réponse et Caroline elle met une autre réponse.
38. Caroline : [Inaudible]
39. Max : de toute façon, Fabien, tu comprends jamais rien
40. Sandra : Je sais c'est fait exprès
41. Max : qu'est-ce qu'il y a ?
42. Caroline : C'est oui, oui, c'est non ?
43. Fabien : et faudrait pas foutre trois jours pour faire un document hé parce que sinon ça va pas ! Non, c'est oui, oui ou...

Apparaît [apparaissent ?] ici des désaccords entre les élèves et des attaques personnelles. Une intervention ironique et le recentrage sur l'activité ont pour effet de ne pas conduire à un conflit et finalement deux élèves se mettent d'accord sous la forme d'une co-élaboration acquiescante pour affirmer qu'ils ne peuvent pas déterminer si cette recherche est reproductible.

L'adoption des rôles d'avocats peut permettre aux élèves de collaborer et de parvenir à un accord collectif. Dans l'extrait qui suit par exemple, les élèves ont collectivement établi que la recherche de l'extrait n° 4 n'est pas valide car il s'agit de souris génétiquement modifiées.

105. Fabien : On a utilisé des souris qui ont été génétiquement modifiées pour augmenter leur...
106. Max : Elles ont été génétiquement modifiées : c'est pas bien.
107. Nathan : Non, mais, ils disent qu'elles ont été génétiquement modifiées.
108. Une fille : Oui.
109. Nathan : Ça fausse tout alors. C'est pas bien.
110. Fabien : Oui, Oui, exactement tu as raison.
111. Nathan : Vous êtes d'accord avec moi, Maître Fabre ?
112. Fabien : Je suis entièrement d'accord avec votre proposition qui est satisfaisante.
113. Nathan : Je suis tout à fait content que nous tombions en accord.
114. Céline : Je me réjouis fortement.
115. Nathan : Monsieur Wilfrid quelle est votre opinion SVP ?
116. Max : Là déjà oui.

Lors de désaccords, l'expression personnelle autoritaire d'un élève sur la base de considérations stratégiques peut conduire à un accord dans le groupe comme dans l'extrait suivant.

145. Nathan : Encore une recherche faussée, je trouve.
146. Caroline : Génétiquement modifiée !
147. Céline : Oui.
148. Nathan : On peut directement annuler ces deux expériences car... elles sont faussées, donc...
149. Fabien : Non, non, elles ne le sont pas. Celle-là, elle va nous aider pour nous défendre puisque les souris au bout de vingt heures par jour pendant dix-huit mois : « aucune augmentation dans le taux de tumeurs du sein de ces souris n'a été observée », donc ça veut dire qu'elles n'ont pas eu d'augmentation de cancer.
150. Céline : Oui, mais tout à l'heure, ce n'était pas la même chose, là c'est le sein, et là, la c'est [inaudible] observée. Donc peut-être que suivant les trucs c'est pas pareil.
151. [...]
152. Nathan : De toute façon, ce n'est pas valide, elles sont génétiquement modifiées.
153. Fabien : Oui, ça, ça va nous aider... donc ça sera valide.

À deux autres reprises, une expression personnelle autoritaire de l'élève tenant un rôle de « leader » a lieu afin de conclure un accord sur un objet discuté dans le groupe. On note également qu'à plusieurs reprises des interventions conflictuelles ont lieu. Ceci indique que les relations entre élèves influent sur les objets de discours collectivement élaborés et structurent la discussion.

DISCUSSION

Différents procédés collaboratifs ont été identifiés dans les groupes de discussion. Les procédés d'élaboration collective des savoirs les plus simples sont par exemple basés sur la répétition ou l'acquiescement : un élève fait une proposition et les autres l'acceptent sans discuter. Il est parfois, à cette occasion, fait usage d'autorité. Des procédés plus complexes se manifestent dans des processus collaboratifs, par exemple lorsqu'un élève fait une proposition, reprise ou développée par un autre, et prolongée dans les successions de tours de parole. Ceci peut avoir lieu entre plusieurs membres du groupe de discussion ou en dyades au sein du groupe et à plusieurs occasions, les élèves ont adopté les rôles du jeu dans ces procédés discursifs. Lors de confrontations de points de vue opposés, les élèves peuvent dans certains cas, développer des procédés complexes pour dépasser leurs désaccords et articuler leurs positions opposées. De même, la prise en compte du discours potentiel de l'autre groupe (position opposée) permet dans certains cas le développement de justifications et une contre-argumentation collaborative entre deux élèves. Dans d'autres cas, au contraire, des confrontations peuvent conduire les élèves à abandonner l'objet de discours et provoquer des tensions interpersonnelles.

Les désaccords parviennent à être résolus par la discussion lorsqu'ils portent sur la stratégie, la formulation d'arguments, l'interprétation des textes mais conduisent à une déstabilisation de l'objet de discours lorsqu'ils sont porteurs d'une exigence épistémologique. Des désaccords sont déplacés sur un registre ontologique et résolus par des procédés collaboratifs, d'autres sont résolus sur la base de considérations stratégiques, d'autres enfin ne parviennent pas à être dépassés. Dans ce cas, les élèves peuvent exprimer individuellement leur point de vue lors de votes ou abandonner l'objet de discours. Des considérations stratégiques peuvent alors permettre (mais pas toujours) l'adoption d'une position collective.

Cette étude souligne l'importance des considérations de nature épistémologique et des connaissances profanes des élèves confrontés à une controverse socio-scientifique. L'analyse montre également que les relations entre les élèves influencent leur discussion collaborative. Les désaccords causant des tensions interpersonnelles, les élèves sont parfois amenés à adopter les positions d'avocats du jeu de rôle pour discuter leurs points de vue et cette stratégie

semble leur fournir un moyen d'éviter d'explorer leurs désaccords.

Nous avons procédé à l'analyse microscopique des échanges langagiers des élèves sans séparer la cognition du social, les connaissances et rôles sociaux étant développés par les élèves dans leurs interactions. Cette analyse a indiqué que les rôles sociaux établis dans les interactions structurent les connaissances élaborées dans les groupes.

Nous avons observé que les processus d'argumentation collaborative sont le plus souvent développés entre deux élèves dans les groupes de discussion. De plus, cette co-construction d'arguments en dyades se produit lorsque les élèves adoptent des rôles sociaux distincts, l'un agissant comme un « *leader* » et l'autre comme un « aide » qui collabore avec le *leader*. Lorsque ces élèves expriment des positions opposées ou lorsque la discussion n'est pas menée par le « *leader* », des confrontations contradictoires apparaissent et ne peuvent pas être dépassées par la discussion. À l'occasion, un élève dans un rôle d'« aide » soulève des critiques et son intervention crée une opportunité d'argumentation collaborative entre l'élève adoptant le rôle de « *leader* » et un autre élève agissant comme « aide ».

Il apparaît que les procédés collaboratifs plus complexes sont mis en œuvre dans le groupe structuré en équipe avec un leader accepté et compréhensif, l'autre groupe ayant des difficultés à réguler ses interactions sur le plan social. De même, Gayford (1992) a montré que les apprentissages individuels et la motivation sont plus importants pour les groupes ayant travaillé en « équipes démocratiques » ou sous une direction qualifiée de « négociatrice » que sous une direction « dominatrice », « dévolutive » ou selon les conseils d'observateurs critiques.

Par ailleurs, bien qu'il soit artificiel de séparer interactions sociales et connaissances collectivement élaborées dans les groupes de discussion (Moje & Shepardson 1998), nous nous centrons ici, à des fins analytiques, sur les rôles sociaux développés par les élèves dans la discussion de groupe afin de questionner différents facteurs pouvant influencer les positions sociales façonnées dans l'interaction.

On remarque que dans les deux groupes de discussion, il n'y a pas de base égalitaire dans la participation à la discussion, certains élèves ayant très peu participé aux échanges (généralement les filles). Le statut académique, estimé par les notes obtenues par les élèves, semble être un facteur d'influence dans

les rôles adoptés par les élèves dans leurs discussions de groupe. Ceci est en accord avec les résultats de Richmond & Striley (1996) : dans le groupe A, les deux élèves se trouvant exclues des discussions sont des élèves dont les notes en sciences physiques sont les plus faibles de la classe ; le *leader* est un élève de niveau élevé. Dans le groupe B, l'élève qui peut être décrit comme un *leader* contesté a obtenu la moyenne la plus faible de la classe.

Si l'on compare avec les moyennes au premier trimestre dans l'ensemble des disciplines, on constate d'autre part que trois filles ont obtenu les trois moyennes les plus élevées, ce qui semble indiquer que ce soient les notes en sciences qui constituent des facteurs d'influence importants dans les positionnements des élèves au cours de cette activité.

De plus, dans la mesure où les pratiques discursives des élèves dépendent de leur contexte de production, il convient de questionner l'impact de la situation pédagogique sur les discours des élèves. Les différences observées dans les procédés d'élaboration collective des savoirs dans les deux groupes de discussion amènent à s'interroger sur l'équité face à l'apprentissage dans ce type d'activités (Kelly *et al.*, 2001). Cela conduit à soulever la question de la stratégie pédagogique des travaux en petits groupes. Kelly et Crawford (1997) suggèrent d'insister sur le rôle médiateur de l'enseignant dans ce contexte d'enseignement particulier. Dans cette étude, l'enseignante intervient à plusieurs reprises pour préciser les consignes de la tâche et veiller à la gestion du temps. Ce faisant, elle interrompt les discussions entre élèves. Par ailleurs, Jiménez-Aleixandre *et al.* (2000) ont montré l'influence de la culture scolaire sur l'argumentation des élèves sur une question socio-scientifique. Comme le souligne Simonneaux (2003, p. 204), « les situations de débat en classe engendrent des difficultés du côté des enseignants qui doivent changer de posture : devenir des gestionnaires de débat au lieu d'être ceux qui savent. » Il est alors important de proposer aux enseignants des formations à la gestion pédagogique de discussions sur des questions socio-scientifiques.

Par ailleurs, Dawes (2004) propose de former les élèves à la pratique du langage lors des discussions de groupe afin de favoriser les discours « d'exploration » (Mercer, 1996). Dans la même veine, Grace (2005) considère que des outils de « guidage » fournissant des règles pour les discussions collaboratives seraient utiles aux élèves pour organiser leurs discussions en groupes. Nous adhérons à ces propo-

sitions et les complétons pour l'étude de questions scientifiques socialement vives. Oulton *et al.* (2004) ont précisé qu'il est difficile d'organiser un débat effectif en classe en référence à Clarke (1992) pour qui nous vivons dans un temps de déclin généralisé de discours et de civilités. Les discussions doivent alors être informées pour ne pas être seulement des « échanges d'ignorance » (Clarke, 1992). On peut ainsi penser que l'étude menée ici à des fins de recherche pourrait servir en classe un travail d'exploration avec les élèves des procédés discursifs afin d'identifier les contraintes qui interviennent dans l'appréhension de la controverse. Confrontés aux incertitudes de « la science en train de se faire » et défendant des points de vue opposés, les élèves ont par exemple tenté de résoudre leurs désaccords dans des procédures de vote. Cela interroge sur les modalités de prises de décision en contexte d'incertitude et face à des discours scientifiques contradictoires. Driver *et al.* (1996) ont montré que lors de l'étude de controverses socio-scientifiques, une minorité d'élèves considère qu'obtenir une certitude scientifique est impossible et considèrent pour la plupart que la controverse sera résolue lorsque suffisamment de données empiriques seront disponibles. Dans ce contexte, les élèves ont tendance à compartimenter les données scientifiques et l'information qu'ils utilisent pour prendre une décision personnelle (Zeidler *et al.*, 2002) et appuient leurs décisions sur d'autres considérations en attente de preuves scientifiques (Albe, 2005). Ceci ne nous apparaît pas sans lien avec une certaine conception des sciences ou une « culture de l'évidence » héritée du scientisme et du positivisme qui concourt à déléguer aux experts scientifiques la décision face à une question incertaine et controversée. Afin de favoriser une « culture de l'argumentation » (Breton, 1996), nous considérons que l'étude par les élèves de controverses socio-scientifiques doit s'accompagner de l'examen explicite en classe de la façon dont des faits scientifiques sont produits au sein d'une communauté et en particulier du rôle de la controverse et de l'argumentation dans cette élaboration. Plusieurs auteurs ont souligné le besoin d'enseignements explicites portant sur l'épistémologie et la sociologie des sciences (Bader, 2004 ; Driver *et al.*, 1996 ; Ryder, 2002). Plus particulièrement, considérer les sciences comme des pratiques sociales et des constructions rhétoriques (Osborne, 2001) avec ses incertitudes, ses limites, les intérêts en jeu ainsi que considérer le

statut, le rôle et les limites des preuves scientifiques nous semblent très important pour l'étude de questions socio-scientifiques.

CONCLUSION

Parmi les débats portant sur l'alphabétisation scientifique dans de nombreux pays, l'étude par les élèves de controverses socio-scientifiques a récemment été développée. Afin de contribuer à la compréhension de la façon dont les élèves appréhendent ces controverses, nous avons réalisé en classe une étude sur la question de la dangerosité des téléphones cellulaires avec des élèves de 1^{re} de série technologique de l'enseignement agricole. Dans une approche micro-ethnographique, nous avons analysé les pratiques discursives des élèves en groupes de discussion.

Cette étude souligne comme d'autres travaux (Alexopoulou & Driver, 1996) les inter-relations complexes entre des facteurs cognitifs, contextuels et sociaux qui influencent l'élaboration de connaissances dans ces situations. Contrairement à des recherches où les discussions entre élèves portent principalement sur des aspects procéduraux des activités (Bianchini, 1997 ; Kittleson & Southerland, 2004), les discussions portent ici sur l'étude des textes et l'élaboration d'arguments. Les élèves se sont ainsi montrés aptes à appréhender cette controverse. Pour Osborne *et al.* (2004), les questions socio-scientifiques offrent la possibilité de développer l'argumentation des élèves. Il semble qu'ici la controverse proposée a permis un grand engagement des élèves, ce qui souligne l'intérêt éducatif du traitement de questions socio-scientifiques en classe. Étant une étude de cas, il n'est pas question de généraliser les résultats obtenus, mais soulignons cependant que les analyses conduites tendent à indiquer que l'étude de controverses socio-scientifiques apparaît comme une activité porteuse de grandes exigences sur le plan épistémologique et social.

Virginie Albe

virginie.albe@educagri.fr

École nationale de formation agronomique (Toulouse)
Unité propre « Didactique des questions scientifiques,
économiques et sociales émergentes »

NOTES

- (1) Voir par exemple le programme du colloque « La crise mondiale des sciences » : Lille, novembre 2005.
- (2) La note de service du ministère de l'Agriculture et de la Pêche n° 2000-2072 en date du 18 juillet 2000 (France, 2000) précise que lors de « l'analyse contradictoire de la fiabilité des connaissances (exemples actuels des OGM, effet de serre, etc.) [...] les étudiants poursuivront à cette occasion le développement de leur conscience citoyenne [...] Les débats offriront un champ d'exercice privilégié à cette méthode ».
- (3) La leçon peut être consultée sur le site internet de l'École nationale de formation agronomique (ENFA), à l'adresse : <http://www.enfa.fr/cdrv/layout.php?ct=pc/disc-pc.htm> (consulté le 7 septembre 2006).
- (4) L'ensemble des *verbatim* peuvent être consultés sur le site internet de l'ENFA, à l'adresse suivante : <http://www.enfa.fr/cdrv/layout.php?ct=pc/disc-pc.htm> (consulté le 7 septembre 2006).

BIBLIOGRAPHIE

- AIKENHEAD G. (2003). « Review of Research on Humanistic Perspectives in Science Curricula ». Paper Presented at *European Science Education Research Association (ESERA) 2003 Conference*. Pays-Bas, 19-23 août 2003.
- ALBE V. & SIMONNEAUX L. (2003). « Procès sur les téléphones mobiles : impact sur la réflexion épistémologique d'enseignants ». In V. Albe, C. Orange & L. Simonneaux (éd.), *Recherches en didactique des sciences et des techniques : questions en débat*. Toulouse : ARDIST & ENFA, p. 253-260.
- ALBE V. (2005). « Positions d'étudiants et d'étudiantes sur une question techno-scientifique controversée : la dangerosité des téléphones cellulaires ». *Revue canadienne de l'enseignement des sciences, des mathématiques et des technologies*, vol. 5, n° 3, p. 361-376.
- ALEXOPOULOU E. & DRIVER R. (1997). « Small Group Discussions in Physics : Peer Interaction Modes in Pairs and Fours ». *Journal of Research in Science Teaching*, vol. 33, n° 10, p. 1099-1114.
- AMIGUES R. (1988). « Travail en groupe des élèves et changement de conception », *Technologies, idéologies, pratiques*, vol. 7, n° 2, p. 117-135 (« L'enseignement des circuits électriques : conceptions des élèves et aides didactiques », éd. par R. Amigues & S. Johsua).
- BADER B. (2003). « Interprétation d'une controverse scientifique : stratégies argumentatives d'adolescentes et d'adolescents québécois ». *Revue canadienne de l'enseignement des sciences, des mathématiques et des technologies*, vol. 3, n° 3, p. 231-250.
- BADER B. (2004). « Disqualification de points de vue critiques face aux sciences : procédés discursifs de jeunes du secondaire ». *Cahiers du Cirade*, n° 3, p. 5-23.
- BARNES D. & TODD F. (1977). *Communication and Learning in Small Groups*. London : Routledge & Kegan Paul.
- BARNES D. (1977). « Talking and Writing in Science Lessons ». *Cambridge Journal of Education* vol. 7, n° 3, p. 138-147.
- BIANCHINI J. A. (1997). « Where Knowledge Construction, Equity, and Context Intersect : Student Learning of Science in Small Groups ». *Journal of Research in Science Teaching*, vol. 43, p. 1039-1065.
- BILLIG M. (1987). *Arguing and Thinking : A Rhetorical Approach to Social Psychology*. Cambridge : Cambridge University Press.
- BLACK P. & ATKIN M. J. (1996). *Changing the Subject, Innovation in Science, Mathematics and Technology Education*. London : Routledge ; Paris : OCDE.
- BRETON P. (1996). *L'argumentation dans la communication*. Paris : La Découverte.
- BROWN J. S. ; COLLINS A. & DUGUID P. (1989). « Situated Cognition and the Culture of Learning. *Educational Researcher*, vol. 17, n° 1, p. 32-42.
- CLARKE P. (1992). « Teaching Controversial Issues ». *Green Teacher*, n° 31, p. 29-32.
- COULTER J. (1979). *The social Construction of Mind : Studies in Ethnomethodology and Linguistic Philosophy*. Totowa [NJ] : Rowan & Littlefield.
- DAWES L. (2004). « Talk and Learning in Classroom Science ». *International Journal of Science Education*, vol. 26, p. 677-695.
- DESAUTELS J. ; LAROCHELLE M. ; GAGNE B. & RUEL F. (1993). « La formation à l'enseignement des sciences : le virage épistémologique ». *Didaskalia : recherches sur la communication et l'apprentissage des sciences et des techniques*, n° 1, p. 79-67.
- DEWHURST D. W. (1992). « The Teaching Of Controversial Issues ». *Journal Of Philosophy of Education*, vol. 26, n° 2, p. 153-163.
- DRIVER R. ; LEACH J. ; MILLAR R. & SCOTT P. (1996). *Young People's Image of Science*. Buckingham : Open University Press.
- DRIVER R. ; NEWTON P. & OSBORNE J. (2000). « Establishing the Norms of Scientific Argumentation in Classrooms ». *Science Education*, vol. 84, n° 3, p. 287-312.
- EDWARDS D. & MERCER N. (1987). *Common Knowledge : The Development of Understanding in the Classroom*. New York : Routledge.
- EISENBERG N. (2000). « Emotion, Regulation and Moral Development ». *Annual Review of Psychology*, vol. 51, p. 665-697.
- FILLON P. & PETERFALVI B. (2004). « L'argumentation dans l'apprentissage scientifique au collège ». *Aster : recherches en didactique des sciences expérimentales*, n° 38, p. 151-184.
- FILLON P. & PETERFALVI B. (2005). « Argumentation et différenciation conceptuelle en sciences ». In *4^{es} rencontres de l'ARDIST : didactique des sciences et des techniques : contributions* : Lyon, 10-15 octobre 2005. Lyon : INRR, p. 151-158.
- FLEMING R. (1986). « Adolescent Reasoning in Socio-Scientific Issues, Part II : Non-Social Cognition ». *Journal of Research in Science Teaching*, vol. 23, n° 8, p. 689-698.

- FOUREZ G. (2002). « Les sciences dans l'enseignement secondaire ». *Didaskalia : recherches sur la communication et l'apprentissage des sciences et des techniques*, n° 21, p. 107-122.
- FRANCE : MINISTÈRE DE L'ÉDUCATION NATIONALE : « L'enseignement des sciences au lycée ». *Bulletin officiel de l'Éducation Nationale*, n° 6, hors-série, 12 Août 1999.
- FRANCE : MINISTÈRE DE L'AGRICULTURE ET DE LA PÊCHE. *Programme du baccalauréat S, rentrée 2000*. Note de service n° 2000-2072, 18 juillet 2000.
- GARCIA-DEBANC C. (1995). « Interactions et construction des apprentissages dans le cadre d'une démarche scientifique ». *Repères : recherches en didactique du français langue maternelle*, n° 12, p. 79-103.
- GAYFORD C. (1992). « Patterns of Group Behaviour in Open-Ended Problem Solving in Science Classes of 15-Year-Old Students in England ». *International Journal of Science Education*, vol. 14, n° 1, p. 41-49.
- GILLY M. (1988). « Interactions entre pairs et constructions cognitives : modèles explicatifs ». In A.-N. Perret-Clermont & M. Nicolet (éd.). *Interagir et connaître : enjeux et régulations sociales dans le développement cognitif*. Cousset [Confédération Helvétique] : Delval, p. 19-28.
- GOFFARD M. & GOFFARD S. (2003). « Interactions entre élèves et résolution de problèmes. *Aster : recherches en didactique des sciences expérimentales*, n° 37, p. 165-187.
- GOFFMAN E. (1974). *Les rites d'interaction*. Paris : Éd. de Minuit.
- GRACE M. M. (2005). « Developing Personal Values and Argumentation Skills Through Decision-Making Discussions About Biological Conservation ». Paper presented at the *ESERA Conference* : Barcelona, 28 août-1^{er} septembre 2005.
- GRISON B. (2004). « Des sciences sociales à l'anthropologie cognitive : les généalogies de la cognition située, *Activités*, vol. 1, n° 2, p. 26-34. Disponible au format PDF sur Internet à l'adresse : <http://www.activites.org/v1n2/grison.pdf> (consulté le 31 mai 2006).
- HILLS G. L. C. (1989). « Students' "Untutored" Beliefs About Natural Phenomena : Primitive Science or Commonsense ? » *Science Education*, vol. 73, n° 2, p. 155-186.
- HIND A. ; LEACH J. & RYDER J. (2001). *Teaching About the Nature of Scientific Knowledge and Investigation on AS/A Level Science Courses*. Technical Report : University of Leeds (United Kingdom). Disponible sur internet à l'adresse : http://www.nuffieldcurriculumcentre.org/go/minisite/Post16TeachingAboutScience/Page_22.html
- HIND A. ; LEACH J. ; RYDER J. & PRIDEAUX N. (2001). *Teaching About the Nature of Scientific Knowledge and Investigation on AS/A Level Science courses*. Leeds : CSSME. Document inédit, disponible sur Internet à l'adresse : <http://www.education.leeds.ac.uk/research/cssme/NuffProjReport.pdf>
- HODSON D. (2003). « Time for Action : Science Education for An Alternative Future ». *International Journal of Science Education*, vol. 25, n° 6, p. 645-670.
- JACOBI D. & PETERFALVI B. (2004). « Interactions langagières : regards didactiques, regards linguistiques ». *Aster : recherches en didactique des sciences expérimentales*, n° 38, p. 3-13.
- JAUBERT M. & REBIÈRE M. (2001). « Pratiques de reformulation et construction de savoirs ». *Aster : recherches en didactique des sciences expérimentales*, n° 33, p. 81-110.
- JIMENEZ-ALEIXANDRE M.-P. (2003). « Argumentation et opérations épistémiques au laboratoire ». *Communication au symposium Argumentation et apprentissage des sciences*, sous la responsabilité d'A. Tiberghien & Ch. Plantin. Lyon : Institut des sciences de l'homme, 7 novembre 2003.
- JIMENEZ-ALEIXANDRE M.-P. ; BUGALLO-RODRIGUEZ A. & DUSCHL R. A. (2000). « "Doing the Lesson" or "Doing Science" : Argument in High School Genetics ». *Science Education*, vol. 84, n° 6, p. 757-792.
- KELLY G. J. & CRAWFORD T. (1996). « Student's Interaction with Computer Representations : Analysis of Discourse in Laboratory Groups ». *Journal of Research in Science Teaching*, vol. 33, n° 7, p. 693-707.
- KELLY G. J. & CRAWFORD T. (1997). « An Ethnographic Investigation of the Discourse Processes of School Science ». *Science Education*, vol. 81, n° 5, p. 533-559.
- KELLY G. J. ; CRAWFORD T. & GREEN J. (2001). « Common Task and Uncommon Knowledge : Dissenting Voices in the Discursive Construction of Physics Across Small Laboratory Groups ». *Linguistics and Education*, vol. 12, n° 2, p. ; 135- 174.
- KITTLESON J. M. & SOUTHERLAND S. A. (2004). « The Role of Discourse in Group Knowledge Construction : A Case Study of Engineering Students ». *Journal of Research in Science Teaching*, vol. 41, n° 3, p. 267-293.
- LAROCHELLE M. (2002). « Rapport au savoir et socialisation à la cité scientifique ». In *Actes des 3^{es} journées d'études franco-québécoises : Paris : 17-18 juin*. Paris : université Paris 5-René Descartes, p. 58-76.
- LAROCHELLE M. & DESAUTELS J. (2001). « Les enjeux des désaccords entre scientifiques : un aperçu de la construction discursive d'étudiants et étudiantes ». *Revue canadienne de l'enseignement des sciences, des mathématiques et des technologies*, vol. 1, n° 1, p. 39-60.
- LEACH J. & LEWIS J. (2002). « The role of Students' Epistemological Knowledge in the Process of Conceptual Change in Science ». In M. Limón & L. Mason (éd.), *Reconsidering Conceptual Change. Issues in Theory and Practice*. Dordrecht : Kluwer, p. 201-216.
- LEMKE J. L. (1990). *Talking science : Language, Learning, and Values*. Norwood [NJ] : Ablex.
- MERCER N. (1996). *The Guided Construction of Knowledge : Talk Amongst Teachers and Learners*. Clevedon : Multilingual Matters.
- MILLAR R. & OSBORNE J. F. [éd.] (1998). *Beyond 2000 : Science Education for the Future*. London : Nuffield Foundation.
- MOJE E. B. & SHEPARDSON D. P. (1998). « Social Interactions and Children's Changing Understanding of Electric Circuits : Exploring Unequal Power Relations in "Peer"-Learning Groups ». In B. Guzzetti & C. Hynd (éd.), *Perspectives on Conceptual Change*. Mahwah [NJ] : L. Erlbaum, p. 225-234.
- MONDADA L. (2005). *Chercheurs en interaction*. Genève : Éd. Le savoir Suisse.
- ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) (1998). *Instrument Design: A Framework for Assessing Scientific Literacy. Report of Project Managers Meeting*. Paris : OECD.

- OSBORNE J. (2001). « Promoting Argument in the Science Classroom : a Rhetorical Perspective ». *Revue canadienne de l'enseignement des sciences, des mathématiques et des technologies*, vol. 1, n° 3, p. 271-290.
- OSBORNE J. ; ERDURAN S. & SIMON S. (2004). « Enhancing the Quality of Argumentation in School Science ». *Journal of Research in Science Teaching*, vol. 41, n° 10, p. 994-1020.
- OULTON C. ; DILLON J. & GRACE M. (2004). « Reconceptualizing the Teaching of Controversial Issues ». *International Journal of Science Education*, vol. 26, n° 4, p. 411-424.
- PATRONIS T. ; POTARI D. & SPILIOPOULOU V. (1999). « Students' Argumentation in Decision-Making on a Socio-Scientific Issue : Implications for Teaching ». *International Journal of Science Education*, vol. 21, n° 7, p. 745-754.
- PLANTIN C. (1996). *L'argumentation*. Paris : Éd. du Seuil.
- PONTECORVO C. [éd.] (1993). *Cognition and Instruction*, vol. 11 n°s 3 & 4.
- RESNICK L. (1988). « Learning in school and out ». *Educational Researcher*, vol. 16, n° 9, p. 13-20.
- RICHMOND G. & STRILEY J. (1996). « Making Meaning in Classrooms : Social Processes in Small-Group Discourse and Scientific Knowledge Building ». *Journal of Research in Science Teaching*, vol. 33, n° 8, p. 839-858.
- ROYAL SOCIETY [GB] (1985). *The Public Understanding Of Science*. London : The Royal Society.
- RYDER J. (2002). « School Science Education for Citizenship : Strategies for Teaching About the Epistemology of Science ». *Journal of Curriculum Studies*, vol. 34, n° 6, p. 637-658.
- SADLER T. D. ; CHAMBERS F. W. & ZEIDLER D. L. (2004). « Student Conceptualisations of the Nature of Science in Response to a Socioscientific Issue ». *International Journal of Science Education*, vol. 26, n°4, p. 387-409.
- SADLER T. D. & ZEIDLER D. L. (2004). « The Morality of Socioscientific Issues : Construal and Resolution of Genetic Engineering Dilemmas ». *Science Education*, vol. 88, n° 1, p. 4-27.
- SADLER T. D. & ZEIDLER D. L. (2005). « Patterns of Informal Reasoning in the Context of Socioscientific Decision Making ». *Journal of Research in Science Teaching*, vol. 42, n° 1, p. 112-138.
- SCHNEEBERGER P. & PONCE C. (2003). « Tirer parti des échanges langagiers entre pairs pour construire des apprentissages en sciences ». *Aster : recherches en didactique des sciences expérimentales*, n° 37, p. 53-82.
- SCHUBAUER-LEONI M.-L. & PERRET-CLERMONT A.-N. (1980). « Interactions sociales et représentations symboliques dans le cadre de problèmes additifs ». *Recherches en didactique des mathématiques*, vol. 1, n° 3, p. 297-350.
- SCHWEIZER D. M. & KELLY G. J. (2001). « An Investigation of Student Engagement in a Global Warming Debate ». In *Actes de la conférence NARST*, Saint-Louis, 25-28 March. Saint-Louis : NARST [CD-Rom].
- SHAMOS M. H. (1995). *The Myth of Scientific Literacy*. New Brunswick [US] : Rutgers University Press.
- SIMONNEAUX L. (2001). « Role-Play or Debate to Promote Students' Argumentation and Justification on An Issue in Animal Transgenesis ». *International Journal of Science Education*, vol. 23, n° 9, p. 903-928.
- SIMONNEAUX L. (2003). « Argumentation dans les débats en classe sur une technoscience controversée ». *Aster : recherches en didactique des sciences expérimentales*, n° 37, p. 189-214.
- SOLOMON J. & AIKENHEAD G. [éd.] (1994). *STS Education : International Perspectives on Reform*. New York : Teachers College Press.
- STRAUSS A. (1992). *La trame de la négociation : sociologie qualitative et interactionnisme / textes réunis et présentés par I. Baszanger*. Paris : L'Harmattan.
- SUTTON C. (1995). « Quelques questions sur l'écriture et la science : une vue personnelle d'outre-manche ». *Repères : recherches en didactique du français langue maternelle*, n° 12, p. 37-51.
- TOULMIN S. (1958). *The Uses of Argument*. Cambridge : Cambridge University Press. [tr. fr. *Les usages de l'argumentation*. Paris : PUF, 1993]
- UNION EUROPÉENNE (1995). *Livre blanc sur l'éducation et la formation : enseigner et apprendre : vers la société cognitive* [COM(95) 590]. Bruxelles : Commission européenne.
- WEISSER M. ; MASCLET E. & REMIGNY M.-J. (2003). « Construction de la compréhension par l'argumentation orale en sciences. Expérience menée au cycle III ». *Aster : recherches en didactique des sciences expérimentales*, n° 37, p. 17-52.
- YANG F.-Y. & ANDERSON O. R. (2003). « Senior High School Students' Preference and Reasoning Modes About Nuclear Energy Use ». *International Journal of Science Education*, vol. 25, n° 2, p. 221-244.
- ZEIDLER D. L. ; WALKER K. A. ; ACKETT W. A. & SIMMONS M. L. (2002). « Tangled Up in Views : Beliefs in the Nature of Science and Responses to Socioscientific Dilemmas ». *Science Education*, vol. 86, n° 3, p. 343-367.

Analyse des gestes professionnels : illustration de « l'œil du maquignon » chez une formatrice en expression corporelle

Monique Loquet

Cet article a pour objectif d'identifier, à travers l'étude fine d'une situation de formation continue destinée aux professeurs, la nature des savoirs pertinents à enseigner en éducation physique et sportive. Cette étude, inscrite dans le champ des recherches didactiques, vise, de plus, à élucider la manière dont s'entremêlent le choix des savoirs à transmettre et le choix des gestes de transmission. L'analyse met en évidence l'ingéniosité de la formatrice à réguler l'activité physique artistique des stagiaires. Elle montre l'intérêt pour les professeurs d'une formation aux gestes professionnels appliqués à la réalité mouvante de la classe.

Descripteurs (TEE) : création artistique, éducation physique, expression corporelle, formation des enseignants, transposition didactique.

Notre étude s'intéresse aux gestes efficaces du professeur d'Éducation physique et sportive (EPS) dans les activités physiques artistiques, référence officielle des programmes scolaires d'EPS. Réglementées sous différentes formes de pratiques au sein de l'École, ces activités développent l'expression d'un « langage poétique du corps [...] évocateur d'émotions et de sens » (France, 2001, p. 46) et participent, avec d'autres disciplines d'expression (arts plastiques, éducation musicale) à l'acquisition d'une culture artistique (France, 2001). Les savoirs pertinents à y enseigner sont peu stabilisés et leur structuration souvent problématique. Comment former à une discipline composée de savoirs au caractère ineffable ? Cette question est abordée avec les outils de la recherche didactique. Nous rendons

compte d'une pratique effective de formation continue en expression corporelle, destinée aux professeurs d'EPS. La situation de formation (1) est animée par une formatrice (2) reconnue comme l'une des meilleures expertes en ce domaine.

PROBLÉMATIQUE

Conceptions courantes de l'expression corporelle

Dans les années 1970, la discipline « expression corporelle (3) » gagne une autonomie éducative au sein de l'EPS française, en défendant des valeurs

alternatives en marge des références sportives (créativité, expressivité) et des méthodes d'enseignement a-modélisantes (respect de l'autonomie des élèves, écoute de l'autre). Dans ce cadre, elle se définit de manière large comme la manifestation de l'expression de soi par le corps à partir d'un « vécu corporel sensible » (Pujade-Renaud, 2003) sans recourir à des techniques codifiées, corporelles ou gestuelles, issus de la danse. La danse quant à elle (ou le mime) consiste à exprimer avec le corps dans un système de contraintes définies par les règles de l'activité. Aujourd'hui dans l'EPS française, il est question d'enseignement de la danse. Toutefois celle-ci reste marquée par l'antinomie technique/expression : la technique apparaît comme une « chose » apportée de l'extérieur à laquelle il faut se conformer, l'expression comme ce qui émerge de l'intimité de la personnalité ou de la corporéité. Que faut-il alors enseigner ? La technique qui apporterait des savoirs, mais serait en soi non-éducative ? L'expression qui en soi respecterait l'élève, mais serait un lieu de non-savoir ? Notre approche didactique considère que la formation « le comique par l'expression corporelle » n'est en soi, ni éducative ni modélisante, seul le traitement par la formatrice, des savoirs en jeu dans l'activité rend celle-ci éducative ou « modélisante ».

Une problématique des savoirs

Les professeurs sont désemparés face à une pratique souvent vécue comme ardue. Ils rencontrent des difficultés (fréquemment exprimées dans les termes suivants : crainte de blocages psychologiques des élèves qui « ne veulent pas dévoiler leur intimité », flottement dans l'orientation de l'enseignement « pour qu'ils ne fassent pas n'importe quoi ! »), qu'ils attribuent spontanément à la nature de la discipline. Sans sous-estimer la part des significations individuelles servant à interpréter ces difficultés, notre analyse se déplace vers les significations modelées par l'institution (Douglas, 1999). Un acteur qui aborde en profane une pratique sociale donnée est confronté à une multitude d'objets qui « n'existent pas » à ses yeux (objets matériels et immatériels, catégories de pensée, lexique et jargon de l'institution concernée) et engendrent un rapport personnel d'incertitude à cette pratique. L'expression corporelle constitue, pour les professeurs non-initiés, une *pratique ésotérique* qui crée chez eux des *besoins en savoirs professionnels*. L'expression « savoir professionnel » est ici utilisée dans un sens large qui inclut l'expérience, les connais-

sances et la puissance d'agir de la formatrice, mobilisées lors des échanges avec les stagiaires. Notre problématique est liée à une volonté « d'extériorité » vis-à-vis de la pratique, autrement dit, de reconnaissance des besoins en savoirs professionnels, afin que « *l'ésotérisme de la pratique cède du terrain devant l'exotérisme obligé des savoirs* » (Chevallard, 1994, p. 174). C'est pourquoi nous souhaitons élucider, à travers l'étude de la situation, les dimensions essentielles de la pratique de formation en expression corporelle. Par ailleurs, estimant que conduire une situation à vocation didactique (formation des professeurs, enseignement scolaire) exige certes, des qualités pédagogiques et psychologiques, mais repose avant tout sur des choix didactiques, nous observerons la manière dont s'entremêlent les choix des savoirs corporels à transmettre et des gestes de transmission.

Étudier les gestes professionnels de transmission

Notre analyse se focalise sur le processus de circulation des savoirs relatifs à « l'expression corporelle » conduit par la formatrice. Les gestes professionnels produit en situation contiennent une part importante d'ingéniosité didactique qui recèle des manières de faire éprouvées et cependant mal connues (Amade-Escot, Loquet & Léziart, 2003). Nous employons le mot geste, au-delà du sens restreint « mouvement du corps, mimique », dans un sens extensif (à la suite de Chevallard, 1999) proche du substantif gestion emprunté au latin *gero* ; *gestum* « prendre en mains, administrer ses affaires ». Chevallard identifie dans l'action professionnelle, un processus de naturalisation qui tend à transformer les *tâches* accomplies au quotidien, en simples *gestes* intériorisés à force d'expériences. Si certains gestes de la formatrice faisant l'objet d'explications aux stagiaires, échappent à la naturalisation, d'autres ne trouvent plus de justification et paraissent incongrus. Leur caractère quasi-automatique peut donc être source de difficultés pour l'enseignant novice, confronté à l'épaisseur de la pratique et à l'opacité des raisons d'agir du formateur. Afin de rendre compte plus finement des manières de faire ingénieuses de la formatrice, nous complétons le terme *geste* (manifestations de la dynamique de l'action : ce que fait la formatrice) par celui d'*œil du maquignon* ou d'*œil expert* (manifestations de la dynamique de l'observation : ce que regarde la formatrice).

Enjeux et questions de la recherche

Nos enjeux de recherche sont à la fois opérationnel, donner aux gestes didactiques une reconnaissance culturelle et un statut d'objet de formation professionnelle, et scientifique, contribuer à la mémoire et à l'identification des savoirs relatifs aux métiers didactiques en activités physiques. Dans le prolongement des travaux déjà engagés (Loquet, Garnier & Amade-Escot, 2002), nous étudions les gestes professionnels de régulation (4), à travers trois questions (5) : (i) quelle progression des savoirs au fil des séquences (régulation *chronogénétique*) ? ; (ii) quelle évolution des responsabilités vis-à-vis des savoirs (régulation *topogénétique*) ? ; (iii) quelle modification du milieu pour apprendre (régulation *mésogénétique*) ? L'application du terme *genèse* aux trois descripteurs, souligne notre intention de saisir le devenir des savoirs du corps enseignés et appris au fil des séquences, c'est-à-dire la *gestion des transformations* de l'activité physique artistique des stagiaires.

MÉTHODE

Le stage « le comique par l'expression corporelle » comprend six séances de trois heures. Parmi l'ensemble des matériaux recueillis, nous nous intéressons à ceux de la quatrième séance dont la position au cœur du stage garantit une densité didactique. Nous avons sélectionné une situation transformatrice de la motricité des stagiaires, particulièrement intéressante pour la compréhension des gestes professionnels. Placée en début de séance, elle développe le thème de « l'exagération comique » en utilisant des musiques « dynamogéniques ». Elle présente, en outre, une efficacité immanente à la situation et un condensé d'interactions, dont l'enchevêtrement des différents registres est quasi-indémêlable « à l'œil nu ». Nous évoquons deux phases méthodologiques de notre recherche : la collecte et le traitement des traces (Schubauer-Leoni & Leutenegger, 2002). La collecte des traces suit cinq étapes :

- (i) Identifier le projet de formation (entretien *ante*). Cet entretien avec le professionnel, avant la séance, vise à cerner ses intentions et à anticiper le fonctionnement en séance (décisions prévisibles, repérage de conduites des stagiaires) ;
- (ii) Analyser l'activité régulatrice de la formatrice dans ses interactions avec les stagiaires en situation (enregistrements audio-vidéo des séquences) ;

- (iii) Effectuer l'analyse *a priori* des savoirs en jeu dans la séquence (enregistrements audio-vidéo). Le chercheur procède par inférence à partir de sources expertes ;

- (iv) Analyser les réactions à chaud du professionnel (entretien *post*). Cet entretien, après la séance, revisite les moments passés, à travers les mots et sensations de l'acteur, dans un but de collecte complémentaire à l'observation *in vivo* (et non dans un souci de mesurer l'écart au projet initial) ;

- (v) Explorer la « mémoire » de l'expérience vécue par les stagiaires (mode écrit et oral). Ce recueil pendant ou à l'issue de la situation, vise à re-construire les données du côté des stagiaires (et non à vérifier les effets de formation).

Le grain descriptif de la situation est le plus fin possible : les paroles des acteurs sont transcrites avec les changements de ton de la voix, accompagnées des mimiques, regards, postures, déplacements, gestuelles segmentaires (6). Nous disposons ainsi d'un ensemble d'observables que nous traitons selon les principes de questionnement réciproque et de rétroaction (Loquet & Amade-Escot, 2000 ; Schubauer-Leoni & Leutenegger, 2002).

Nous rendons compte de la pratique formatrice en croisant deux niveaux de description. Au premier niveau, l'activité régulatrice est envisagée selon « *la sémantique naturelle de l'action* » (Sensevy, Mercier & Schubauer-Leoni, 2000, 276), c'est-à-dire tel que le ferait un familier de l'action. Au second niveau, elle est analysée à travers d'une part, quatre catégories descriptives de l'action du professeur – définir, réguler, dévoluer, instituer, selon le modèle emprunté à Sensevy, Mercier & Schubauer-Leoni, 2000 ; Sensevy, à paraître –, d'autre part, les trois dimensions génétiques des savoirs en jeu (avancée temporelle des savoirs, dynamique des responsabilités, évolution des objets du milieu). Cette double description imbriquée s'efforce de respecter la pratique (analyse *ascendante* grâce aux catégories susceptibles d'être utilisées par le formateur) et de tester le modèle théorique (analyse *descendante* grâce aux catégories théoriques).

RÉSULTATS

Les résultats se situent à l'articulation de deux temporalités :

- (i) la *temporalité linéaire de l'action* (tableau I) découpée selon trois critères : modalité d'interven-

Tableau I. – **Déroulement temporel de la situation « exagération comique » (vingt-deux premières minutes) et découpage selon trois critères**

Tps min.	Modalités d'intervention	Indices de coupure des tâches et sous-tâches	Tâches effectives (ce qu'il y a à faire)
1	Groupe en dispersion	F : je commence par la fin... je n'en dis pas plus F : vous êtes curieux / faites comme vous voulez	vous êtes des gens curieux
2			
3		La musique démarre...	le curieux est obstiné
4			
5		F : terminé ↓ malheureusement il va falloir le reprendre/	rythmique obsessionnelle tête motrice espaces de regard géométriques
6			
7		F : maintenant que... dites-vous que...	
8		F : ton tempo est... voilà tu t'es trouvé un tempo	
9		F : ta tête est... voilà tu y es...	
10		F : alors, maintenant donnez-vous...	
11		F : arrêtez-vous une minute...	
12	Par deux en dispersion	F : vous allez vous mettre par deux... vous manipulez	manipulez la tête d'un copain
13			
14			
15		F : maintenant qu'on vous a bien manipulé...	
16			
17		F : stop ooh mesdames et messieurs / je suis triste	la curiosité motive les mouvements
18	F : il y a un thème c'est... allez c'est parti ↓		
19			
20	F : tu y es ↓		
21			
22	S : on ressemble à un singe... F : bon / c'est bon ↓	jouez la curiosité	

tion du formateur, tâche effective, indice de coupure à l'initiative d'un acteur. Ce découpage (dans les termes des professionnels) scande la succession chronologique des activités occupant collectivement la formatrice et les stagiaires ;

(ii) la *temporalité des savoirs* en jeu dans la situation, beaucoup plus sinueuse, faite de tours et détours : les méandres de la circulation des savoirs sont reconstitués à l'aide du modèle théorique.

La situation (trente-trois minutes) est reconstituée en trois phases. Après avoir présenté les traits essentiels de la phase initiale (cinq minutes), nous développons la phase centrale (dix-sept minutes) où se

concentrent les gestes régulateurs. Ces gestes et regards sont interprétés à partir des expressions originales des acteurs en situation (encadrés).

Définir la situation et la dévoluer aux stagiaires (5 min.)

Dans cette phase initiale, la formatrice assume une double nécessité : la *définition* et la *dévolution* de la situation, visant respectivement à délimiter l'activité des stagiaires et à les responsabiliser vis-à-vis des savoirs enseignés. Nous suivons le fil des interactions.

Annnonce du projet et réticence

Le projet de « transmettre » des connaissances est d'emblée annoncé, mais celles-ci ne sont pas nommées (sec. 20) :

F : *je commence par la fin ↓ / c'est-à-dire par trois aspects qui sont une synthèse d'éléments qu'on n'a pas encore vus ou qu'on a déjà vus [...] faut comprendre là hein ↑ [rire] donc ↓ je commence par la fin ↓*

Cet énigmatique « commencement par la fin » trouve deux sources : (i) didactique, la parole du formateur s'efface devant l'importance accordée au faire des stagiaires. Les connaissances sont donc gardées au secret (7) ; (ii) épistémologique, il est une image de l'activité que la formatrice souhaite transmettre, celle où les « éléments » (dissociation segmentaire : bras, buste, tête...) comptent moins que la « synthèse » : exprimer corporellement, selon Matho, « c'est avant tout **représenter** donc il faut aller à l'essence / sinon on fait de la gym en musique [évoque les cours de gymnastique d'entretien] » (entretien post).

Cadre du jeu : l'essentiel

Le cadre initial du jeu est énoncé (min. 1) :

F : *je vais vous mettre une musique qui est absolument a-ga-çan-te [...] vous êtes des gens curieux cu-ri-eux et ces personnes curieuses vont tomber peu à peu dans l'obsession de la curiosité [...] je vous laisse pour le moment je n'en dis pas plus / vous savez simplement que vous avez le droit d'aller à l'exagération comique ↓ non je ne vous en livre pas plus / vous vous débrouillez avec ça.*

La définition du jeu combine les deux origines : (i) didactique, le cadre laisse de la liberté aux stagiaires et la responsabilité d'une recherche autonome, soutenue par la musique, à l'endroit où ils doivent s'exprimer. La dévolution est une condition d'entrée dans le jeu ; (ii) épistémologique, le cadre cerne ce qui est l'essentiel, dont l'analyse *a priori* met à jour les enjeux : l'action sollicitée est sémiocinétique et s'exerce à propos d'un double objet, l'un objectif et descriptif (être un curieux), l'autre subjectif et à charge émotionnelle (être en état obsessionnel).

Partage de références

À qui pourrait s'inquiéter d'un éventuel désarroi des stagiaires (comment exprimer corporellement une « obsession de curiosité » ?), l'analyse montre que

cette entrée n'est lâche qu'en apparence. Elle est sous-tendue par deux types de *références communes* partagées par le groupe. La première référence renvoie à des comportements inscrits dans la vie courante (« les gens curieux ») que chacun peut imaginer sans mal (elle éveille des impressions d'enfance chez certains stagiaires qui d'emblée miment l'enfant fouineur). La « curiosité » attendue est toutefois illustrée par Matho, en sous-main, au travers d'indications motrices précises (min. 2) :

F [semblant chercher quelque chose autour d'elle] : *vous êtes cu-ri-eux [répète en articulant] / quoi ↑ quoi ↑ quoi ↑ [mime un fureteur, tourne la tête à droite, à gauche]*

La seconde référence rappelle un passé didactique commun :

F : *comme on a déjà vu / ménager peu à peu vos effets / ne donner jamais tout de suite / vous vous souvenez ↑ il faut que ceux qui regardent se disent mais où ça va aller tout ça ↑*

Le partage de références (imaginaire, « déjà vu ») allège la définition du jeu, évitant aux stagiaires d'y entrer par le désordre ou la surcharge.

Place aux stagiaires

La suspension des gestes de définition laisse la place aux stagiaires : place à l'étonnement, questionnement ou « manque à savoir » délibéré ; place surtout à l'action motrice (min. 3) :

F : *il faut aller jusqu'au bout de ce qu'on va appeler l'exagération comique*

S : *en parole ↑*

F : *non y'a pas de parole ↓*

S : *ah en geste ↑*

F [voix théâtrale] : *je n'en sais rien si c'est des gestes / de la ges-tu-elle [gestes des mains] ↑ ou du cor-po-rel [ondulation du buste] ↑ discussion ↑*

F [la musique se développe] : *vous êtes cu-ri-eux ↓ faites comme vous voulez / allez répartissez-vous ↓*

Par ce geste dévolutif, la formatrice se retire du milieu et incite les stagiaires à l'occuper. Chacun explore l'espace des solutions expressives dont il dispose. Nous en décrivons quelques traits. Les stagiaires « font » les curieux, déployant une activité que nous qualifions de réaliste. Ils mènent une explora-

tion assidue de ce qui est caché réellement à leurs yeux, passant d'une cachette à l'autre (l'un d'eux fouille dans un sac, un autre inspecte le dessus d'une armoire). Nous suivons en particulier deux stagiaires, nommés ici Ber et Jac, qui lors des séances précédentes restaient en périphérie du groupe : Ber se laisse visiter les poches, Jac ausculte les tiroirs d'un bureau. En fin d'évolution musicale, Matho « reprend la main » s'exclamant telle une comédienne (min. 4) :

F : *aaah mal-heur* [voix théâtrale, main sur le front] / *je crois qu'il va falloir le* [l'exercice] *reprendre* / [voix professorale] *ah mais j'ai fait exprès* ↓
 S : *avec des consignes* ↓
 F [acquiesce] : *on va se donner des p'tites consignes* ↓

Une reprise de définition plébiscitée par les stagiaires, s'impose. C'était anticipé et même provoqué : les stagiaires jouent bien le jeu, mais ne jouaient pas tout à fait au « bon » jeu.

Collaboration didactique

La définition du jeu serait inopérante sans la *collaboration* des stagiaires. Grâce à leurs réalisations coopératives, même approximatives, le formateur va désigner la « bonne » direction escomptée. L'exemple suivant décrit la manière dont Matho utilise le concours de Jac pour recentrer la définition et faire avancer le temps didactique (8). Manifestement quelque chose dans son activité est emblématique de ce qui ne doit pas être conservé (min. 5) :

F [joue l'intriguée et interpelle Jac qui fouine encore en vrai curieux] : *qu'est-ce qu'il fait lui* ↑
 S [la classe se tourne vers le bureau] : *ben il est curieux* ↓
 F : *ah oui* ↓ / [fait mine de s'excuser] *je sais que je commence par la fin que j'ai pas fait les préalables* ↓

Déchargeant « son » collaborateur d'une éventuelle culpabilité liée à l'échec, elle classe sa solution sans l'exclure :

F [en aparté à Jac] : *tu me fais une autre sorte de comique mais pas celui que je voulais / mais c'en est un autre d'accord* ↓

La formatrice emmène alors la classe tâtonner dans un milieu moins réaliste. L'objet de savoir est objet de devinette (9) :

F : *aaah* ↓ / *c'est comme un jeu de de-vi-nette* ↓ / *le-cu-ri-eux* ↓ / *dites-vous c'est un obstiné* ↓ [désarticule mains, buste, tête dans différentes directions spatiales] / *tout-l'in-té-resse-au-ni-veau-cu-rio-sité* [mouvements de tête saccadés, tempo scandé de la voix] / *en fait* [ton confidentiel] *c'est presque un malade mental* / *c'est un ob-se-ssio-nnel* [se tapote le front en cadence] / *commencez à vous déplacer en vous disant / il est cu-ri-eux-de-tout* [tempo de la voix] *allez-y* ↓

La formatrice donne des indices corporels d'entêtement. Accentués par le tempo de la voix et la cadence des mots, ils rendent désormais superflu tout support musical. Cet ensemble d'indications élève l'activité créatrice des stagiaires au diapason de la formatrice. À ce moment, les gestes de définition sont jugés suffisamment clairs pour chacun : la dévolution s'opère. Les stagiaires entrent dans la direction de recherche indiquée par Matho. Il s'agit bien de : (i) déréaliser le travail de curiosité vers un gain d'abstraction ; (ii) convertir l'idée d'obsession en mouvements corporels.

Réguler la situation (17 min.)

Dans cette phase centrale, une nouvelle nécessité est assumée : la régulation des *rappports* qu'établissent les stagiaires au jeu préalablement défini et dévolu. Elle vise à développer les stratégies créatives adéquates. Nous suivons le fil des interactions :

Trois règles stratégiques

Un emmêlement de règles surgit *in situ* (min. 6) :

F [évolue dans le groupe en dispersion] : *curieux de tout en-haut-à-droite-en-bas-derrière* [tempo de la voix et mouvements de tête dans les quatre directions] / [...] *quand il y a obstination il y a régularité-ryth-mique* ↓ [répétition cadencée] / *ta-ta-ta* [accentue chaque pas marché avec conviction] / [...] *maintenant que vous avez votre tempo* ↓ *vous vous dites* ↑ *c'est votre tête qui est toujours le moteur du mou-ve-ment* / [marche et voix cadencées, tête impliquée dans la cadence]

Variété spatiale des mouvements, régularité rythmique de la marche, impulsion du mouvement par la tête : ces trois règles renvoient, à l'évidence, aux « trois aspects » tenus secrets par la formatrice. Inspirées des principes d'enseignement morphocinétiques en usage dans les arts de scène, elles constituent

son *plan d'organisation de l'expression corporelle* (Fromantel, Bertrand & Dumont, 1968). Ainsi le jeu du curieux exige que sur un *tempo* régulier de locomotion, les stagiaires improvisent des mouvements de la tête – siège de la caricature du curieux – sur lesquels s'insère la notion d'espace comme étendue géométrique à investir.

Avoir l'œil et accompagner

Une forte propension à l'*observation et l'accompagnement* des stagiaires se manifeste chez la formatrice sans cesse sur le qui-vive. Les gestes formateurs trouvent à nouveau deux origines : (i) nécessité didactique, la formatrice, garantissant aux acteurs une forte autonomie, assure en même temps une proximité didactique avec chacun. Les trois règles stratégiques (*tempo*, tête, espace) sont reprises une à une, en lien avec ce qu'elle leur voit faire ; (ii) conception épistémologique, le travail des trois règles n'entame pas l'unité de la « synthèse », il y contribue. Ainsi l'accompagnement se prolonge à travers chaque transition : dès qu'une règle est potentiellement maîtrisée (« *oui ça vient c'est bon* »), la suivante est abordée sur-le-champ (« *alors maintenant...* ») s'unissant à la précédente.

La première règle (*tempo*) consiste à produire une rythmique : elle permet de se dégager du réalisme expressif pour suggérer l'obsession (min. 8) :

F [marche au sein du groupe, en bon guetteur observe chacun] : [emboitant le pas d'une première stagiaire à allure ordinaire] *ton tempo est un peu lent / tu es trop proche de la réalité / faut un tempo plus obsessionnel tu vois* ↓ [à tous] *on quitte la réalité* [accentue le tempo des mots, des pas et des mains] *c'est ob-ses-sio-nnel voilà hop hop //* [tapote sur l'épaule d'un second stagiaire] *attention là tu as cassé ton tempo / tu joues naturaliste //* [accompagne la marche d'un troisième] *oui ça va / tu as un tempo c'est bon //* [à tous sans rompre son tempo d'accompagnement] *tac-voilà-cha-cun-s'est-trouvé-un-tempo-oui-c'est-bon* [articulation syllabique, gestes cadencés des avant-bras, index déictiques battant la mesure].

En vertu de cette indissociabilité que nous voyons fermement assurée, *tempo* d'accompagnement didactique et *tempo* d'expression corporelle, il s'ensuit une transition sans rupture vers la deuxième règle (tête motrice) et une combinaison avec celle-ci (min. 9) :

F [déambule en cadence parmi les stagiaires, toujours à l'affût] : *la-tête-est-moteur-du-mouvement* ↓ *si vous tournez* ↑ [« décroche » la tête vers l'arrière et entame un demi-tour] [...] *si vous baissez* ↑ [esquisse par la tête une descente du corps vers le sol] [...] *si vous regardez en l'air* ↑ [impulse par la tête un étirement du corps vers le plafond] [...] *c'est la tête qui entraîne le mouvement* ↓ / *la tête est lo-co-mo-ti-ve-de-tout* [entrechoque les mots, index pointé appuyant les gestes, en mesure] *et non pas vos pieds* ↓ *vous y êtes* ↑ [...] [guette la production d'un premier stagiaire conforme au sens indiqué] *voilà ça y est tu y es là* [pose la main sur l'épaule d'un second stagiaire arrêté] *si tu démarres pour marcher c'est la tête qui va entraîner ton démarrage* [...] [pointe l'index vers les jambes d'un troisième stagiaire s'apprêtant à tourner] *ne tourne jamais parce que tes jambes veulent tourner c'est la tête qui te fait tourner* ↓ [regard circulaire sur le groupe] *oui ça vient c'est bon /*.

De la même façon sans rompre les tempi, est intégrée la troisième règle : variété spatiale du mouvement (min. 10) :

F [enchaine deux mouvements corporels avec changements de direction impulsés par la tête vers l'arrière, à droite puis à gauche] : *alors maintenant donnez-vous des directions plus compliquées / il y a eu le-haut-le-bas-devant-derrière / prenez des obliques bas / des obliques haut* [plonge dans les directions indiquées] *géo-mé-tri-que-ment / variez vos espaces de regard géométriques / ça vient oui c'est bon* ↓

La formatrice oriente ainsi le travail du groupe vers une complexification « douce » de l'activité corporelle.

Entrée sur le territoire d'un stagiaire

Si la « synthèse » dépasse par ses exigences, les règles analytiques qui la composent, les trois règles possèdent chacune les marques de l'ensemble dont elles se détachent. Or la confrontation aux règles entraîne des résistances motrices chez les stagiaires. La formatrice va s'efforcer de les débusquer. Ses gestes régulateurs s'individualisent. Dans l'exemple suivant, Matho entre plus avant sur le « territoire » d'un stagiaire. Ber éprouve des difficultés à « descendre au sol » (c'est-à-dire à se déplacer au sol avec des appuis autres que les pieds ; les « descentes »

ou « passages au sol » sont des thèmes de travail habituels en expression corporelle) selon les deux règles, tête motrice et espaces de regard géométriques (min. 11) :

F [à Ber amorçant une « descente au sol »] : *tes pieds te conduisent ↓ tes pieds n'ont pas le droit de te conduire / c'est la tête ↓ qui entraîne ton corps /*

Ber [fait une nouvelle tentative] : *mais je ne vois plus rien ↓*

F : *mais **noon** ↓ là c'est ton corps qui a **baissé** ↓ c'est **ta tête** qui entraîne ton corps ↓*

Ber : *ben oui mais si je veux descendre ↑*

F : *c'est la tête qui va t'entraî / [s'interrompt tout à coup et claironne dans la salle] tututut arrêtez-vous une minute [prend à la volée la décision de suspendre le travail individuel]*

La formatrice perçoit dans la difficulté éprouvée par Ber, la présence d'un *obstacle* typique de la motricité (10). Ainsi elle confirmera qu'elle « *avait l'œil sur ce point faible du groupe* » analysé comme « *problème sur la dissociation corporelle* » et affronté intentionnellement : « *je l'ai vu j'ai fait exprès aussi cette cassure tout à coup* » (entretien post).

Nous identifions l'obstacle comme suit (analyse a priori) : la règle « tête motrice » exige un modification de la fonction habituelle du regard et provoque un déséquilibre postural, vécu négativement par les acteurs. Habituellement, le regard fournit aux individus des repères extérieurs stables pour abaisser le corps jusqu'au sol : il ajuste avec précision le placement du corps au dessus de l'appui des mains ou du genou au sol. Ici la tête et, avec elle, le regard sont impliqués dans une création morphocinétique ; ce dernier ne peut donc contrôler la descente corporelle jusqu'au sol et les informations sensori-motrices usuelles sont insuffisantes pour suppléer à la perturbation visuelle. La tâche prescrite par la formatrice (impulser le corps par la tête), d'apparence anodine, est en réalité contre-intuitive, d'autant qu'elle s'effectue dans des espaces non habituels, hors de l'espace familier de locomotion.

L'examen avisé du comportement de Ber permet de réguler l'activité collective. Cette fois la régulation, explicitement ciblée sur l'obstacle, est mésogénétique. Jugeant sans doute que demander oralement la modification ou même la montrer, ne suffira pas à l'obtenir, la formatrice opère l'aménagement du milieu (mésogénèse) suivant :

F [à toute la classe] : *vous allez vous mettre par deux [...] / viens [place ses paumes de main sur les tempes d'un stagiaire] je te tiens la tête / [au groupe] pour lui faire comprendre je lui fais hop la tête va là-bas [impulse un changement de direction par la tête] / [au stagiaire manipulé] maintenant tu vas regarder en bas et c'est ta tête qui va entraîner le corps d'accord ↑ [au groupe] manipulez-vous vous pensez haut-bas-devant-derrière-à-droite-à-gauche-oblique-oblique-oblique-oblique [pointe le doigt dans les dix directions citées] / c'est déjà pas mal [rire] hein ↑*

L'aménagement du milieu introduit une aide tactile extérieure comme moyen de faire *ressentir* aux stagiaires l'impulsion isolée de la tête et systématiser l'exploration dans les directions spatiales inaccoutumées. Cette régulation mésogénétique s'accompagne d'un changement de position (*topos*) du formateur qui le place en « position haute » et distancie vis-à-vis des stagiaires (surplomb topogénétique) pour mieux évaluer leur activité. Ainsi, alors que l'aide manipulatoire tend à dévier les stagiaires de l'activité expressive, la formatrice maintient le contrôle de la définition initiale du jeu (min. 12) :

F [accompagne un duo dont le manipulé « s'abandonne » à la manipulation] : *la tête est lo-co-mo-tive [chantonne] en pensant **quand même** que vous êtes **cu-ri-eux***

L'étau se resserre sur l'essentiel

La formatrice, toujours en position de guetteur, juge que le thème « curiosité » est momentanément déstabilisé (en tant qu'objet du milieu) et a cessé d'exister comme organisateur de l'activité des stagiaires. Elle impose, dans une mise en scène théâtralisée, le retour du thème dans le milieu (min. 18) :

F [tel un Monsieur Loyal donnant la réplique à des clowns en piste] : *oo'oh je suis **triste** ↓ / oooh **mesdames et messieurs** je suis triste ↓*

S [déçus] : *noon ↓*

F [voix théâtrale] : *il y a un thème c'est la cu-ri-osi-té*

S [se remémorent] : *aaah ↑ oui ↓*

F : *et je vous vois faire [voix théâtrale] comme des profs de gym' des mouvements de tête / [change de tonalité avec fermeté] c'est pas bouger la tête pour la bouger / [voix professorale] c'est **parce que** je suis **cu-ri-eux** [scande le mot avec les mains] n'oubliez pas le thème / vous êtes cu-ri-eux ↓ [aux aguets cherche autour d'elle].*

S : *mais les pieds* ↑

F : *aaaah* [crescendo théâtral] *le pied est le symbole de votre obsession là* ↓ *tatatata* [mime le curieux par un mouvement de tête et son obsession par une marche cadencée]

C'est le moment d'une accélération chronogénétique décisive. La formatrice attend « *une exigence de qualité de jeu [...] il faut jouer **juste** / redevenir naturel dans le thème [...] je veux que ça soit d'une grande **lisibilité** / c'est-à-dire que peu à peu tu dois enlever tous les mouvements superflus pour garder **l'essentiel** / alors là* ↑ *c'est là **l'état** / ça commence à se resserrer* » (entretien *post*). L'échange suivant entre Matho et les stagiaires révèle ce que cet essentiel implique : affronter le problème de la *conversion sémiocinétique*. Ce n'est pas « du » mouvement qui doit être donné à voir, c'est une expression scénique de l'état de curiosité (min. 19) :

F [à la classe, appuyant la main sur l'épaule d'un stagiaire] : *le mouvement est motivé par* ↑

S [poursuit la phrase] : *la curiosité* ↓

F : [voix théâtrale] *il a gagné/ le mouvement est motivé par la curiosité* [se reprend vivement] *ooh non* ↓ [voix professorale] *c'est faux / c'est la **curiosité** qui motive le mouvement / **invertissons** le problème* ↓

La formatrice par ses régulations insiste sur la qualité du « jeu », sans abandonner les règles morphocinétiques préalablement échafaudées. Toujours « en piste » au centre du groupe et aux aguets, elle s'assure que les stagiaires jouent bien en acteur la saynète du curieux, tout en gardant une vigilance sur les moyens de l'action travaillés antérieurement (min. 21) :

F : ***jouez le** / faut être **curieux** / voilà c'est bon* [à un premier stagiaire] *attention c'est la tête qui entraîne ton corps* / [pose le doigt sur l'épaule de plusieurs stagiaires au passage] *voilà ça vient / tes pieds sont après ton corps hein te trompes pas [...]* *ça y est* ↓ *c'est bon* ↓ / ***joue bien** la curiosité quand même* ↓ [approuve les réalisations] *t'es curieux / [chantonne comme une ritournelle enfantine] **cu-ri-eux cu-ri-euse*** ↓

Finalement la mutation mouvement/thème s'opère, comme l'indique l'échange suivant avec Ber, toujours en position de collaborateur didactique (min. 22) :

Ber : *on se disait finalement qu'on ressemblait à un singe / en fait c'est p'têt ça la base*

F [approuve] : *ça s'appelle l'exagération comique* ↓

Ber : *de ressembler à un singe* ↑

F : *éventuellement oui* ↓

Ber : *les singes si on les regarde bien / ils sont obligés de regarder autour d'eux* ↓

F : *oui / on est dans l'exagération comique* ↓ [à ce moment la formatrice lève le bras et arrête le travail] *bon c'est bon* ↓

La boucle est bouclée : l'action motrice créatrice a été mise en route par le thème de la curiosité (min. 1), c'est au tour du thème de « prendre le dessus » sur le mouvement (min. 22). Les gestes régulateurs du formateur sont alors suspendus.

DISCUSSION

Notre objectif était de dévoiler l'ingéniosité d'une formatrice à travers l'étude minutieuse de ses gestes régulateurs en situation. Celle-ci nous porte à trois types de remarques :

Une triple mise en scène

Nous constatons, tout au long de la situation, que Matho assume trois positions : comédienne, professeur et formatrice. Les gestes régulateurs liés à chaque position se distinguent assez facilement, car ils possèdent une identité propre : (i) ceux de la comédienne se préoccupent du jeu de l'acteur-mime ; (ii) ceux du professeur visent à transformer la motricité des stagiaires (qu'elle prend « *comme des enfants* ») ; (iii) ceux de la formatrice s'attachent à transmettre ce qu'elle appelle la « *la théorie pratique* » des choix d'enseignement. Mais ce n'est pas la superposition de ces gestes professionnels variés qui constitue la « substance » didactique. Nous observons leur combinaison subtile, sorte de fusion ordonnée par la communication des savoirs « d'expression corporelle ». Plus qu'une mécanique gestuelle, il s'agit d'une dynamique interactive, où chacun des acteurs collabore de sorte qu'aucun n'arrive aux savoirs en jeu sans la coopération de l'autre.

Ce constat de complexité nous amène à questionner les conceptions « naturalistes » de l'appren-

tissage. Pourrait-on apprendre « l'expression corporelle » comme on apprend sa langue maternelle par imprégnation « naturelle » ? Suffirait-il d'une participation progressive à cette activité, sans décalage de méthode ou de vocabulaire, et sans travail transpositif ? Dans un débat avec les stagiaires portant sur « l'improvisation », la formatrice tranche : « *surtout ne dites jamais aux élèves / aujourd'hui on s'exprime ↓ sujet d'improvisation ↑ on va faire le clown ou le burlesque ↓* ». Nous affirmons à la suite de Brousseau (1996) que « *le milieu culturel ne peut être "naturellement" enseignant sans être localement didactique* » (1996, p. 78). Expert d'une APSA, le formateur ou le professeur d'EPS ne peut se passer de situation didactique. Ainsi que le montre notre étude, il ne suffit pas à Matho d'être artiste et de se comporter comme telle avec les stagiaires (en comédienne montrant les émotions et les comportements qu'elle veut mettre en scène). Si Matho se contentait de produire elle-même ses questions et réponses d'actrice, elle proposerait des contenus que les stagiaires ne pourraient jamais atteindre et les priverait d'une puissance d'agir. Elle leur fait donc de la place, leur donne du temps, les confronte à un milieu où ils rencontrent des ignorances et où les questions n'ont *a priori* pas de réponse. L'artiste assume alors les fonctions didactiques de professeur.

Plan d'organisation et improvisation didactique

La formatrice s'en tient non seulement à un plan d'organisation bien éprouvé, mais trouve également dans ce canevas, en situation, une liberté d'*improvisation didactique*. Matho s'appuie sur des gestes didactiques qui intègrent le stagiaire dans sa propre démarche et tiennent à la fois (selon ses propres expressions) de la « *marche de l'équilibriste* », de la « *ruse du renard* », du « *sac à malice* » et de « *l'œil du maquignon* », comme l'indique cet extrait de l'entretien post : « *c'est le maquignon qui sans arrêt voit le point fort et le point faible / il rassure sur le point fort sans jamais mettre le doigt sur le point faible / mais il utilise le point faible pour amener d'autres interventions [...] si par exemple il a un cheval qui a du mal à trotter / il sait qu'il a des qualités athlétiques / il va lui faire faire autre chose pour l'amener à trotter* ».

Confronté à une réalité multiple et changeante, « l'œil » s'ouvre dans toutes les directions à la fois : « *c'est sans arrêt le maquignon qui essaie d'être entièrement en regard ouvert sur les gens en permanence / oublier qu'on existe en tant que professeur / c'est-à-dire que toi tu n'existes pas ↓ sans arrêt tu es*

en train / plus ils sont nombreux plus c'est fatigant ↓ mais plus tu vas aller regarder leurs points forts leurs points faibles » (entretien post).

Ne pouvant saisir le groupe dans une forme intelligible immédiate et unique, la formatrice se montre elle-même plus mobile et « roublarde » encore que ses stagiaires : « *il faut être comme des renards on ne va pas attraper des poules en plein jour et à midi ↓ faut comme les posséder / on fait leur bonheur malgré eux ↓ [...] par exemple les stagiaires / en tant que maquignon je peux dire / ils aiment **jouer** ↓ donc je n'ai rien fait au niveau du **jeu comédien** / comme on dit la bête est bonne / par contre ils ont eu des problèmes quand j'ai fait "le curieux" / je l'ai fait exprès aussi ↓ sur la **dissociation corporelle** ↓ la faiblesse je l'ai **reprise par un autre bout** / je change mon fusil d'épaule / quand j'ai vu que ça marchait bien / je suis passé sur autre chose / c'est ça sans arrêt l'œil du maquignon* » (entretien post).

Mener la classe « comme on mène la chasse »

La formatrice mène la classe comme elle mènerait la chasse, multipliant les ruses pour attraper les plus résistants et les rabattre dans l'enclos (11). La métaphore s'applique à la formatrice qui s'efforce avec finesse d'envelopper et encercler les stagiaires dans une zone de collaboration didactique. Cette finesse qui est à l'œuvre dans la classe (sûreté du coup d'œil, habileté à repérer points faibles et points forts, agilité à saisir l'occasion et à viser juste...) relève d'un mode de connaissance qui échappe à la formulation rigoureuse : « *l'œil du maquignon c'est un peu péjoratif comme terme ↓ je n'ai pas d'autre mot que ça / c'est pas considérer les gens comme des animaux du tout ↓ ni les mépriser ↓ mais comment dire ↑ parce que c'est pas scientifique hein ↓ il n'y a pas de recette / ce n'est pas dans les livres ↓* » (entretien post).

Éduquer cet œil chez les stagiaires est néanmoins essentiel. Il faut apprendre au plus vite aux jeunes apprentis enseignants à conduire une classe en déchiffrant indices et signes. La formatrice évoque de ce point de vue, les formations qu'elle mène avec des enseignants débutants : « *je leur dis alors qu'est ce que vous avez vu ↑ au début ils voient que les-bons-les-mauvais ↓ je dis non ↓ en chaque bon il y a du mauvais et en chaque mauvais il y a du bon / [cite quelques critères] alors **ça** j'avais pas vu / et ils apprennent peu à peu à entendre les élèves / à **repérer*** » (entretien post).

CONCLUSION

Si l'œil avisé est si important pour le formateur, alors il nous faut accorder un intérêt particulier à une forme d'intelligence de la ruse (que M. Détienne & J.-P. Vernant (1974) nomment la *mêtis*), qui vise toujours l'efficacité pratique. Or, toutes ces compétences dont est faite la ruse du formateur, sont habituellement ramenées au niveau des qualités personnelles et psychologiques de l'individu, ou des routines acquises avec l'âge, quand elles ne sont pas laissées dans l'ombre et le silence.

L'intérêt méticuleux porté à la gamme des gestes didactiques du formateur désavoue les descriptions

de classe réalisées à partir d'une dichotomie entre l'intelligible et le sensible. D'ordinaire groupées par couples, ces catégories contrastées s'ajustent les unes aux autres pour former un système d'antinomies : d'un côté, le domaine prescriptif du *savoir* jugé figé et rigoriste, des techniques, de l'immuable, du limité ; de l'autre, le domaine créatif de l'*être*, jugé quant à lui mobile et multiple, des émotions, du devenir, du multiple, de l'illimité. Former aux gestes professionnels nous détache de ces catégories beaucoup trop schématiques et générales pour saisir à elles seules le sens de l'activité du formateur. Ce qui caractérise « l'œil du maquignon » c'est justement d'opérer, à l'inverse, une fusion entre tous ces plans.

NOTES

- (1) La situation étudiée prend place au sein d'un dispositif de formation « Le comique par l'expression corporelle » mis en place lors du « stage d'Éducation physique scolaire » au Centre universitaire pour la formation continuée en éducation physique (CUFO-CEP) à Louvain-La-Neuve (Belgique) en 2002.
- (2) La formatrice, Mathilde Dumont dite « Matho » de son nom de scène, est à la fois enseignante, artiste spécialiste de mime et d'expression corporelle et experte en formation dans le domaine des activités physiques artistiques. Elle a été professeur à l'École normale supérieure d'EPS et a fondé à Paris, avec Monique Bertrand dite « Pinok », une compagnie « pré-professionnelle » de théâtre et mime (Compagnie Pinok et Matho, 1962 ; Théâtre l'École Mouvement et Pensée, 1965 ; Tremplin Théâtre des Trois Frères, 1990).
- (3) Le terme « expression corporelle » a été adopté entre 1968 et 1974, lorsque C. Pujade-Renaud publie un livre portant ce titre, qui fédère un mouvement contestataire dans l'EPS, la danse et le théâtre (Entretien avec C. Pujade-Renaud, 2003).
- (4) Par gestes de régulation nous entendons ceux « que le professeur produit en vue d'amener l'élève à élaborer des stratégies gagnantes » (Sensevy, Mercier & Schubauer-Leoni, 2000, p. 269). Ces gestes ne sont pas seulement cybernétiques (au sens de « modification après mesure d'un écart ») mais comportent des interactions didactiques riches et évolutives entre les professeur et les élèves.
- (5) Ces trois questions visant à étudier les phénomènes de contrat didactique en classe, sont issues de la théorie de la transposition didactique (Chevallard, 1985) et retravaillées pour l'étude des pratiques effectives « ordinaires » (Sensevy, Mercier & Schubauer-Leoni, 2000 ; Sensevy, 2002 ; Schubauer-Leoni & Leutenegger, 2002 ; Sensevy, à paraître).
- (6) Les scripts respectent le caractère oral des échanges (d'après Leutenegger-Rhiis, 1999). Les paroles sont notées en italiques, les gestes, regards, comportements et actions motrices en écriture standard et entre parenthèses. Les mots en gras et ceux dont les syllabes sont détachées renvoient respectivement, à une accentuation de l'intonation et à une insistance de la part de l'acteur ; les pauses (de plus ou moins longue durée) sont indiquées par des barres obliques (/ ou // ou ///) ; le ton montant ou descendant par une flèche ↑ ou ↓ ; S désigne le(s) stagiaire(s) et F le formateur.
- (7) Ceci évoque la « réticence » typique des interactions didactiques en classe : « le professeur doit « transmettre » à l'élève des connaissances mais il ne peut les lui transmettre directement » (Sensevy, 2004, p. 337).
- (8) La fabrication, par le professeur, d'élèves collaborateurs « chronogènes » a été mise en évidence dans d'autres travaux (par exemple, Amade-Escot, Verscheure & Uchan, 2004).
- (9) Selon Brousseau (1986), envisager l'enseignement comme la dévolution à l'élève de la responsabilité d'une recherche autonome, conduit le professeur à affronter un paradoxe fondamental : « tout ce qu'il entreprend pour faire produire par l'élève les comportements qu'il attend, tend à priver ce dernier des conditions nécessaires à la compréhension et à l'apprentissage de la notion visée : si le maître dit ce qu'il veut, il ne peut plus l'obtenir » (Brousseau, 1996, p. 59). Les savoirs vont donc devoir « s'avancer sous un masque » pour être « devinés » par l'élève en cours d'action.
- (10) La notion centrale d'obstacle en EPS est analysée comme « une résistance de la *motricité usuelle* des élèves face aux exigences du milieu » (Loquet, 1996, p. 25).
- (11) La métaphore de l'*enclos* est utilisée également par Joshua & Félix (2002) pour désigner le « milieu pour l'étude ». « Le milieu peut être imaginé sous la forme d'un enclos, borné par des frontières. Il comporte à l'intérieur ce qui est considéré comme convenable pour bâtir des apprentissages nouveaux [...] Calibré trop petit, cet enclos sera une prison, qui empêchera une prise de sens quant aux objets nouveaux traités. Calibré trop large, l'enclos contiendra inévitablement trop de « trous » ingérables, et le sens vagabondera. La plupart des enseignements se jouent entre ces deux modes extrêmes. » (2002, p. 93).

BIBLIOGRAPHIE

- AMADE-ESCOT C. & LOQUET M. (2003). *Approches didactiques pour penser l'intervention dans les APSA : l'apport de la didactique comparée*. Symposium 10^e Congrès de l'ACAPS, Toulouse.
- AMADE-ESCOT C. ; LOQUET M. ; LÉZIART Y. (2003). « Collaborations scientifiques didacticiens-chercheurs : dix ans après, quels "effets" sur les pratiques enseignantes ? » Communication au *Symposium Questions de didactique comparée: rencontres du Réseau éducation formation (REF)* : Genève (Suisse), 18, 19 et 20 septembre 2003.
- AMADE-ESCOT C. ; VERSCHEURE I. ; UCHAN K. (2004). « Inégalités de genre en éducation physique et sportive : intérêt de l'approche didactique ». In M. Loquet & Y. Léziart (éd.), *Cultures sportives et artistiques : formalisations des savoirs professionnels, pratiques, formations, recherches*. Actes du 2^e Colloque international de l'Association pour la recherche sur l'intervention en sport (ARIS). Rennes : PUR, p. 151-159.
- BROUSSEAU G. (1986). « Fondements et méthodes de la didactique des mathématiques ». *Recherches en didactique des mathématiques*, vol. 7, n° 2, p. 33-115.
- BROUSSEAU G. (1996). « L'enseignant dans la théorie des situations didactiques ». In *Actes de la VII^e école d'été de didactique des mathématiques*. Clermont-Ferrand : IREM d'Auvergne.
- CHEVALLARD Y. (1985). *La transposition didactique, du savoir savant au savoir enseigné*. Grenoble : Éd. La Pensée sauvage.
- CHEVALLARD Y. (1994). « Les processus de transposition didactique et leur théorisation ». In G. Arzac, Y. Chevallard, J.-L. Martinand & A. Tiberghien (éd.), *La transposition didactique à l'épreuve*. Grenoble : Éd. La Pensée Sauvage.
- CHEVALLARD Y. (1999). « L'analyse des pratiques enseignantes en théorie anthropologique du didactique ». *Recherches en didactique des mathématiques*, vol. 2, n° 19, p. 221-266.
- Contre Pied* (2003). n° 13 : « Danse avec les autres ! ». Paris : Centre national d'étude et d'information « EPS et Société ».
- DETIENNE M. & VERNANT J.-P. (1974). *Les ruses de l'intelligence : la métis des Grecs*. Paris : Flammarion.
- DOUGLAS M. (1999). *Comment pensent les institutions*. Paris : Éd. La Découverte.
- FRANCE : MINISTÈRE DE L'ÉDUCATION NATIONALE (2001). *Enseigner au collège. Éducation Physique et Sportive. Programmes et Accompagnement*. Paris : CNDP.
- FROMANTEL M. ; BERTRAND M. & DUMONT M. (1968). « Études dynamiques ». *Revue EPS*, n° 91, p. 57-59.
- JOSHUA S. & FÉLIX C. (2002). « Le travail des élèves à la maison : une analyse didactique en termes de milieu pour l'étude ». *Revue française de pédagogie*, n° 141, p. 89-97.
- LEUTENEGGER-RHIIS F. (1999). *Contribution à la théorisation d'une clinique pour la didactique. Trois études de cas en didactique des mathématiques*. Thèse de doctorat : sciences de l'éducation, université de Genève [coll. Didactique comparée].
- LOQUET M. (1996). *EPS au collège et Gymnastique rythmique sportive*. Paris : INRP.
- LOQUET M. & AMADE-ESCOT C. (2000). « L'avancée des savoirs dans la classe : Analyse d'une séance ordinaire : atelier méthodologique ». Communication au 1^{er} Colloque de l'Association pour la recherche sur l'intervention en sport (ARIS) : Grenoble, 14, 15 et 16 décembre 2000.
- LOQUET M. ; GARNIER A. & AMADE-ESCOT C. (2002). « Transmission des savoirs en activités physiques, sportives et artistiques dans des institutions différentes : enseignement scolaire, entraînement sportif, transmission chorégraphique ». *Revue française de pédagogie*, n° 141, p. 99-109.
- PUJADE-RENAUD C. (1974/1981). *Expression corporelle, langage du silence*. Paris : ESF.
- PUJADE-RENAUD C. (2003). « Entretien avec Claude Pujade-Renaud », *Contre Pied*, n° 13, p. 74-76.
- SCHUBAUER-LEONI M.-L. & LEUTENEGGER F. (2002). « Expliquer et comprendre dans une approche clinique/expérimentale de didactique ordinaire ». In F. Leutenegger & M. Saada-Robert (éd.), *Expliquer et comprendre en sciences de l'éducation*. Bruxelles : De Boeck, p. 227-251.
- SENSEVY G. ; MERCIER A. & SCHUBAUER-LEONI M.-L. (2000). « Vers un modèle de l'action didactique du professeur. A propos de la course à vingt ». *Recherche en didactique des mathématiques*, vol. 3, n° 20, p. 263-304.
- SENSEVY G. (2002). « Des catégories pour l'analyse comparée de l'action du professeur : un essai de mise à l'épreuve ». In P. Venturini, C. Amade-Escot & A. Terrisse (éd.), *Études des pratiques effectives : l'approche des didactiques*, Grenoble : Éd. La Pensée sauvage, p. 25-46.
- SENSEVY G. (2004). « Discours et Action : quels types de relation ? » In M. Loquet & Y. Léziart (éd.), *Cultures sportives et artistiques : formalisations des savoirs professionnels, pratiques, formations, recherches : Actes du 2^e Colloque ARIS*. Rennes : PUR, p. 335-338.
- SENSEVY G. (à paraître). « Des catégories pour décrire et comprendre l'action du professeur ». In G. Sensevy & A. Mercier (éd.), *Agir ensemble : l'action didactique conjointe du professeur et des élèves*. Rennes : PUR.

La construction sociale du jugement inspectoral : la définition du métier d'enseignant vue à travers une analyse de rapports d'inspection en EPS

*Marie-Paule Poggi, Nathalie Wallian,
Mathilde Musard*

Nous faisons l'hypothèse que le jugement inspectoral ne répond pas seulement à des normes pédagogiques et didactiques mais qu'il se construit également à partir de contraintes sociales telles que le type de public scolaire et les caractéristiques socio-culturelles des enseignants. L'analyse de contenu de cent-onze rapports d'inspection nous permet d'examiner la nature des modèles d'intervention valorisés par l'institution. Le discours inspectoral tend à uniformiser ses attentes en invoquant des principes généraux prônant principalement le recours aux méthodes actives et à un traitement différencié des élèves. Cependant, ces exigences se font plus pressantes en établissements défavorisés, auprès des femmes et des jeunes enseignants. Par ailleurs, on ne sait rien de ce que les élèves apprennent au cours d'une leçon. En ignorant les pratiques réelles des élèves, l'inspecteur ne se prive-t-il pas d'une dimension essentielle à l'évaluation : l'efficacité de l'enseignement ?

Descripteurs (TEE) : conduite de la classe, éducation physique, expérience de l'enseignement, inspection, programme d'études.

INTRODUCTION

Au-delà d'une obligation purement administrative, l'inspection tend à constituer un acte d'évaluation et de formation. Cette triple caractéristique fait d'elle un révélateur intéressant des conceptions en Éducation physique et sportive (EPS) et des différentes facettes constitutives de l'activité enseignante dans cette discipline. Notre recherche s'est organisée selon une double perspective : d'une part, participer à l'ana-

lyse des pratiques enseignantes à partir d'une entrée toute particulière, celle du discours institutionnel et, d'autre part, approcher les processus de construction sociale du jugement inspectoral. L'objectif poursuivi est donc de s'intéresser à une forme, parmi d'autres, de description et d'évaluation des pratiques enseignantes. Cette évaluation légitime ces pratiques ou au contraire les exclue du champ des pratiques reconnues. Le contenu des rapports d'inspection constitue alors une trace intéressante à analyser.

Reprenons les deux pistes de réflexion évoquées et tout d'abord celle relative à l'analyse des pratiques enseignantes. On peut dire que par leur dimension à la fois descriptive, évaluative et prescriptive, ces rapports nous informent sur le point de vue des inspecteurs à propos des activités concrètement mises en œuvre par l'enseignant dans sa classe. Ils sont également révélateurs des exigences officielles en provenance de l'institution ainsi que des processus sous-jacents aux pratiques enseignantes (1) tout au moins tels que les inspecteurs les définissent. A ce titre, ils se situent à l'articulation entre *curriculum* réel, *curriculum* formel et *curriculum* caché (Perrenoud, 1984). Leur analyse constitue une entrée possible parmi d'autres dans l'étude des pratiques enseignantes. En fonction d'informations puisées dans ce qui fait la réalité d'une classe pour l'inspecteur, celui-ci déroule un récit à la fois descriptif, explicatif et compréhensif de l'activité des élèves et du mode d'intervention de l'enseignant. Bien entendu, ce regard n'est pas neutre socialement puisqu'il émane d'un acteur qui, comme le précise le rapport Monteil (1999), joue un rôle d'« évaluateur pédagogique » et de « cadre, vecteur des politiques académiques, déclinées à partir des orientations nationales ». Ce « porte parole autorisé » (Bourdieu, 1982), du fait de son statut particulier, est en droit de définir « l'usage pédagogique » légitime (Voluzan, 1975).

Au-delà de notre analyse descriptive et compréhensive des pratiques à partir du discours institutionnel, cette recherche s'est également donnée pour objet d'approcher les processus de construction sociale du jugement inspectoral. Nous faisons l'hypothèse que ce jugement ne répond pas seulement à des normes pédagogiques et didactiques mais qu'il se construit également à partir de contraintes sociales telles que le type de public scolaire, les caractéristiques socio-culturelles des enseignants ou encore le cadre de référence personnel des inspecteurs (2). Notons toutefois que les variations liées au profil des inspecteurs n'ont pas été étudiées dans le cadre de ce travail compte tenu des caractéristiques de notre échantillon mais elles mériteraient sans doute qu'on y porte une attention particulière. En effet, si l'inspecteur reste le représentant mandaté par l'institution, il n'en demeure pas moins un acteur disposant d'une marge d'interprétation et d'appropriation des normes officielles au regard à la fois des caractéristiques locales d'enseignement (projet d'établissement, projet pédagogique EPS) mais également de ses orientations et valeurs personnelles.

CADRE THÉORIQUE ET PROBLÉMATIQUE

Notre réflexion s'inscrit dans le cadre théorique de la sociologie du *curriculum* (Forquin, 1989 ; Young, 1971), qui étudie les mécanismes par lesquels le savoir est sélectionné, organisé et évalué dans les établissements scolaires. Comme le souligne D. Lawton, cité par J.-C. Forquin, le *curriculum* désigne « les sélections faites par les écoles au sein de la culture, il détermine la façon dont les priorités sont décidées et mises en pratique ». Cette définition nous guidera tout au long de notre exposé. Plus précisément ici, il s'agit d'identifier la nature des savoirs et des compétences promus au statut de culture légitime en EPS en nous appuyant sur le discours d'un acteur clé de l'institution.

Plus globalement, ce projet est lié aux recherches en sociologie de l'éducation qui s'intéressent au poids de variables sociologiques externes telles que les caractéristiques socioculturelles des enseignants et des élèves et à leur influence sur la détermination des contenus, l'orientation des démarches pédagogiques ou encore la définition du métier. Pour résumer, nous nous interrogerons sur la nature des modèles pédagogiques et didactiques valorisés, préconisés, prescrits et/ou, au contraire rejetés (proscrits) par les inspecteurs. Les différents axes de la problématique s'articulent autour de quatre éléments :

1) Quels jugements, constitués en norme pédagogique officielle, les inspecteurs portent-ils sur l'élève et ses activités en EPS mais également sur le travail de l'enseignant en situation d'intervention (3) ? Au fond, quels sont les modèles didactiques et pédagogiques valorisés par l'institution qui vont constituer en quelque sorte cet « usage pédagogique légitime » ?

2) De quelle façon la composition sociale des publics scolaires ainsi que les caractéristiques socio-culturelles des enseignants pèsent-elles sur la définition de ce modèle pédagogique ?

3) Quels sont les modèles produits localement ? La question qui se pose ici est celle de la fabrication opérée localement par l'inspection d'une partie des normes pédagogiques dans un contexte de décentralisation où les initiatives engagées précisément par le local sont massivement encouragées. Quel est le rôle des principaux acteurs locaux que sont inspecteurs et enseignants, dans la fabrication du pédagogique ? Plus globalement, il s'agit d'étudier la façon dont s'opère la mise en œuvre locale de la politique éducative nationalement définie ou comme le souligne A. Henriot-van Zanten (1990), d'analyser

les « processus de réappropriation des directives nationales par des groupes locaux ».

4) Quelle contribution spécifique, une différenciation des modèles pédagogiques et didactiques valorisés par l'institution, apporte-t-elle aux processus d'inégalité entre les élèves ? Mais ce dernier versant est peu exploré dans ce travail.

Nous ferons donc l'hypothèse que les rapports d'inspection diffusent des modèles d'intervention qui varient selon le cadre de référence des inspecteurs, le lieu d'implantation des établissements scolaires et le profil socioculturel des enseignants. Nous voulons mettre en rapport l'analyse de contenu des rapports d'inspection avec des données externes telles la tonalité sociale dominante du public scolaire et les caractéristiques socioculturelles des enseignants. Nous interrogerons ce discours local au regard de l'esprit du discours national afin de repérer des décalages ou des proximités entre des prescriptions ministérielles prônant la transmission d'une culture commune, d'un « socle commun de connaissances » et un discours local particulier contraint de s'adapter. On peut craindre que ces ajustements aux caractéristiques locales d'enseignement soient producteurs d'inégalités ou du moins participent au creusement des inégalités déjà existantes entre les établissements scolaires. Le local sera ici étroitement associé à la notion de contexte (4) dans lequel l'activité de l'inspecteur s'exerce, contexte constitué par un ensemble de contraintes liées aux caractéristiques propres du milieu étudié : celles des établissements, des enseignants et des inspecteurs. Nous pourrions ainsi observer l'influence de ces phénomènes contextuels sur la définition des profils d'intervention attendus par les inspecteurs.

DES RAPPORTS D'INSPECTION : DE QUOI S'AGIT-IL ?

Selon les travaux des spécialistes de l'analyse du discours, les rapports d'inspection peuvent entrer dans la catégorie des « discours contraints » par opposition aux interactions orales spontanées (Maingueneau, 1998). Ils ont pour caractéristique d'être soumis à des « routines » ou des normes, ils se déroulent selon des modèles stabilisés qui obéissent à des contraintes fortes qui ne peuvent que légèrement évoluer. Notre *corpus* répond à ces règles de production qui rendent les données relativement homogènes. Les inspecteurs concernés se sont fixés localement un certain nombre de règles dans la limite

des contraintes institutionnelles qui s'imposent à eux. De fait, la trame générale suivie est toujours la même. Trois parties composent les rapports :

- l'introduction précise en quelques lignes le choix de l'activité physique support, le moment dans le cycle, le descriptif des différents temps de la leçon ;
- le corps du rapport, partie la plus conséquente, consiste en une analyse critique de la leçon constituée d'une suite d'assertions positives, négatives et prescriptives ;
- enfin, en quelques lignes, la conclusion dresse un portrait de l'enseignant à partir de ses caractéristiques personnelles et professionnelles : son investissement dans la vie de l'établissement et hors établissement, son parcours personnel, ses projets, ses difficultés personnelles, etc.

Au-delà de cet aspect purement formel, le rapport d'inspection constitue un rite social particulier que nous ne pouvons comprendre qu'en examinant de plus près les conditions sociales de sa production. En effet, il s'impose comme un discours d'institution c'est-à-dire comme « la parole officielle du porte parole autorisé s'exprimant en situation solennelle ». Comme tout « discours d'autorité » produit par un « porte parole autorisé », il n'est en mesure de « définir l'usage légitime que parce qu'il est prononcé par une personne légitime, en situation légitime et dans les formes légitimes » (Bourdieu, 1982) et qu'il est reconnu comme tel par ses récepteurs. Sa puissance performative se mesurera sans doute à l'aune de cette reconnaissance qui variera selon l'âge, le statut ou encore le grade de l'enseignant.

Enfin, d'un point de vue pédagogique, ces rapports se trouvent à l'interface entre *curriculum* formel et *curriculum* réel. L'inspecteur se livre à une interprétation de la réalité, produit à la fois de son système de valeurs et des contraintes institutionnelles qui pèsent sur lui. Son discours ne révèle qu'une image déformée de ce qui fait le quotidien de la classe en EPS. Ne sont retenus que quelques traits « idéal-typiques » (Weber, 1971) constitutifs du profil d'intervention de l'enseignant. Il ne s'agit en aucun cas d'un compte rendu fidèle et exhaustif des événements constitutifs d'une séance d'EPS.

Finalement, comme le souligne Plaisance (1986), ils constituent « une sorte de procès-verbal de l'usage pédagogique et même un modèle pédagogique, à la fois sous forme de description et de formulation de jugements et de conseils ». Leur contenu devrait nous renseigner sur la conception des savoirs à transmettre en EPS, sur la nature des changements à produire

chez l'élève et chez l'enseignant, bref, sur la façon de concevoir le métier d'enseignant. Pour résumer, on peut dire que ces rapports sont révélateurs des modèles pédagogiques et didactiques prônés par l'institution sous la forme de descriptions, de jugements critiques (positifs ou négatifs) et de prescriptions.

MÉTHODOLOGIE

Notre corpus est constitué de cent onze rapports d'inspection rédigés par deux inspecteurs hommes issus d'une même académie. Les enseignants, cinquante femmes et soixante et un hommes, exercent tous en collège. Le *corpus* est homogène du point de vue du niveau de *cursus* considéré.

La majorité des inspectés est composée de certifié(e)s, 78,4 %, pour 8,1 % d'agrégés ou bi-admissibles et 13,5 % de chargés d'enseignement, maîtres auxiliaires ou PEGC. On trouve à peu près à parts égales des enseignants en début, milieu et fin de carrière, respectivement 27,9 %, 29,7 % et 35,1 % (5). L'athlétisme et la gymnastique ainsi que les sports collectifs constituent la part la plus importante des activités physiques enseignées (6). Enfin, l'échantillon compte à peu près autant d'établissements favorisés, 42,3 %, que défavorisés, 45,9 % (7).

Nous avons soumis ces rapports à une analyse de contenu visant à repérer la nature des modèles pédagogiques et didactiques valorisés par l'institution et donc produits par l'école. L'ensemble du *corpus* a été traité à partir d'une analyse quantitative (comptage du nombre de jugements neutres, favorables, défavorables et prescriptifs) ainsi que d'une analyse de contenu thématique plus qualitative.

Procédures de dépouillement

Nous avons opté pour une analyse thématique telle qu'a pu la définir L. Bardin (1983). Les thèmes sont considérés comme des « noyaux de sens », des « unités de signification complexes, de longueur variable » (*Ibid.*). Le parti pris pour mettre en œuvre cette analyse est donc d'ordre sémantique et non lexical. C'est bien le sens et non la forme qui a organisé le découpage des textes. Les unités d'enregistrement sont repérées à partir de mots ou de groupes de mots. On peut trouver dans une phrase un même thème répété deux fois, il sera alors comptabilisé deux fois. La phrase sert d'unité de contexte, on dénombre en son sein le nombre de thèmes présents. Les pourcentages ont été calculés en rapportant le nombre de jugements étudiés sur le nombre total de jugements. Nous avons procédé à un relevé exhaustif de toutes les expressions constitutives du jugement inspectoral. Ces dernières ont été traitées à l'aide de deux grilles.

Pour chaque sous-catégorie, nous avons distingué : ce qui est dévalorisé, valorisé, prescrit.

La première (grille 1), de type plutôt quantitatif, a permis de classer les discours en fonction de leur orientation : neutre, approbative, critique ou encore prescriptive. La seconde (grille 2), à orientation qualitative, a permis de préciser l'objet sur lequel portaient ces différents types de jugements (la prise en compte des élèves, les qualités personnelles de l'enseignant, la gestion et l'organisation, etc.) que nous avons associée à une orientation du jugement de type critique, positive ou prescriptive. Ces informations qualitatives et quantitatives sur la nature du jugement inspectoral ont été mises en relation avec différentes variables indépendantes classiques telles que :

Grille 1. – Orientations des jugements portés par l'inspecteur

Type de <i>curriculum</i>	Contenu	Orientations	Appellation
<i>curriculum</i> réel décrit/jugé	ce qui se fait (constat sans engagement)	neutre	jugements assertifs neutres
	ce qui est bien fait	approbative	jugements assertifs favorables
	ce qu'il aurait fallu faire, ce qui est mal fait	critique	jugements assertifs défavorables
<i>curriculum</i> prescrit	ce qu'il faudrait faire ou ne pas faire	pratique	édications pratiques

Grille 2. – **Objets du discours inspectoral (des illustrations se trouvent en annexe I)**

Catégories	Sous-catégories
Prise en compte des élèves	<ul style="list-style-type: none"> – diversifier les itinéraires d'apprentissage, prendre en compte les caractéristiques des élèves ; – observer les élèves, contrôler les acquis ; – faire participer les élèves aux apprentissages ; – développer le temps de pratique
Qualités personnelles	<ul style="list-style-type: none"> – caractéristiques morales, affectives et relationnelles ; – caractéristiques intellectuelles ; – caractéristiques professionnelles et institutionnelles
Gestion/organisation Communication/outils	<ul style="list-style-type: none"> – maîtrise des aspects gestionnaires et organisationnels (gestion du temps, de l'espace, répartition des rôles, rotation dans les ateliers) ; – la communication et les outils (passation des consignes, gestion de la sécurité, <i>media</i> utilisés, etc.)
Savoirs et culture	<ul style="list-style-type: none"> – connaissances, savoirs, compétences ; – valeurs et attitudes
Recours au <i>curriculum</i> formel	<ul style="list-style-type: none"> – conformité aux programmes officiels, respect de la sécurité, devoir d'évaluation, etc.
Analyse de sa pratique	<ul style="list-style-type: none"> – conception, planification ; – capacité à analyser son action, à exploiter les bilans
Analyse des Activités physiques sportives artistiques (APSA)	<ul style="list-style-type: none"> – respect de la logique de l'APSA
Activité des élèves	<ul style="list-style-type: none"> – activités des élèves liées à l'apprentissage de savoirs et savoir faire ; – attitudes et comportements des élèves

- la tonalité sociale dominante du collège établie à partir des statistiques sur la répartition en catégories socioprofessionnelles des élèves, nous avons classé les établissements en tenant compte de l'écart entre les statistiques d'établissement et la moyenne académique ;
- le sexe, le grade et l'échelon de l'enseignant concerné.

nelles des enseignants sont reconnues positivement. La volonté d'encouragement paraît manifeste. L'inspection veut clairement s'inscrire dans une démarche formative plus que sommative. L'intervention de l'inspecteur vise à aider, encourager et guider l'intervenant. La dimension prescriptive est également privilégiée avec 26 % des jugements qui y sont consacrés. L'inspection répond là à une de ses missions traditionnelles consistant à orienter et guider les pratiques. Finalement, deux logiques animent les inspecteurs : l'encouragement et l'aide à la décision.

RÉSULTATS

La direction des jugements : neutre, favorable, défavorable ou prescriptive

Les jugements favorables dominent (50,9 %), on compte environ un quart de discours prescriptifs (26,1 %) et un peu plus de 10 % de déclarations défavorables. Globalement, les compétences profession-

L'objet des jugements : la prise en compte des élèves, l'activité des élèves, les caractéristiques personnelles des enseignants, les instructions officielles, l'analyse de sa propre pratique...

Près d'un tiers des discours concerne la prise en compte des élèves par les enseignants, quelle que soit l'orientation des déclarations (positive, négative).

tive ou prescriptive) ; la préoccupation principale concerne donc la place occupée par les élèves dans la démarche d'enseignement. Dans ce domaine, trois sous thèmes sont tout particulièrement évoqués avec insistance :

1) diversifier les itinéraires d'apprentissage, prendre en compte les caractéristiques des élèves et viser la réussite de tous (14,1 %) ;

2) observer les élèves, contrôler les acquis : réaliser des bilans, organiser la connaissance des résultats (6,6 %) ;

3) faire participer les élèves aux apprentissages (5,2 %) ;

4) et, plus secondairement, développer le temps de pratique : favoriser l'activité de tous les élèves, mettre rapidement les élèves en activité, engagement

Tableau I. – L'objet des jugements quelle que soit l'orientation du discours

	Effectif	%
Prise en compte des élèves	775,0	29,2
diversifier les itinéraires d'apprentissage, prendre en compte les caractéristiques des élèves	374,0	14,1
observer les élèves, contrôler les acquis	175,0	6,6
faire participer les élèves aux apprentissages	138,0	5,2
développer le temps de pratique	88,0	3,3
Caractéristiques des enseignants	647,0	24,4
caractéristiques morales, affectives et relationnelles	245,0	9,2
caractéristiques intellectuelles	69,0	2,6
caractéristiques professionnelles et institutionnelles	333,0	12,5
Gestion/organisation, communication/outils	529,0	19,9
maîtrise des aspects gestionnaires et organisationnels	366,0	13,8
la communication et les outils	163,0	6,1
Savoirs et culture	117,0	4,4
connaissances, savoirs, compétences	96,0	3,6
valeurs et attitudes	21,0	0,8
Recours au curriculum formel	22,0	0,8
Analyse de sa pratique	385,0	14,5
conception, planification	271,0	10,2
capacité à analyser son action, à exploiter les bilans	75,0	2,8
autre	39,0	1,5
Analyse des APSA	47,0	1,8
Activité des élèves	134,0	5,0
activité des élèves liées à l'apprentissage de savoirs et savoir-faire	75,0	2,8
attitudes et comportements des élèves	59,0	2,2
<i>Ensemble</i>	2 656,0	100,0

Tableau IIa. – L'objet des jugements selon l'orientation du discours

Ce qui est valorisé		
	N	%
Caractéristiques des enseignants	574	39,3
Prise en compte des élèves	250	17,1
Gestion/organisation, communication/outil	250	17,1
Analyse de sa pratique	244	16,7
Activité des élèves	80	5,5
Analyse des APSA	36	2,5
Savoirs et culture	17	1,2
Recours au <i>curriculum</i> formel	8	0,5
	1 459,0	100,0

Tableau IIb. – L'objet des jugements selon l'orientation du discours

Ce qui est dévalorisé		
	N	%
Prise en compte des élèves	131	35,6
Gestion/organisation, communication/outil	75	20,4
Analyse de la pratique	73	19,8
Caractéristiques des enseignants	38	10,3
Activité des élèves	26	7,1
Savoirs et culture	19	5,2
Recours au <i>curriculum</i> formel	4	1,1
Analyse des APSA	2	0,5
	368,0	100,0

moteur des élèves, donner du temps aux apprentissage, veiller à ce que les élèves s'exercent suffisamment, etc. (3,3 %).

Trois autres thèmes font également l'objet de toute l'attention des inspecteurs : les caractéristiques des enseignants, l'organisation et la communication dans la classe, l'analyse de sa pratique professionnelle. En revanche, quatre catégories sont quasiment exclues

Tableau IIc. – L'objet des jugements selon l'orientation du discours

Ce qui est prescrit		
	N	%
Prise en compte des élèves	394	47,5
Gestion/organisation, communication/outil	204	24,6
Savoirs et culture	81	9,8
Analyse de la pratique	68	8,2
Caractéristiques des enseignants	35	4,2
Activité des élèves	28	3,4
Recours au <i>curriculum</i> formel	10	1,2
Analyse des APSA	9	1,1
	829,0	100,0

du discours inspectoral : l'activité des élèves, les savoirs et la culture, l'analyse des APSA, le recours au *curriculum* formel. Le discours des inspecteurs ne balaie donc pas l'ensemble des questions liées à l'intervention en EPS mais, au contraire, reste ciblé sur des préoccupations particulières dont les savoirs et l'activité réelle des élèves semblent exclus.

L'analyse de l'orientation des jugements permet de repérer les pôles de centration des inspecteurs. Les réserves et les critiques concernent principalement les difficultés des enseignants à prendre en compte les caractéristiques de leurs élèves (35,6 % des discours critiques) et plus précisément leur incapacité à diversifier les itinéraires d'apprentissages (20,1 % des discours critiques contre 3 % pour « développer le temps de pratique », 9,5 % pour « observer les élèves des élèves » et 3,2 % pour « faire participer les élèves aux apprentissages »). Logiquement, c'est également ce domaine qui est visé en premier par les prescriptions. Les discours laudateurs visent d'abord les caractéristiques personnelles et professionnelles de l'enseignant, ensuite sa démarche pédagogique et en troisième lieu ses capacités plus techniques de gestion de la classe et de communication des informations.

Ces résultats appellent quatre commentaires

1) Premièrement, une minorité de rapports s'attache à l'activité réelle des élèves en situation, aux savoirs

en jeu et à l'analyse de l'APSA. L'inspection concerne d'abord et surtout l'enseignant. Nos résultats convergent avec ceux de A. Si Moussa (2000) qui notait à propos de rapports établis dans le premier degré que « l'observation de l'inspecteur porte davantage sur la façon dont le maître enseigne que sur la manière dont les élèves se comportent et apprennent ». Dans notre corpus, la tendance est similaire, puisqu'on trouve peu d'informations sur ce qui est réellement acquis par les élèves, sur les stratégies mises en œuvre par ces derniers pour apprendre ou encore sur les attitudes et comportements manifestés en classe. Pourtant, près de la moitié des discours prescriptifs (47,5 %) adressés aux enseignants incite à davantage d'efficacité dans ce domaine (faire participer les élèves aux apprentissages, diversifier les itinéraires d'apprentissage, etc.). Les stratégies d'intervention se trouvent au cœur du dispositif d'analyse des pratiques au détriment d'une analyse des stratégies d'apprentissage développées par les élèves (5 % des discours). Les rapports d'inspection se montrent également extrêmement discrets sur le type et/ou le niveau de savoirs et de compétences acquis ou non acquis voire en cours d'acquisition chez les élèves (4,4 % des discours). Les enseignants sont ainsi dépourvus de précieuses informations utiles à la régulation de leurs pratiques d'intervention. Ce constat est sans doute inhérent à la spécificité des pratiques d'observation et d'évaluation en EPS. Le caractère éphémère des réponses des élèves prive inspecteurs et enseignants d'une trace tangible, à l'instar des autres disciplines d'enseignement, sur laquelle le discours des uns et des autres pourrait s'appuyer à l'issue de la leçon. Il faut saisir en situation des informations sur la motricité des élèves et apprécier dans l'instant des productions motrices furtives et évolutives, impossibles à réitérer à l'identique, sauf recours, logiquement lourd, à la vidéo. Un retour en arrière, une relecture *a posteriori* des productions des élèves appuyée sur des preuves effectives s'avèrent donc difficiles. Par ailleurs la liberté de mouvement, constitutive de l'activité des élèves en cours d'EPS, donne à voir une pluralité de mises en forme de cette activité aussi bien au niveau de la gestion de l'espace que des formes de groupement qui rendent plus complexe qu'ailleurs l'observation menée par l'inspecteur.

2) Dans un second point, on note une tendance à la personnalisation des jugements. 39,3 % des discours laudateurs valorisent les caractéristiques personnelles des enseignants : sur ces 39,3 %, on compte 15,1 % de caractéristiques morales, affectives et relationnelles (ambition, bonne volonté, simplicité, naturel, qualités humaines, etc.), 4,4 % de caracté-

ristiques intellectuelles (inventif, réfléchi, rigoureux) et 19,8 % de caractéristiques professionnelles (maîtrise professionnelle, expertise pédagogique, acteur de la vie scolaire de premier ordre). Les traits de personnalité sont ainsi amalgamés aux compétences professionnelles. Le rapport Monteil (1999) fait état de ce phénomène de psychologisation qu'il qualifie « d'extrapolations abusives attachées à l'évaluation », extrapolations qui ne « recouvrent aucune réalité scientifique » et qui « ne rendent pas compte d'une connaissance effective des pratiques ». On note que le recours à ce type d'arguments est principalement utilisé à des fins d'encouragement, il ne représente que 7 % des discours critiques qui ne concernent, dans ce cas, que les caractéristiques professionnelles et non strictement personnelles.

3) Troisièmement, contrairement à ce qu'on aurait pu attendre, l'exigence de conformité des pratiques aux prescriptions officielles n'est évoquée qu'à la marge (moins de 1 % des jugements). Le rappel des directives institutionnelles n'est pas utilisé comme argument d'autorité, de contrainte ou de rappel à l'ordre. L'obligation de conformité aux programmes officiels, le respect de la sécurité, ou encore le devoir d'évaluation ne sont que timidement rappelés. En fait c'est plutôt l'esprit ou l'orientation générale des textes officiels qui transparait en filigrane dans le contenu des discours critiques et notamment le principe qui vise à placer l'élève au centre du dispositif éducatif.

4) Enfin, les discours critiques déplorent principalement le manque de prise en compte des élèves (35,7 %), puis les difficultés de gestion et de communication (30,6 %) ainsi que les résistances ou les incapacités à analyser sa propre pratique (26,5 %). Finalement, la norme pédagogique prônée se décline en trois principes qui se trouvent en continuité avec le discours officiel national :

- diversifier, prendre en compte les caractéristiques des élèves ;
- observer les élèves et contrôler leurs acquis ;
- faire participer les élèves à leurs apprentissages.

En revanche on est peu renseigné sur ce qu'ont réellement appris les élèves.

Cette orientation s'inscrit pleinement dans l'esprit des textes officiels qui incitent à « prendre en compte des élèves différents dans un collège pour tous » ou encore à « diversifier les méthodes d'enseignement pour aiguïser l'appétit d'apprendre » (France, 1999), bref qui arborent clairement la volonté de mettre

« l'élève au centre du dispositif éducatif » tout en s'appuyant sur une « conception républicaine des méthodes actives » (France, 2002). Nos résultats montrent que le discours local prend le relais du discours national ; en fait les réappropriations particulières propres au niveau local apparaissent lorsque l'on s'intéresse aux variations selon la tonalité sociale dominante du public scolaire ou selon les caractéristiques socioculturelles des enseignants.

LES VARIATIONS SELON LES CARACTÉRISTIQUES SOCIALES DES ENSEIGNANTS ET DES ÉLÈVES

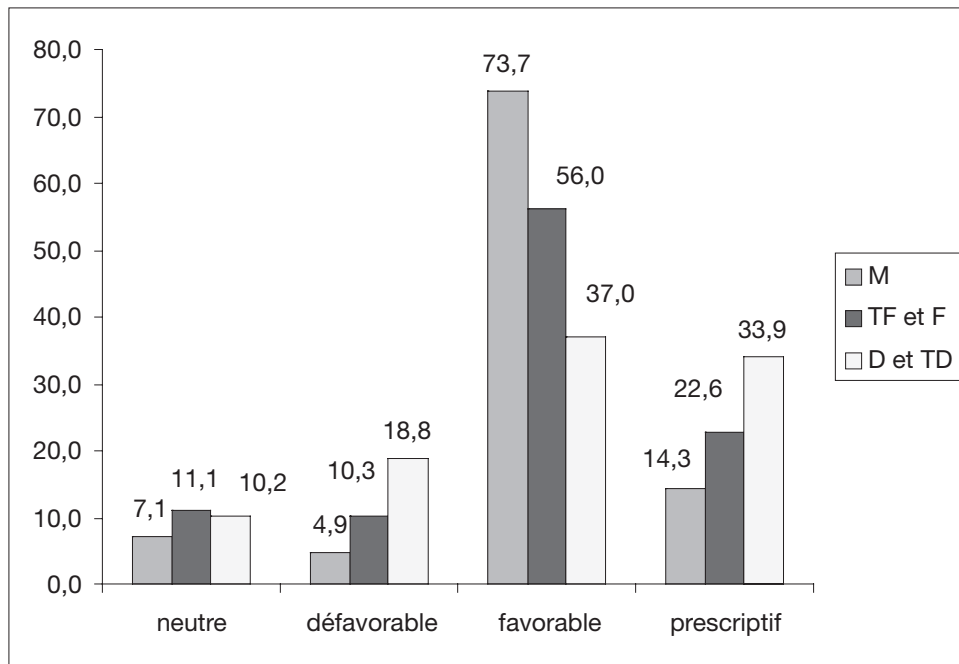
La tonalité sociale dominante de l'établissement

Quel que soit le type de jugement, le discours de l'inspecteur est toujours plus développé quantitativement auprès d'enseignants exerçant en établissements défavorisés ou mixtes. Le nombre de jugements émis en moyenne par type d'établissements est

respectivement de 25,3 % et 26,8 % contre 21,5 % en milieu favorisé. Les inspecteurs sont moins réactifs à ce qui se passe dans ce type de quartier. En ce qui concerne la nature des discours développés, on note que les enseignants exerçant en établissements favorisés sont plus souvent jugés de façon positive contrairement à leurs collègues des établissements défavorisés plus souvent exposés à des jugements critiques ou prescriptifs.

C'est principalement une mauvaise prise en compte des élèves qui fait défaut aux enseignants exerçant en établissements favorisés (44,6 %). Alors qu'en quartiers défavorisés, les reproches se répartissent entre une mauvaise prise en compte des élèves (30,5 %) et des difficultés d'organisation et de communication (21,2 %). Il en est de même pour les prescriptions. L'effort de différenciation des pratiques qu'exige l'institution semble moins soutenu en milieu favorisé. Face à un public déjà bien familiarisé avec les formes d'apprentissage scolaires, différencier les pratiques ne semble pas constituer une priorité. Ce contexte favorable engendre une certaine routine de fonctionnement qui nuit à un effort de réflexion sur

Figure 1. – Répartition des jugements selon la tonalité sociale dominante de l'établissement



Légende : pour « D » il faut lire *défavorisé*, pour « TD » il faut lire *très défavorisé*, pour « F » il faut lire *favorisé*, pour « TF » il faut lire *très favorisé*, pour « M » il faut lire *mixte* (les résultats détaillés se trouvent en annexe II).

Tableau III. – Répartition des jugements selon la tonalité sociale dominante de l'établissement

Variables	Valorisé				Prescrit				Dévalorisé			
	D TD		F TF		D TD		F TF		D TD		F TF	
	N	%	N	%	N	%	N	%	N	%	N	%
Prise en compte des élèves	92,0	15,3	115,0	18,0	230,0	46,9	123,0	48,4	62,0	30,5	54,0	44,6
Caractéristiques des enseignants	240,0	40,0	249,0	39,0	28,0	5,7	7,0	2,8	33,0	16,3	5,0	4,1
Gestion/organisation, communication/outils	111,0	18,5	115,0	18,0	111,0	22,7	56,0	22,0	43,0	21,2	19,0	15,7
Savoirs et culture	6,0	1,0	7,0	1,1	41,0	8,4	39,0	15,4	12,0	5,9	6,0	5,0
Recours au <i>curriculum</i> formel	3,0	0,5	4,0	0,6	7,0	1,4	2,0	0,8	4,0	2,0	0,0	0,0
Analyse de sa pratique	104,0	17,3	96,0	15,0	41,0	8,4	23,0	9,1	38,0	18,7	27,0	22,3
Analyse des APSA	16,0	2,7	14,0	2,2	7,0	1,4	2,0	0,8	2,0	1,0	0,0	0,0
Activité des élèves	28,0	4,7	38,0	6,0	25,0	5,1	2,0	0,8	9,0	4,4	10,0	8,3
Ensemble	600,0	100,0	638,0	100,0	490,0	100,0	254,0	100,0	203,0	100,0	121,0	100,0

Légende : Pour « D » il faut lire *défavorisé*, pour « TD » il faut lire *très défavorisé*, pour « F » il faut lire *favorisé*, pour « TF » il faut lire *très favorisé* (8).

sa propre pratique, 22,3 % des discours désapprobateurs dénoncent des faiblesses dans la conception des séances et dans la capacité à analyser son action.

D'autre part, l'attention portée par l'inspecteur aux contenus à transmettre est plus marquée en établissements favorisés (15,4 % en établissements favorisés contre 8,4 % en établissements défavorisés des discours prescriptifs). Face aux difficultés liées à la conduite de la classe (21,2 % des discours dévalorisant la gestion et la communication), le souci d'améliorer les contenus auprès d'élèves de milieu défavorisé n'apparaît pas primordial. Ce renoncement peut contribuer à une accentuation des inégalités d'accès à la culture scolaire entre différents types d'élèves. Enfin, dans tous les cas, les qualités personnelles et professionnelles des enseignants constituent la première source de valorisation du travail des enseignants. Cependant, ces mêmes qualités personnelles font nettement plus l'objet de critiques en établissements défavorisés (16,3 % en établissements défavorisés contre 4,1 % en établissements favorisés).

Quelles conclusions tirer de ces résultats ?

Sans nul doute, exercer en établissement défavorisé augmente de manière assez significative les

chances d'obtenir une appréciation négative, la sévérité des inspecteurs s'exerce avec davantage de force auprès de ceux qui interviennent dans des conditions déjà peu confortables. Cependant, autour du même thème fédérateur de la différenciation des pratiques particulièrement prisé par l'inspection, le discours des inspecteurs dénonce les carences manifestes dans ce domaine dans des établissements pourtant peu exposés à un exercice difficile de cette différenciation. En quartier défavorisé, les reproches sont d'un autre ordre. Les caractéristiques personnelles des enseignants, leur capacité à analyser leur pratique, bref leur investissement personnel et professionnel dans le métier représente une part relativement importante (respectivement 16,3 % et 18,7 % des critiques adressées dans ce type d'établissements) du discours critique des inspecteurs. Enseigner en contexte difficile constitue un terrain propice à des « décrochages professionnels » que les inspecteurs ne manquent pas de pointer tout en reconnaissant à ces mêmes acteurs des compétences en matière de prise en compte des caractéristiques des élèves. Enfin, on peut craindre que la moindre attention accordée aux contenus transmis en établissements défavorisés conduise à accroître les inégalités d'accès au savoir entre les différentes catégories sociales d'élèves.

Les caractéristiques des enseignants

Les femmes font plus souvent que les hommes l'objet de jugements prescriptifs. On ménage les femmes, on les accompagne plus qu'on ne les juge. Les jugements critiques et favorables s'adressent davantage aux hommes. Les fonctions du jugement inspectoral varient selon le sexe de l'enseignant : il joue un rôle de guide pour les femmes, et plutôt de contrôle/évaluation pour les hommes.

Les hommes sont jugés défaillants dans le domaine de la prise en compte des caractéristiques des élèves et du point de vue de leurs propres caractéristiques personnelles et professionnelles, tandis que les femmes sont invitées à améliorer leurs pratiques du côté de la gestion, l'organisation et la communication.

Enfin, les jugements défavorables concernent les enseignants les moins gradés. Les agrégés semblent bénéficier d'un étiquetage favorable. Les jugements favorables augmentent au fur et à mesure que la carrière avance. Inversement, les jugements prescriptifs, particulièrement abondants en début de carrière, décroissent avec l'ancienneté. On peut noter toutefois que les enseignants en fin de carrière ne sont pas épargnés par les jugements défavorables. Ce sont eux qui sont le plus concernés par ce type de jugement. Le degré d'exigence des inspecteurs est accru auprès d'enseignants anciens jugés expérimentés. On peut également supposer que les inspecteurs les sentent moins disposés à mettre en œuvre les principes des méthodes actives qui organisent en filigrane le discours institutionnel local.

Les enseignants concernés par une guidance plus prononcée à visée explicitement prescriptive se trouvent parmi les femmes, les moins gradées, les plus jeunes et ceux exerçant en établissements défavori-

sés. Seraient-ils ciblés comme une population fragile nécessitant davantage d'accompagnement et de guidance dans l'exercice du métier ? Quoi qu'il en soit, l'exercice du pouvoir hiérarchique s'accroît auprès de ces populations.

Enfin, avant de conclure, une remarque s'impose. Les limites méthodologiques de l'enquête invitent à la prudence. Au regard des taux de significativité des tests statistiques utilisés ou encore du profil de notre échantillon, les résultats obtenus restent attachés à une situation particulière propre à une académie et à deux inspecteurs dotés de propriétés spécifiques. La recherche gagnerait à être élargie à des ensembles d'établissements, d'enseignants et d'inspecteurs plus diversifiés. Cependant, notons que les rapports d'inspection constituent des documents confidentiels difficiles d'accès au chercheur (10) qui ne peut guère mener ses investigations au-delà de ce dont il dispose.

CONCLUSION

Alors même que les enseignants inspectés sont exhortés à mieux prendre en compte leurs élèves, les élèves singuliers disparaissent dans le discours inspectoral au profit de l'évocation de principes didactiques et pédagogiques généraux qui placent toutefois l'apprenant au cœur de la démarche d'enseignement. En ignorant partiellement les pratiques réelles des élèves, l'inspecteur ne se prive-t-il pas et ne prive-t-il pas du même coup l'enseignant inspecté d'une dimension essentielle à l'évaluation et l'accroissement de l'efficacité de l'enseignement ? L'inspecteur ne formule aucun avis sur cette question cruciale : qu'apprennent réellement les élèves durant la leçon d'EPS ? On peut toutefois faire l'hypothèse que ce qui guide l'inspecteur c'est davantage le rappel de principes généraux exhortant à mieux prendre en compte les élèves, à organiser plus efficacement la classe et à analyser sa propre pratique que la transmission de « recettes pédagogiques » immédiatement applicables en situation. En contre partie, le discours inspectoral tend à s'uniformiser autour de principes généraux prônant principalement le recours aux méthodes actives et à un traitement différencié des élèves. Cette façon de faire écarte tout risque d'une réappropriation trop localisée des directives nationales, démarche localiste dont on connaît les effets inégalitaires, mais elle tend alors à produire une analyse standardisée des pratiques sans prise en compte des contextes particuliers d'enseignement.

Tableau IV. – Répartition des jugements selon le sexe des enseignants (9)

	Féminin		Masculin	
	N	%	N	%
Neutre	163,0	10,6	181,0	9,4
Défavorable	177,0	11,5	275,0	14,3
Favorable	727,0	47,4	1 036,0	53,8
Prescriptif	468,0	30,5	434,0	22,5
Ensemble	1 535,0	100,0	1 926,0	100,0

Cependant, des différenciations apparaissent à un autre niveau. Même si globalement le jugement inspectoral s'inscrit dans une logique d'encouragement et d'accompagnement, on a pu remarquer que sa sévérité s'accroît auprès des populations enseignantes les plus exposées aux difficultés d'enseignement (jeunes, moins gradés, exerçant en établissements défavorisés), ce qui ajoute sans doute à la précarité de l'intervention dans ces contextes d'enseignement. Ces distorsions favorisent un traitement inégalitaire des enseignants mais également, à terme, des élèves.

Enfin, le processus d'intervention est davantage pris en considération dans sa dimension personnelle que dans une perspective didactique. La survalorisation du poids des caractéristiques personnelles des enseignants dans la mise en œuvre de leur démarche d'enseignement est révélatrice de cette orientation alors que, dans le même temps, la référence aux contenus d'enseignement et aux compétences spécifiques visées disparaît. Les savoirs pratiques et théoriques que l'élève doit construire dans l'APSA support en vue de progresser ne sont que rarement évoqués.

Reste à savoir quel impact ce « discours d'autorité » peut avoir sur les pratiques réelles autrement dit, le rapport d'inspection est-il en mesure de constituer un outil de transformation des pratiques ? On peut effectivement s'interroger sur la puissance pragmatique de ce langage institutionnel dont le but avoué est de faire évoluer la réalité pédagogique. Depuis les travaux d'Austin (1970), on sait que tout énoncé linguistique vise, non seulement à transmettre des informations, mais également à tenter d'agir sur son interlocuteur. Mais il faut également noter que l'efficacité performative de ce discours ne peut s'exercer qu'avec la complicité ou la collaboration des enseignants qui reconnaissent dans l'inspecteur le dépositaire mandaté, autorisé du discours pédagogique légitime. Comme le souligne P. Bourdieu, « l'efficacité symbolique des mots s'exerce dans la mesure où celui qui la subit reconnaît celui qui l'exerce comme fondé à l'exercer ». Le discours inspectoral, en tant

que dépositaire d'une autorité reconnue et acceptée, est en mesure d'imposer sa vision légitime du pédagogique, et ce d'autant plus efficacement qu'il relaie, comme cela est le cas dans notre étude, le discours officiel national. Mais la distribution socialement différenciée des prescriptions, la place discrète accordée aux pratiques réelles des élèves ou encore la centration excessive sur les caractéristiques personnelles des enseignants nuisent peut-être à la transparence et la légitimité du message délivré. Cependant, on peut faire l'hypothèse que la convergence entre le projet à visée explicitement formative des inspecteurs et les attentes des enseignants principalement tournées vers la demande « de conseils plutôt que de jugements » (Sénore, 2000) contribuera, en situation, à dépasser ces limites. Il n'en reste pas moins que le temps écoulé entre les visites, la fréquence parfois insuffisante (variable selon les académies) de ces dernières, bref les conditions concrètes d'exercice de l'inspection atténuent quelque peu la portée de son action, d'autant plus que l'on sait que les conceptions et les croyances des enseignants évoluent et se transforment lentement.

Marie-Paule Poggi

marie-paule.poggi@univ-fcomte.fr

Université de Franche-Comté, Besançon

Laboratoire de sémio-linguistique, didactique et informatique/
Groupe de recherche en intervention sur les activités
physiques et sportives (LASELDI/GRIAPS)

Nathalie Wallian

nathalie.wallian@univ-fcomte.fr

Université de Franche-Comté, Besançon

Laboratoire de sémio-linguistique, didactique et informatique/
Groupe de recherche en intervention sur les activités
physiques et sportives (LASELDI/GRIAPS)

Mathilde Musard

mathilde.musard@univ-fcomte.fr

Université de Franche-Comté, Besançon

Laboratoire de sémio-linguistique, didactique et informatique/
Groupe de recherche en intervention sur les activités
physiques et sportives (LASELDI/GRIAPS)

NOTES

(1) Nous entendons par intervention « tout acte professionnel mettant en œuvre des compétences, des savoirs d'expérience et théoriques au service d'un objectif, dans différents champs (l'école, le club sportif, les loisirs actifs, la rééducation, la psychomotricité...) au profit de publics variés, aux différents âges de la vie » (Drouin *et al.*, 1988). Pour l'EPS, Durand précise qu'il convient d'entendre « toute action et situation individuelle ou collective à destination d'une ou plusieurs personnes engagées dans une activité physique, sportive ou artistique, et visant des modifications de cette activité » (Durand *et al.*, 1998).

(2) Nous entendons par processus sous-jacents aux pratiques enseignantes l'ensemble des éléments constitutifs d'une stratégie d'enseignement : statut accordé à l'élève, démarche d'enseignement privilégié, la nature des savoirs transmis, etc.

(3) Cette perspective pourrait être envisagée dans un travail ultérieur dans le cadre d'une approche compréhensive (entretiens au sein d'un échantillon contrasté d'inspecteurs).

(4) Cette notion de « contexte » est à distinguer de celles de situation, environnement ou encore milieu. Pour plus de précisions,

le lecteur pourra utilement consulter Bautier (2004) ainsi que Henriot-van Zanten (1990).

(5) Nous avons opté pour le découpage suivant :

début de carrière	De l'année de stage jusqu'à cinq ans d'ancienneté
milieu de carrière	six à huit ans d'ancienneté
fin de carrière	neuf à onze ans d'ancienneté

(6) Répartition des rapports selon les APSA servant de support à l'enseignement

APPN	Art	Athlétisme	Autre	Combat	Gym	Natation	Raquette	Sports collectifs	
5	4	24	2	9	27	3	11	26	111
4,5	3,6	21,6	1,8	8,1	24,3	2,7	9,9	23,4	100

(7) Nous disposons pour chaque établissement de deux informations : le pourcentage de Professions et catégories socioprofessionnelles (PCS) favorisées et le pourcentage de PCS défavorisées. Ces pourcentages ont été comparés aux pourcentages académiques pour ces deux indicateurs. Les établissements ont ensuite été répartis en cinq groupes selon les écarts à cette moyenne.

Écart à la moyenne académique	PCS favorisées	PCS défavorisées
- 35 à - 10	TD	TF
- 9.9 à - 3	D	F
- 2.9 à 3.9	Mixte	Mixte
4 à 10	F	D
10.1 à 51	TF	TD

Lecture du tableau : lorsque l'écart entre le pourcentage de PCS défavorisées de l'établissement et de l'académie est compris entre

- 35 et - 10, nous avons décidé que la tonalité sociale dominante de l'établissement était TF

Puis nous avons comparé pour chaque établissement les résultats obtenus dans les deux types de classification (PCS favorisées et PCS défavorisées) :

- pour cinquante-trois établissements les deux indicateurs coïncident (treize TD, neuf D, vingt-deux TF, trois F, six M)
- pour huit établissements le TF est nuancé par un F, ils sont alors classés en F
- pour onze établissements le TD est nuancé par un D, ils sont alors classés en D
- pour vingt-six établissements tous les indicateurs sont nuancés par un M, ils sont alors classés soit en F (14), soit en D (9), soit en TD (3)
- pour les huit établissements restants, un peu atypiques, nous avons tranché de la façon suivante : cinq associent D et TF, ils ont été classés en D, un associe F et TD, il a été classé en D et sept associent D et F, ils ont été classés en M.

Ce qui conduit, pour notre échantillon, à la répartition suivante :

Défavorisée	Favorisée	Mixte	Très défavorisée	Très favorisée	
35	25	13	16	22	111
31,5	22,5	11,7	14,4	19,8	100

(8) Le seul χ^2 significatif est obtenu lors du croisement de la variable PCS établissement et jugements valorisés dans le domaine de la prise en compte des élèves, tous les autres croisements ne donnent lieu à aucun chi carré significatif.

(9) Aucune différence n'est significative.

(10) Nous profitons de cette occasion pour remercier les deux inspecteurs qui nous ont accordés leur confiance en nous confiant les rapports d'inspection qui constituent notre corpus.

BIBLIOGRAPHIE

- AUSTIN J. L. (1970). *Quand dire c'est faire*. Paris : Éd. du Seuil.
- BARDIN L. (1983). *L'analyse de contenu*. Paris : PUF.
- BAUTIER É. (2004). « Contexte, situation et milieu ». In J.-F. Marcel & P. Rayou (dir.), *Recherches contextualisées en éducation*. Lyon : INRP, p. 223-234.
- BOURDIEU P. (1982). *Ce que parler veut dire*. Paris : Fayard.
- DROUIN D. ; GODBOUT P. & TOUSSIGNANT M. (1988). *La supervision de l'intervention en éducation physique*. Montréal : G. Morin.
- DURAND M. ; EUZET J.-P. & REFUGGI R. (1998). « L'intervention dans le domaine du sport et de la motricité humaine ». Communication présentée au colloque international *Recherche sur l'intervention en EPS et en sport* : Antibes-Juan-les-Pins, 16, 17 et 18 décembre 1998.
- FORQUIN J.-C. (1989). *École et culture : le point de vue des sociologues britanniques*. Bruxelles : De Boeck ; Paris : INRP.
- FRANCE : MINISTÈRE DE L'ÉDUCATION NATIONALE (1999) « La mutation des collèges : un collège pour tous et pour chacun ». *Bulletin officiel de l'Éducation nationale*, n° 23, 10 juin 1999 [supplément].
- FRANCE : CONSEIL NATIONAL DES PROGRAMMES (2002). *Qu'apprend-on au collège ? Cahier des exigences pour le collégien : pour comprendre ce que nos enfants apprennent*. Paris : CNDP ; Éd. XO.
- HENRIOT-VAN ZANTEN A. (1990). *L'école et l'espace local*. Lyon : Presses universitaires de Lyon.
- MAINGUENEAU D. (1998). « Les tendances françaises en analyse du discours ». Compte rendu de la conférence donnée à l'Université d'Osaka le 12 novembre 1998.
- PERRENOUD P. (1984). *La fabrication de l'excellence scolaire : du curriculum aux pratiques d'évaluation : vers une analyse de la réussite, de l'échec et des inégalités comme réalités construites par le système scolaire*. Genève : Droz.
- PLAISANCE É. (1986). *L'enfant, la maternelle, la société*. Paris : PUF.
- MONTEIL J.-M. (1999). *Propositions pour une nouvelle approche de l'évaluation des enseignants. Rapport à Monsieur le ministre de l'Éducation nationale, de la Recherche et de la Technologie*. Paris : MENRT.
- SÉNORE D. (2000). *Pour une éthique de l'inspection*. Paris : ESF.
- SI MOUSSA A. (2000). « Les rapports d'inspection du 1^{er} degré, contenus et finalités : un exemple à la Réunion ». *Revue française de pédagogie*, n° 133, p. 75-86.
- VOLUZAN, J. (1975). *L'école primaire jugée*. Paris : Larousse.
- WEBER, M. (1971). *Économie et société : 1. Les catégories de la sociologie*. Paris : Plon.
- YOUNG M. F. D. (1971). *Knowledge and Control : New Directions for the Sociology of Education*. London : Macmillan.

ANNEXES

Annexe I – Grille 2 : Objets du discours inspectoral. Analyse de contenu

Pour constituer cette grille, la procédure « par tas » a été préférée à celle « par boîtes » (Bardin, 1983) c'est-à-dire que les catégories ont été constituées au fur et à mesure de la lecture du contenu corpus de façon à n'ignorer aucun élément constitutif du jugement inspectoral.

Prise en compte des élèves

– développer le temps de pratique :

augmenter le volume d'activité des élèves, donner du temps aux apprentissages, favoriser une participation plus conséquente des élèves, mettre rapidement l'ensemble de la classe en action, veiller à ce que les élèves s'exercent suffisamment, quantifier avec précision l'activité de l'élève ;

– diversifier les itinéraires d'apprentissage, prendre en compte les caractéristiques des élèves :

aider tous les élèves à réussir, viser la réussite de tous, adaptation aux élèves des procédures et des contenus, donner du sens à ce que font les élèves, ajuster les situations, pertinence de la démarche au regard des caractéristiques des élèves, prise en compte du contexte spécifiques de la classe, mettre en valeur les conduites de réussite de vos élèves par des approches différenciées ;

– observer les élèves, contrôler les acquis :

ne pas esquisser le bilan effectué avec les élèves, observables à construire, mise en œuvre d'une co-évaluation, exploiter au maximum ses analyses et observations pour en dégager des perspectives pour la suite des apprentissages, mettre en place un dispositif de suivi et de contrôle, observations des élèves ;

– faire participer les élèves aux apprentissages :

pédagogie dialoguée, intégrer le projet des élèves, promouvoir une démarche active d'apprentissage, faire apprécier aux élèves le comment de la réussite, interroger les élèves pour analyser leur action, mettre l'élève en projet, faire participer l'élève à son apprentissage en le sollicitant pour observer, apprécier ses résultats, réfléchir à son action.

Qualités personnelles

– caractéristiques morales, affectives et relationnelles :

aisance liée à l'écoute des élèves, ambitieuse, attachement au métier, bonne volonté, calme et patience, capacité à gérer les phénomènes relationnels, volonté,

épanouie, qualités humaines, pédagogie souriante, dynamisme, enthousiasme, présence, naturel, simplicité, engagement, rayonnement, attitude positive, ressources de sa personnalité ;

– caractéristiques intellectuelles :

réfléchi, attitude sérieuse, consciencieux, permanence de la réflexion, rigueur, précision, rigoureux et exigeant, réalisme ;

– caractéristiques professionnelles et institutionnelles :

compétences pédagogiques, compétences professionnelles, efficacité de l'action de l'enseignant, expertise pédagogique, investissement dans le cadre du collège, maturité pédagogique, maîtrise professionnelle, excellent pédagogue, travail d'apprenant, acteur de la vie scolaire de premier ordre, responsabilités collectives, enseignement solide.

Gestion/organisation Communication/outils

– maîtrise des aspects gestionnaires et organisationnels (gestion du temps, de l'espace, répartition des rôles, rotation dans les ateliers) :

répartition des rôles entre les élèves, rotation dans les ateliers, aisance liée à la gestion, la conduite de la leçon, cohérence objectif / mise en œuvre, adéquation moyens par rapport aux objectifs, rigueur au niveau de la gestion de la séance, rigueur dans le respect des consignes, aisance dans la gestion des différents paramètres de la leçon situations judicieuses, parfaitement adaptées, excellentement organisées, mise en œuvre laborieuse, maîtrise de la classe (aspects disciplinaires), régulation de la leçon ;

– la communication et les outils (passation des consignes, gestion de la sécurité, media utilisés, ...) :

précision, clarté des attentes du professeur, passation des consignes, media utilisés, utilisation des fiches de travail, organiser l'observation des élèves, réduire le temps d'exposé au tableau, utilisation d'un outil formalisé et critérié comme aide précieuse pour l'enseignant.

Savoirs et culture

– connaissances, savoirs, compétences :

enrichir les connaissances et la compréhension des élèves, favoriser l'acquisition de nouvelles compétences, dimension ludique ne doit pas être occultée, consacrer la fin de la séance à des jeux aquatiques et relais, réduire la course d'élan, versant acroba-

tique en gym, mini-combat à thèmes en lutte, enrichir le savoir nager global des élèves par des modes d'entrée variés dans l'eau, reconsidérer la complexité de certaines figures gymniques, dégager un temps d'échauffement à base d'échanges de balle et de déplacements, apprentissage de la parade, aller vers plus de précision dans le geste, plus de rigueur dans les postures, pour rejoindre les principes fondamentaux de la gym, compétences spécifiques doivent créer un versant acrobatique ;

– valeurs et attitudes :

autonomie, responsabilité, ouverture culturelle, construction d'un code de vie avec les élèves, formation aux différents rôles sociaux, décliner des objectifs éducatifs prioritaires, construction du code de référence.

Recours au curriculum formel

– conformité aux programmes officiels, respect de la sécurité, devoir d'évaluation, etc. :

adéquation avec les programmes de la discipline.

Analyse de sa pratique

– conception, planification :

préparation des séances, inscrire son action dans le projet pédagogique et le projet d'établissement, énoncer avec pertinence les compétences attendues, formaliser son action, formaliser les CE, compétences visées, etc., construire un projet de cycle par activité, la référence à un projet de cycle permet de contextualiser son action ;

– capacité à analyser son action, à exploiter les bilans :

réflexion sur sa pratique ;

– autres :

se tenir informé (FPC), se perfectionner, enrichir ses compétences spécifiques en EPS, être auteur de l'évolution de la discipline, refuse de présenter CAPEPS, ne sollicite pas l'aide de ses collègues, insuffisance des connaissances disciplinaires, se doter de nouvelles connaissances sur la discipline.

Analyse des activités physiques sportives artistiques (APSA)

– respect de la logique de l'APSA :

respect de la logique de l'APSA, respect des fondamentaux de l'APSA, respect des éléments directeurs de l'activité support.

Activité des élèves

– activités des élèves liées à l'apprentissage de savoirs et savoir-faire :

stratégies mises en œuvre par les élèves, les élèves savent ce qui est à faire et ceci est déterminant dans un processus d'apprentissage, les élèves peuvent progresser, réussite constatée des élèves, les élèves apprennent et progressent, les élèves n'ont rien appris, bonne réalisation constatée de tous les élèves ;

– attitudes et comportements des élèves :

les élèves apprennent et réalisent différents rôles sociaux.

**Annexe II – Répartition des jugements selon la tonalité sociale dominante de l'établissement
(pour « D » il faut lire défavorisé, pour « TD » il faut lire très défavorisé, pour « F » il faut lire favorisé,
pour « TF » il faut lire très favorisé, pour « M » il faut lire mixte)**

	M		TF et F		D et TD	
	N	%	N	%	N	%
JAN	46	7,1	141	11,1	157	10,2
JAD	32	4,9	131	10,3	289	18,8
JAF*	480	73,7	714	56,0	569	37,0
JP	93	14,3	288	22,6	521	33,9
	651	100	1 274	100	1 536	100

* χ^2 significatif

NOTE DE SYNTHÈSE

Climat motivationnel instauré par l'enseignant et implication des élèves en classe : l'état des recherches

*Philippe Sarrazin, Damien Tessier,
David Trouilloud*

Cet article a pour objectif de présenter l'état des recherches sur les pratiques enseignantes et leurs effets sur la motivation des élèves. Effectuée principalement dans le cadre de deux paradigmes théoriques – la théorie de la motivation à l'accomplissement et la théorie de l'autodétermination – la recension de ces travaux montre d'une part les conséquences positives d'un climat de maîtrise et les effets controversés d'un climat de compétition, et d'autre part les bienfaits d'un style soutenant l'autonomie des élèves au contraire d'un style contrôlant. Néanmoins, l'application à la classe des principes éducatifs issus de ces études, requiert de la part des enseignants une certaine prudence. La complexité du contexte naturel d'enseignement par rapport aux situations étudiées en laboratoire, les contraintes qui s'imposent à l'enseignant, qui réduisent sa liberté pédagogique et qui le poussent à être contrôlant envers ses élèves, ainsi que les limites méthodologiques des études par questionnaire, constituent les principales réserves à avoir quant à la validité écologique de ces recherches.

Descripteurs (TEE) : autonomie, conduite de la classe, élève, motivation d'accomplissement, motivation pour les études, style pédagogique.

« **M**otiver, motiver, comment les motiver » titrait *Le Monde de l'Éducation* dans son numéro d'avril 2004, se faisant l'écho de la fameuse « question 8 » du « Grand débat sur l'école » organisé par Luc Ferry le ministre de l'Éducation nationale de l'époque. Lors des 20 590 réunions de ce grand débat, la *motivation* est arrivée largement en tête des thèmes abordés (1). S'affirmant comme la première préoccupation des enseignants et des parents, la (dé)motivation des élèves est en passe de devenir une priorité pédagogique. Face à l'évolution des nouveaux publics

scolaires caractérisés, entre autres, par l'augmentation des comportements en marge des normes sociales ou par des tendances consuméristes (Truong, 2003), l'enseignant semble devoir être plus que jamais un « motivateur ».

Sensibles à cette préoccupation sociale grandissante, les travaux en psychologie de l'éducation se sont multipliés ces vingt dernières années sur le thème de la motivation. La *Revue française de pédagogie* a d'ailleurs récemment consacré un numéro à certains de ces travaux (2006, n° 155). Dans ceux-ci, le rôle de l'enseignant et plus précisément du « climat motivationnel » qu'il met en place, sont particulièrement soulignés. De par les comportements qu'il met en œuvre, les activités d'apprentissage qu'il propose, la nature de ses interactions avec les élèves, le climat psychologique qu'il aménage, les *feedbacks* qu'il délivre, l'enseignant est susceptible de « faire la différence », en terme d'apprentissage et plus généralement de réussite scolaire des élèves (voir Good & Brophy, 2000). Des recherches ont d'ailleurs quantifié cet « effet-enseignant » (2) dont le poids est estimé à environ 16 % sur les acquisitions scolaires (Duru-Bellat & Mingat, 1994).

L'objectif de cette revue de littérature est précisément de montrer que le climat motivationnel instauré par l'enseignant dans sa classe constitue une variable importante à analyser pour comprendre l'investissement motivationnel des élèves à l'école. Plus précisément, l'ambition est de comprendre comment les différentes dimensions du climat motivationnel facilitent ou au contraire entravent la motivation des élèves envers les apprentissages scolaires. Après un bref historique des études sur les comportements de l'enseignant, nous développerons les deux paradigmes théoriques à l'intérieur desquels les travaux sur ce thème ont été pour la plupart réalisés. Nous aborderons en premier lieu les travaux effectués dans le cadre de la théorie des buts d'accomplissement (Elliot & Dweck, 2005). Nous soulignerons les conséquences positives d'un climat de maîtrise et les effets controversés d'un climat de compétition. Nous porterons ensuite un autre regard sur le climat motivationnel au travers de la théorie de l'autodétermination (Deci & Ryan, 2002). Nous montrerons que les évolutions de cette théorie ont permis le passage d'une conception bidimensionnelle (le contrôle *contra* le soutien de l'autonomie) à une vision multidimensionnelle du climat. Enfin, après une synthèse de l'ensemble de ces travaux, nous évoquerons dans une troisième partie les précautions que les enseignants devraient prendre pour appliquer au contexte de la classe les principes éducatifs mis en exergue dans ces études.

LES PREMIÈRES ÉTUDES SUR LES COMPORTEMENTS DE L'ENSEIGNANT

C'est au travers du concept de *style* que les premières recherches sur l'intervention ont éclairé l'étude des comportements de l'enseignant. Ce terme a commencé à être utilisé dans les années cinquante pour catégoriser les comportements des enseignants réellement mis en œuvre dans l'action (Altet, 1993). Il a d'abord été employé de façon opératoire pour faciliter l'identification des pratiques réelles des enseignants dans leur classe et devenir ainsi un outil descripteur de pratiques, utilisable en formation d'enseignants. Le courant des recherches nord-américaines des années 1950-1970 s'est efforcé de caractériser des styles globaux de façon dichotomique et d'examiner leurs effets sur l'apprentissage des élèves. Les styles direct-indirect de N. A. Flanders (1970), formel, informel ou mixte de S. I. Bennett (1976), ou autoritaire, démocratique et laisser-faire de K. Lewin, R. Lipitt et R. K. White (1939) sont caractéristiques de ces recherches. Par la

suite, les travaux des années 1970-1980 s'attachèrent en revanche à l'analyse des styles d'enseignement dans une perspective multidimensionnelle, la réalité pédagogique se laissant difficilement saisir par une vision dichotomique, chaque dimension étant constituée par une combinaison de comportements. Par exemple, les travaux de S. I. Bennett (1976) ont permis de repérer douze styles à partir de diverses dimensions de la conduite de la classe (la façon d'organiser la classe, les stratégies d'intervention, les techniques de motivations, etc.).

En résumé, dans la recherche en éducation, ce concept de style a été employé selon deux acceptions différentes : 1) soit quelque chose de « global », c'est-à-dire un instrument de généralisation, de classification pour recenser des comportements caractéristiques, des traits communs à plusieurs enseignants ; 2) soit au contraire quelque chose de « différencié », c'est-à-dire un outil de singularisation pour repérer le style personnel à chaque enseignant, sa manière d'être et de faire, sa facture propre. Tous ces travaux, qu'ils s'inscrivent dans l'une ou l'autre de ces deux acceptions, mettent en évidence qu'il n'existe pas de styles purs définis dans l'absolu, mais des combinaisons variées, des styles intermédiaires (Postic & De Ketele, 1988), chacun représentant plutôt « le reflet d'un mode d'enseignement dominant » (Provencher, 1983).

Depuis une trentaine d'années, des travaux de plus en plus nombreux se sont intéressés aux liens entre l'environnement d'apprentissage de la classe orchestré par l'enseignant, les buts et motivations des élèves et les conséquences de ceux-ci, par exemple, en terme d'effort manifesté, de stratégies utilisées ou de résultats obtenus. Dans cette ligne de recherche, le concept de style a été moins systématiquement utilisé au profit d'autres termes comme ceux d'organisation de la classe, de structuration des buts, de style motivationnel, d'environnement véritable ou perçu. À l'instar de certains auteurs, nous utiliserons dans cet article le concept générique de « climat motivationnel » pour faire allusion à l'environnement psychologique de la classe, qui oriente les buts et les motivations de l'élève (Ames, 1992a). Cette variable est assez similaire à celle de « contexte », utilisée dans les travaux récents en psychologie de l'éducation (Monteil & Hugué, 2002 ; Drozda-Senkowska & Hugué, 2003). Dans ces derniers, le contexte correspond à la fois à l'environnement du sujet et à son état interne. L'état interne fait référence aux différents éléments récupérés en mémoire par l'individu placé dans une situation donnée et qui sont susceptibles de faciliter ou d'inhiber les performances. Ainsi le climat motivationnel peut se définir comme étant un « contexte » dans la mesure où il se caractérise non seulement par l'environnement d'apprentissage mis en place par l'enseignant, ses comportements observables (par ex. son style), mais également par le fait qu'il est susceptible d'activer certains états internes chez l'élève en fonction des perceptions qu'il a de cet environnement. Dans la mesure où les études que nous allons passer en revue, sont centrées plus exclusivement sur les aspects « motivationnels » du contexte, nous préférons utiliser dans la suite de cet article le terme de « climat motivationnel », à celui plus général de contexte.

Notre préoccupation est de montrer comment les études actuelles se sont emparées de la question du rôle catalyseur de l'enseignant dans la motivation et les performances scolaires des élèves. Pour y répondre, nous avons recensé à partir des bases de données *PsychInfo*, *Francis* et *Current Content*, les articles portant sur cette question, contenant les mots clés « climat motivationnel » (*motivational climate*), « structuration des buts de la classe » (*class goal structure*) et « comportements de l'enseignant » (*teacher's behaviors*), publiés depuis 1990. Environ 200 études ont satisfait ces critères, la plupart s'inscrivant dans deux théories majeures de la motivation : la théorie des buts d'accomplissement

(Elliot & Dweck, 2005) pour environ 80 % des publications et la théorie de l'auto-détermination (Deci & Ryan, 2002) pour le reste. Nous allons à présent passer en revue ces différents travaux.

LA THÉORIE DES BUTS D'ACCOMPLISSEMENT (TBA) : BERCEAU DU CONCEPT DE « CLIMAT MOTIVATIONNEL »

Postulats théoriques

Initialement développée dans le champ de l'éducation grâce aux propositions de C. Ames (1992a & 1992b), C. S. Dweck (1999), M. L. Maehr (Maehr & Nicholls, 1980 ; Maehr & Braskamp, 1986), J. G. Nicholls (1984 & 1989) et plus récemment A. J. Elliot (1997), la théorie des buts d'accomplissement est devenue un paradigme théorique majeur en psychologie sociale au cours de ces quinze dernières années (pour une revue détaillée de ces travaux, voir Elliot & Dweck, 2005).

Dans ce qu'il a d'essentiel, ce paradigme théorique repose sur le postulat selon lequel l'un des buts prioritaires des individus dans les contextes d'accomplissement tels que l'école est de « développer ou manifester – à soi ou aux autres – une compétence élevée et d'éviter de paraître incompetent » (Nicholls, 1984, p. 328). La plupart des théoriciens (Ames, 1992a ; Dweck & Legett, 1988 ; Nicholls, 1989) s'accordent sur l'existence de deux manières de manifester sa compétence, renvoyant chacune à un ou deux buts déterminés : soit en se comparant aux autres et en faisant preuve d'une supériorité (certains comme J. G. Nicholls parlent d'implication de l'ego, d'autres comme A. J. Elliot de but de performance), soit en progressant dans le temps (J. G. Nicholls parle ici d'implication dans la tâche et A. J. Elliot de but de maîtrise). Les conséquences des buts impliquant l'ego et de maîtrise ont été examinées dans un grand nombre d'études conduites en situation de classe et en laboratoire. Dans l'ensemble, les résultats montrent une supériorité du but de maîtrise par rapport au but impliquant l'ego. Par exemple, les buts de maîtrise sont associés à un choix de tâche de défi (Ames & Archer, 1988), à l'implication de l'élève dans le processus d'apprentissage (Nicholls *et al.*, 1989), et à des stratégies de travail efficaces (Nolen & Haladyna, 1990). Par contraste, les chercheurs ont souligné les conséquences non adaptatives du but impliquant l'ego telles que des émotions négatives et la réduction de l'effort (Ames, 1992a ; Hidi & Harackiewicz, 2000).

Parallèlement à la mise en évidence des conséquences singulières de ces buts, les travaux se sont intéressés aux antécédents de ces derniers. La théorie présuppose en effet que le contexte dans lequel évolue l'individu constitue l'un des déterminants principaux de ses états motivationnels. Par exemple, selon J. G. Nicholls (1984 & 1989), un but impliquant l'ego a plus de chance d'être suscité, quand les tâches (surtout si elles impliquent des habiletés socialement valorisées) sont présentées comme des tests (faisant référence à des normes situant le bon, le moyen et le mauvais), dans une situation de compétition ou de comparaison sociale, ou quand la facette publique du Soi est activée, en particulier en présence d'une audience, d'une caméra, etc. D'un autre côté, un but de maîtrise a plus de chance d'être suscité dans des contextes qui minimisent la facette publique du Soi et les évaluations sociales, et qui simultanément mettent l'accent sur le processus d'apprentissage, la maîtrise de tâches adaptées au niveau de l'individu, l'investissement et les progrès.

Des dizaines d'études ont manipulé expérimentalement les buts des participants en leur fournissant certaines informations juste avant qu'ils ne s'engagent dans une tâche (Elliot & Harackiewicz, 1996 ; voir Dweck & Leggett, 1988 ; Harackiewicz, Barron & Elliot, 1998 pour une revue de cette littérature). Généralement, l'information apportée par les expérimentateurs incitait les participants soit à faire de leur mieux sur la tâche, à y prendre du plaisir ou à essayer de progresser (un but de maîtrise), soit à se comparer aux autres ou à des normes (un but impliquant l'ego). Par exemple, pour induire un tel but auprès d'enfant de CM2, A. J. Elliott et D. S. Dweck (1988) ont annoncé aux participants de leur étude, que leur performance serait filmée et jugée relativement à celle de leurs camarades. Pour susciter un but de maîtrise, les expérimentateurs ont au contraire insisté sur l'intérêt que présentait la tâche pour « affûter »/développer leur raisonnement, quelque chose qui pouvait les aider dans leur travail scolaire.

Les résultats de ces études en laboratoire ont démontré à maintes reprises la facilité avec laquelle il semblait possible de manipuler le but poursuivi. Dès lors, s'interrogeant sur les similarités entre le contexte du laboratoire et celui de l'école, certains chercheurs ont tenté de voir s'il en était de même en classe. En mettant l'accent sur certaines informations, les consignes de l'enseignant sont susceptibles de créer un climat particulier, à même de faciliter la poursuite d'un but particulier (Ames, 1992a & 1992b ; Ames & Archer, 1988 ; Maehr & Midgley, 1991 ; Urdan, Kneisel & Mason, 1999).

Climat de maîtrise *contra* climat de compétition

On doit à C. Ames (1992a & 1992b), les premiers travaux sur les pratiques de l'enseignant susceptibles de créer une structure de buts ou un climat motivationnel particulier. C. Ames et ses collaboratrices ont décrit différentes dimensions du processus d'enseignement à même de créer, d'après les termes de l'auteur, un « climat de maîtrise » ou un « climat de compétition ». L'enseignant instaure un climat de maîtrise quand ses interventions sont principalement orientées vers l'apprentissage, les progrès personnels et la valorisation du travail et des efforts. Par contraste, il installe un climat de compétition quand la comparaison sociale, la compétition interpersonnelle, et le résultat final sont valorisés.

Pour distinguer les éléments constituant ces climats, C. Ames s'est appuyée sur les variables utilisées par J. L. Epstein (1988) pour organiser les interactions en classe. Celles-ci sont résumées dans l'acronyme TARGET, pour Tâche (nature et structure du matériel à apprendre), Autorité (délégation ou non de certaines prises de décisions aux élèves), Reconnaissance (renforcement des comportements valorisés par l'enseignant), Groupement (organisation et fréquence des regroupements), Évaluation (modalité et standard pour la notation) et Temps (gestion du temps d'apprentissage). Ces variables sont à la fois des formalisations théoriques et le produit de convergences empiriques issues de l'étude de la littérature (voir Ames, 1992b). Dans ce cadre, le climat de compétition se caractérise comme une structure dans laquelle 1/ tous les élèves pratiquent la même tâche, 2/ l'enseignant prend toutes les décisions concernant ce qu'il y a à apprendre et l'installation du matériel, 3/ les encouragements sont destinés à valoriser les meilleures performances, 4/ des groupes de niveau sont explicitement organisés, 5/ l'évaluation revêt un caractère public et se fonde sur des standards sociaux de performance tel qu'un barème ou le niveau de la classe, 6/ le temps imparti pour réaliser les tâches est défini par l'enseignant et ne tient généralement pas compte des différences dans le rythme d'apprentissage des élèves.

À l'inverse, il y a climat de maîtrise 1/ quand les élèves ont la possibilité de choisir entre différentes tâches en fonction de leur niveau d'habileté, quand l'activité est structurée pour favoriser la variété, et le défi individuel dans l'apprentissage, 2/ quand les élèves peuvent choisir l'objet d'apprentissage sur lequel ils vont s'investir ou ont la possibilité de prendre des initiatives dans le processus d'apprentissage, 3/ quand l'enseignant reconnaît et encourage les efforts et les progrès réalisés et qu'il considère les erreurs non comme des indicateurs de moindre habileté, mais comme des éléments consubstantiels de l'apprentissage, 4/ quand les formes de groupement sont flexibles et hétérogènes (tâches individualisées, groupes de besoin, groupes hétérogènes valorisant la coopération), 5/ quand l'évaluation est délivrée de manière confidentielle et fondée sur des standards personnels de performance tels que les apprentissages et les progrès réalisés, la participation, les efforts et l'investissement en classe, 6/ quand la gestion du temps d'apprentissage est flexible et permet aux élèves de travailler à leur rythme. Le tableau I résume les différentes dimensions d'un climat de maîtrise et les stratégies qui devraient être valorisées par l'enseignant pour le mettre en place.

Conséquences pour l'élève d'un climat motivationnel de maîtrise contra un climat de compétition

Pour examiner les effets du climat de la classe sur la motivation et la réussite des élèves, les chercheurs se sont basés principalement sur des études utilisant des questionnaires. Dans celles-ci, la « perception » du climat de la classe (susceptant la maîtrise et/ou la compétition), la motivation des élèves (ego *contra* tâche) et différentes variables affectives (le plaisir en classe), cognitives (les stratégies d'études utilisées, la confiance en soi) et comportementales (le temps consacré à faire son travail) ont été mesurées par des outils appropriés. Quelques études plus rares ont tenté de renforcer la validité des résultats en allant observer directement dans les salles de classes si les comportements « véritables » (et non « perçus ») de l'enseignant étaient reliés à une motivation particulière des élèves (Meece, 1991). Des études encore plus rares, ont tenté de manipuler le climat de la classe en jouant sur les dimensions du TARGET afin de tester l'impact de celles-ci (Goudas *et al.*, 1995). Les résultats de ces recherches sont dans l'ensemble, assez consistants avec ceux des études destinées à examiner les conséquences des buts d'accomplissement.

Dans l'une de leurs premières études sur le climat motivationnel, C. Ames et J. Archer (1988) se sont intéressées à la signification qu'accordaient les élèves à l'environnement scolaire et à ses effets sur l'adoption de stratégies d'études et de comportements particulièrement adaptés à celui-ci. Les résultats ont suggéré que le climat perçu de la classe pouvait conduire à des comportements motivationnels facilitant les apprentissages lorsqu'il rendait saillants les buts de maîtrise. S'appuyant sur les travaux antérieurs effectués dans le domaine de l'apprentissage (Weinstein, Schulte & Palmer, 1987), ces chercheurs ont montré que les élèves qui percevaient de manière dominante un climat de maîtrise ont rapporté utiliser des stratégies d'études plus efficaces (un traitement plus en profondeur du matériel à apprendre, plutôt qu'un traitement superficiel de type « par cœur »), s'être orientés vers des tâches de défi, avoir une attitude plus positive envers le cours, et croire fortement à l'effet déterminant de l'effort sur la réussite. À l'inverse, les élèves qui ont perçu de manière saillante un climat de compétition ont eu tendance à se focaliser sur leur niveau de compétence par rapport à celui des autres élèves (et non sur l'effort) et à attribuer leurs échecs à un manque de compétence.

Tableau 1. – **Résumé des éléments d'un climat de maîtrise et des stratégies susceptibles de le nourrir**
(d'après Maehr & Midgley, 1991)

Domaines du <i>TARGET</i>	Focalisation	Stratégies
Tâche	Valeur intrinsèque de l'apprentissage	<ul style="list-style-type: none"> • Réduire la prégnance des incitateurs extrinsèques (promesse de récompense) • Proposer des exercices qui posent des défis à tous les élèves • Insister sur les buts et objectifs de l'apprentissage • Insister sur le plaisir d'apprendre
Autorité	Participation des élèves dans les prises de décisions concernant l'apprentissage et plus généralement l'école	<ul style="list-style-type: none"> • Fournir des opportunités de développer la responsabilité et l'indépendance • Développer les compétences à s'auto-réguler
Reconnaissance	Type et utilisation des reconnaissances et récompenses	<ul style="list-style-type: none"> • Offrir à tous les élèves des opportunités d'être reconnus (pas uniquement les plus forts) • Reconnaître le progrès personnel dans la maîtrise des contenus à apprendre • Reconnaître les efforts réalisés dans une large gamme d'activités d'apprentissage (et pas uniquement dans les activités les plus « nobles »)
Groupement	Interaction entre élèves, compétences sociales, valeurs	<ul style="list-style-type: none"> • Construire un environnement d'acceptation et de valorisation de tous les élèves • Élargir les formes d'interaction sociale, en particulier pour les élèves en difficulté • Développer les compétences sociales • Encourager les valeurs humaines • Construire un environnement dans lequel tous les élèves ont le sentiment d'apporter une contribution significative
Évaluation	Types et modalités d'évaluation/notation	<ul style="list-style-type: none"> • Augmenter le sentiment de compétence et d'efficacité des élèves. • Faciliter la prise de conscience des progrès réalisés • Faire accepter aux élèves que l'échec fait partie de l'apprentissage (et même de la vie)
Temps	Gestion du temps pour atteindre les buts fixés	<ul style="list-style-type: none"> • Améliorer le nombre d'exercices réalisés (en particulier par les élèves les plus « lents ») • Améliorer les capacités à planifier et à s'organiser, en bref, à s'auto-gérer. • Faire en sorte que ça soient les tâches d'apprentissage et les besoins des élèves qui dictent les temps de travail et non d'autres paramètres comme les programmes à bouclier à tout prix

Dans l'ensemble les résultats des études subséquentes sont congruents avec les travaux originels de C. Ames et J. Archer (1988). La perception d'un climat de maîtrise est généralement reliée positivement à des variables motivationnelles, affectives, cognitives et comportementales favorables aux acquisitions scolaires (pour une revue de ces travaux, voir Galand *et al.*, 2006 ; Kaplan & Maehr, 2002 ; Meece *et al.*, 2006 ; Ntoumanis & Biddle, 1999 ; Urda *et al.*, 1999). Sur le plan *motivationnel*, le climat de maîtrise est corrélé positivement avec les buts de maîtrise (Anderman & Midgley, 1997 ; Midgley & Urda, 1995 ; Wolters, 2004), la motivation intrinsèque (Biddle *et al.*, 1995 ; Cury *et al.*, 1996 ; Escarti & Gutierrez, 2001 ; Meece, 1991) et le désir de pratiquer les tâches en dehors de l'école (Escarti & Gutierrez, 2001 ; Wolters, 2004). Au niveau *affectif*, il est relié à des affects positifs (Urda & Midgley, 2003), à une satisfaction envers l'apprentissage (Biddle *et al.*, 1995 ; Cury *et al.*, 1996 ; Escarti & Gutierrez, 2001 ; Meece, 1991), et à une faible anxiété (Papaioannou, 1994 & 1995). Au niveau *cognitif*, il prédit positivement l'utilisation de stratégies d'études efficaces (Wolters, 2004), la recherche autonome de l'aide du professeur (Karabenick, 2004 ; Ryan *et al.*, 1998), la préférence pour des tâches de défi, dont la difficulté excède les possibilités du moment de l'élève, mais dont la réussite semble possible s'il s'en donne les moyens (Sarrazin, Famose & Cury, 1995 ; Treasure & Roberts, 2001 ; Ames, 1992a), la perception d'un traitement équitable de l'enseignant entre garçons et filles (Cervello *et al.*, 2004), une habileté perçue élevée (Escarti & Gutierrez, 2001 ; Midgley *et al.*, 1995 ; Treasure, 1997 ; Urda & Midgley, 2003) et une utilisation réduite de stratégies d'auto-handicap comme la procrastination (Wolters, 2004) ou les comportements d'évitement (Turner *et al.*, 2002). Sur le plan *comportemental* enfin, le climat de maîtrise est lié positivement à l'effort consenti dans le travail et à la persévérance face aux difficultés (Wolters, 2004). Toutefois, en ce qui concerne les performances des élèves (attestées, par exemple, par leurs notes), les résultats sont un peu plus inconsistants. Si certaines études révèlent des corrélations positives (Midgley & Urda, 2001 ; Wolters, 2004), d'autres font état de corrélations non significatives (Anderman & Anderman, 1999 ; Anderman & Midgley, 1997), voire négatives (Murdock *et al.*, 2001). Cette inconsistance entre le climat de maîtrise et les performances des élèves constitue l'une des « anomalies » de la théorie des buts d'accomplissement (Meece *et al.*, 2006). Elle s'explique en partie par la manière avec laquelle sont mesurées les performances académiques des élèves. Grant et Dweck (2003) ont récemment montré que les buts de maîtrise étaient fortement reliés aux performances académiques des élèves quand les tâches à réaliser contiennent des défis d'apprentissage importants ou quand elles requièrent la maîtrise de contenus complexes, en bref, quand elles nécessitent des stratégies de traitement en profondeur des exercices. En revanche, quand les exercices à réaliser ne sont pas particulièrement difficiles, les buts de maîtrise ne sont pas reliés à la performance. Or la plupart des études ne tiennent pas particulièrement compte du niveau de difficulté des tâches (Meece, Herman & Mc Combs, 2003), alors que cette variable semble constituer un modulateur important des liens buts de maîtrise-performance.

Concernant les corrélats d'un climat de compétition, les résultats sont moins consistants que ceux du climat de maîtrise. Dans l'ensemble, ils tendent à faire ressortir des liens modérés mais négatifs avec des variables bénéfiques aux apprentissages scolaires. Par exemple, la perception d'un climat de compétition est négativement reliée à une préférence pour des tâches de difficulté optimale (Sarrazin *et al.*, 1995 ; Treasure & Roberts, 2001), à la satisfaction et la motivation intrinsèque (Papaioannou, 1995 ; Treasure & Roberts, 2001), à la perception d'un traitement équitable de l'enseignant entre garçons et filles et à l'adoption de comportements disciplinés en classe (Cervello *et al.*, 2004).

En outre, cet environnement favorisant la comparaison sociale entretient des liens positifs avec des variables plutôt défavorables aux apprentissages scolaires, comme une forte anxiété (Papaioannou, 1994 & 1995), le recours à des stratégies d'auto-handicap (Midgley & Urda, 2001), la tricherie (Anderman, Griesinger & Westerfield, 1998), la diminution des efforts et de la persévérance notamment face à des tâches considérées comme difficiles (Wolters, 2004). Concernant les liens entre le climat de compétition et les résultats scolaires des élèves (les notes par exemple), les résultats sont plus inconsistants. Si la théorie prédisait une relation négative, peu de travaux ont confirmé un tel lien (Sarrazin *et al.*, 2002) ; plusieurs études ne font ressortir aucun lien entre ces deux variables, et il arrive même parfois qu'un tel climat soit positivement relié à la performance des élèves (Midgley & Urda, 1995 ; Murdock *et al.*, 2001). Nous avancerons plus loin (*infra*, « Le climat de compétition : des effets controversés ») quelques explications à ces résultats.

Créer un climat motivationnel de maîtrise : une piste pour optimiser l'engagement des élèves ?

Dans le cadre de la théorie des buts d'accomplissement, et d'une manière générale dans le champ de la motivation, peu de travaux ont adopté une procédure quasi-expérimentale visant à manipuler le climat en situation naturelle d'enseignement afin d'en tester les effets sur la motivation et les résultats des élèves. Les rares études existantes ont cherché à comparer les effets (en terme de motivation intrinsèque, de perception de compétence, d'efforts consentis, de degrés de satisfaction en classe, etc.) d'un climat de maîtrise *contra* un climat de compétition, ou les effets d'un programme de formation des enseignants à la mise en œuvre d'un climat de maîtrise avec des classes « contrôles » dont les professeurs n'avaient pas bénéficié d'une telle formation.

Par exemple, D. A. Weigand et S. Burton (2002) ont induit expérimentalement un climat de maîtrise en différenciant la difficulté des tâches d'apprentissage en fonction du niveau des élèves, en leur laissant choisir l'ordre selon lequel ils souhaitaient résoudre ces tâches, en complimentant les élèves sur la base de leurs progrès et non de leurs performances absolues, en atténuant les possibilités de comparaison sociale, en notant les progrès et les efforts réalisés et en passant autant de temps avec chaque élève indépendamment de son niveau. Les résultats révélèrent que, par rapport aux élèves du groupe contrôle ayant suivi un enseignement traditionnel, ces conditions de travail mises en place au sein du groupe expérimental ont induit une propension plus marquée à poursuivre des buts de maîtrise, une perception de compétence plus élevée et un plus haut niveau de satisfaction.

D'autres travaux, à l'image de M. Goudas *et al.* (1995) ont utilisé les dimensions du TARGET pour créer l'un ou l'autre des deux climats afin de comparer leurs effets. N'utilisant que trois dimensions (tâche, autorité et temps), ils ont comparé les effets d'un climat d'enseignement qualifié de « directif » à ceux d'un climat dit « différencié ». Dans un climat directif, la plupart des décisions sont prises par l'enseignant : le type de tâche, la durée de l'exercice et le degré de difficulté. C'est un climat faiblement orienté vers la maîtrise. *A contrario*, un climat différencié est un environnement facilitant l'apprentissage. Les élèves ont des choix à faire, une variété d'activités est proposée, le degré de difficulté et le rythme d'apprentissage sont déterminés par l'élève. Les résultats ont montré des effets plus positifs de ce type de climat par rapport au climat directif, en particulier une plus grande motivation intrinsèque et des intentions plus fermes de participer aux leçons suivantes.

Les travaux les plus marquants dans ce domaine sont ceux de E. M. Anderman, M. L. Maehr et C. Midgley (Anderman & Urdan, 1995 ; Anderman, Maehr & Midgley, 1999 ; Maehr & Anderman, 1993 ; Maehr & Midgley, 1996 ; Midgley & Maehr, 1999). Ces auteurs et leurs collaborateurs se sont engagés dans plusieurs projets destinés à réformer l'école et les pratiques éducatives en Amérique du Nord, à partir de la théorie des buts d'accomplissement. Pour résumer, ces chercheurs ont travaillé avec des enseignants, des parents et des administrateurs à la fois dans des écoles élémentaires et des collèges afin d'examiner et de modifier les pratiques éducatives dominantes. Des groupes de travail se sont réunis pendant plusieurs années afin d'apporter un regard critique sur les politiques et pratiques éducatives à la lumière des buts de maîtrise et impliquant l'ego. Utilisant un plan quasi expérimental, l'étude de C. Midgley et M. L. Maehr (1999) par exemple, a montré que les enseignants pouvaient changer leurs pratiques et faciliter la poursuite de buts de maîtrise chez les élèves. Ces changements entraînaient plusieurs conséquences positives parmi celles soulignées plus haut.

Une autre étude (Anderman *et al.*, 1999) est partie du constat d'une accentuation plus marquée sur les buts impliquant l'ego lors du passage de l'école primaire (*elementary school*) au collège (*high school*). L'objectif de cette étude était d'impulser une réforme du fonctionnement de l'établissement afin d'aider les enseignants à minimiser de tels buts et à valoriser un climat de maîtrise. Les résultats révélèrent l'efficacité de ce programme de recherche. Contrairement aux élèves de l'établissement contrôle – à l'intérieur duquel les expérimentateurs ne sont pas intervenus – les élèves de l'établissement expérimental n'ont pas rapporté de perception accrue d'un climat de compétition pendant la phase de transition.

Comme nous venons de le voir dans cette partie, les conclusions des recherches révèlent les effets bénéfiques d'un climat de maîtrise sur plusieurs variables susceptibles de favoriser les apprentissages scolaires (les stratégies d'études utilisées, l'effort), tout en gardant à l'esprit que ses liens avec les résultats scolaires ne sont pas toujours confirmés sauf quand les exercices à réaliser nécessitent un travail en profondeur. Concernant le climat de compétition les résultats sont plus nuancés.

Le climat de compétition : des effets controversés

À s'intéresser de plus près au climat de compétition, on découvre que cet environnement ne produit pas systématiquement des effets néfastes sur les conduites des élèves. Le plus souvent, il est faiblement négativement, ou pas relié du tout aux comportements favorables aux apprentissages. Par exemple, dans l'étude de C. Ames et J. Archer (1988), la perception d'un climat de compétition n'était pas reliée aux stratégies d'études des élèves, ni à la recherche de tâche de défi. Dans celles de N. Digledis *et al.* (2003) et de A. Escarti et M. Gutierrez (2001), ce climat ne prédisait pas l'effort consenti par les élèves dans l'apprentissage, et dans celle de A. Papaioanou, H. W. Marsh et Y. Theodorakis (2004) il ne prédisait pas non plus la motivation intrinsèque des élèves. Autrement dit, si les effets de ce type de climat ne sont pas positifs (comme c'est le cas avec un climat de maîtrise), ils ne sont pas non plus toujours négatifs. Il arrive même parfois qu'il soit relié positivement aux performances scolaires, comme nous l'avons souligné plus haut (Midgley & Urdan, 1995 ; Murdock *et al.*, 2001).

Ces résultats ont interpellé plusieurs spécialistes du domaine comme S. Hidi et J. M. Harackiewicz (2000) qui plaident pour un examen plus systématique de

ce type de climat : « Les conséquences positives des buts de compétition ont été jusqu'à présent sous estimées, et nous pensons qu'il est crucial de considérer la possibilité que les buts de compétition puissent promouvoir des comportements d'accomplissement adaptatifs dans certains contextes éducatifs » (p. 164). En définitive, les effets du climat de compétition restent une question en suspens. Trois pistes se dégagent pour éclairer les résultats constatés.

L'adéquation entre l'orientation des buts et le climat (l'hypothèse de « l'ajustement climat-motivation » ou *matching hypothesis*). Les travaux de C. S. Dweck (1986 ; Dweck & Legett, 1988) ont suggéré d'étudier les interactions qui existent entre les buts dominants des individus et le climat motivationnel dans lequel ils évoluent. Partant de cette idée générale, certains chercheurs (Linnenbrinks, 2005) ont formulé l'hypothèse selon laquelle l'adéquation entre le climat de la classe et le type de motivation préférentielle de l'élève (impliquant l'ego *contra* maîtrise) pourrait avoir des effets bénéfiques. Autrement dit, les élèves pourraient être davantage motivés quand ils évoluent dans des contextes qui sont en adéquation avec les buts qu'ils poursuivent de manière préférentielle : un élève qui cherche à apprendre et à progresser sera plus motivé dans un climat de maîtrise que dans un climat de compétition. Symétriquement, un élève qui aime la comparaison sociale sera plus motivé dans un contexte de compétition que dans un climat de maîtrise. Même s'ils ne sont pas nombreux, les travaux sur ce thème confirment en partie cette hypothèse (Duda, 2001 ; Harackiewicz & Elliot, 1998 ; Standage, Duda & Ntoumanis, 2003b).

Les climats motivationnels composites (l'hypothèse du rôle « tampon » du climat de maîtrise ou *buffering hypothesis*). Plusieurs travaux (Goudas & Biddle, 1994 ; Linnenbrinks, 2005) ont montré que les climats motivationnels de maîtrise et de compétition n'étaient pas mutuellement exclusifs. Un enseignant peut à la fois mettre l'accent sur la maîtrise des tâches et la comparaison sociale entre élèves. Autrement dit, il est tout à fait possible de percevoir dans la classe une accentuation élevée placée simultanément sur la maîtrise et la compétition. Des études ont donc examiné les effets combinés des deux types de climat. De ces travaux semble ressortir l'idée que la perception d'un climat de compétition n'est pas nuisible à la motivation tant qu'elle est associée à la perception concomitante d'un climat de maîtrise ; ce dernier jouant en quelque sorte un rôle de « tampon », atténuant les effets potentiellement négatifs d'un climat de compétition.

Rôle modulateur du sentiment de compétence dans un climat compétitif. La théorie des buts d'accomplissement a connu ces dix dernières années un renouvellement théorique important dont les développements les plus récents ont conduit à la formulation d'un modèle « 2 x 2 buts » (Elliot & McGregor, 2001). En effet, plusieurs auteurs (Elliot, 1997 ; Middleton & Midgley, 1997 ; Skaalvik, 1997) ont présumé que les buts de performance et de maîtrise pouvaient être scindés en deux modalités distinctes : l'approche et l'évitement. Le but de performance-approche est dirigé vers la démonstration de sa compétence en référence à des critères normatifs (l'individu cherche à démontrer sa supériorité par rapport aux autres ou par rapport à une norme) ; le but de performance-évitement est dirigé vers l'évitement de toute manifestation d'incompétence ; le but de maîtrise-approche consiste à rechercher le progrès, la compréhension profonde de la tâche ; et le but de maîtrise-évitement consiste à éviter de ne pas comprendre, de ne pas apprendre, en bref, de régresser. Selon ce modèle, le sentiment de compétence constitue l'un des antécédents des buts. Ainsi la perception d'un niveau élevé de compétence personnelle induit une attitude positive qui pousse l'individu à rechercher le succès par l'intermédiaire des deux buts d'approche : le but de maîtrise-approche et le but de performance-approche. À l'inverse, la perception d'un faible niveau

de compétence conduit la personne à anticiper la confrontation avec l'échec et à développer des buts d'évitement. Pour A. J. Elliot et J. M. Harackiewicz (1996) les résultats inconsistants concernant le but de performance s'expliquent en partie par l'absence de différenciation des modalités d'approche et d'évitement. Dans le cadre des études sur le climat, il est possible de penser que l'absence de liens entre le climat de compétition et les variables dépendantes mesurées dans les études (Ames & Archer, 1988 ; Digledis *et al.*, 2003 ; Escarti & Gutierrez, 2001) soit due à la non prise en compte du sentiment de compétence des élèves. Autrement formulé, ce type de climat pourrait avoir des effets différents chez les élèves en fonction de leur sentiment de compétence. Ceux qui ont une compétence perçue élevée sont susceptibles de déployer un *pattern* de comportements plutôt positifs (un effort soutenu) tant que cela leur permet de démontrer leur supériorité sur les autres ; à l'inverse ceux qui ont une faible compétence perçue risquent de manifester des comportements d'évitement dont les conséquences sont plutôt négatives (auto-handicap, renoncement). Néanmoins, très peu d'études ont examiné le rôle modulateur joué par la perception de compétence dans un climat compétitif. Les connaissances dans ce domaine n'étant pas très étoffées, il est encore délicat de tirer des conclusions.

Dans cette première partie, nous avons passé en revue les travaux effectués sur le climat motivationnel dans le cadre de la théorie des buts d'accomplissement, qui peut être considérée comme le « berceau » des travaux sur cette thématique. La recherche révèle la supériorité du climat de maîtrise sur celui de compétition. En effet, le climat de maîtrise est systématiquement associé à des réponses motivationnelles, cognitives, affectives, et comportementales « positives » pour les apprentissages (persévérance, motivation intrinsèque, perception d'auto-efficacité, etc.). Ainsi, favoriser une telle orientation motivationnelle contribue à optimiser la réussite scolaire des élèves. À cet égard, le TARGET en tant que cadre de structuration des comportements de l'enseignant s'est révélé d'une grande valeur heuristique. À l'inverse, le profil des réponses associées au climat de compétition est beaucoup moins consistant. L'étude de la littérature révèle des relations complexes entre le climat de compétition et les réponses motivationnelles et comportementales des élèves en classe. Pour éclaircir ces réflexions nous avons tenté une investigation des modulateurs potentiels de ces relations. Il est apparu tout d'abord que les élèves sont plus motivés dans les environnements correspondants à leur définition personnelle de la compétence (Standage, Duda & Ntoumanis, 2003b), c'est-à-dire lorsqu'il y a adéquation entre leurs dispositions motivationnelles (leurs orientations vers la tâche ou vers l'ego) et leurs perceptions du climat motivationnel (climat de maîtrise ou de compétition). Nous avons ensuite souligné que la perception d'un climat de compétition n'était pas nuisible à la motivation tant qu'elle était associée à la perception concomitante d'un climat de maîtrise. Enfin, nous avons évoqué grâce au modèle quadridimensionnel (Elliot & McGregor, 2001) l'effet potentiellement modulateur de la perception de compétence sur le climat de compétition. Néanmoins, ces réflexions demeurent à ce jour des hypothèses théoriques étant donné le peu d'études qui se sont réellement penchées sur ces variables modulatrices.

En définitive, malgré les avancées substantielles de la théorie des buts d'accomplissement, certaines questions demeurent en suspens concernant les effets des climats motivationnels. Par exemple, il subsiste des zones d'ombres à propos des effets du climat de compétition : ses effets sont-ils systématiquement négatifs ? Ses effets peuvent-ils être positifs chez les élèves qui ont une forte compétence perçue dans la matière et/ou une forte propension à rechercher la comparaison sociale (une forte orientation motivationnelle vers l'ego) ? Dans le

cas de la perception de climats motivationnels composites, quelles sont les configurations les plus adaptées et pour quelles variables cibles considérées ? Autant de questions qui représentent des pistes de travail pour des recherches futures.

Le chapitre suivant sera consacré à la présentation des études réalisées dans le cadre de la théorie de l'autodétermination qui constitue un autre regard sur le climat motivationnel.

LA THÉORIE DE L'AUTODÉTERMINATION (TAD) : UN AUTRE REGARD SUR LE CLIMAT MOTIVATIONNEL

Postulats théoriques

La théorie de l'autodétermination (Deci & Ryan, 2002 ; voir Sarrazin et Trouilloud, 2006 pour une présentation de ces travaux en langue française) s'est imposée, ces dernières années, comme un cadre heuristique pour rendre compte des comportements des individus dans différents contextes dont l'école. Selon cette approche, différents types de motivation peuvent être repérés et classés en fonction de leur degré d'autodétermination (Deci & Ryan, 2000). Une motivation est dite « autodéterminée » quand l'activité est réalisée spontanément et par choix. À l'inverse, la motivation est « non autodéterminée » quand l'individu réalise une activité pour répondre à une pression externe ou interne, et qu'il cesse toute implication dès que celle-ci diminue. En contexte scolaire, les élèves font preuve d'une motivation autodéterminée :

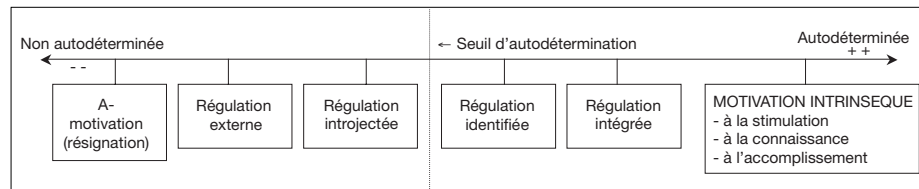
- quand ils s'engagent dans les tâches ou les matières scolaires pour des raisons intrinsèques à celle-ci, que ce soit pour le plaisir qu'ils ressentent en la pratiquant (les auteurs parlent de motivation intrinsèque aux sensations), pour le sentiment de maîtrise qu'ils en retirent (les auteurs parlent de motivation intrinsèque à l'accomplissement), ou pour la satisfaction d'apprendre quelque chose de nouveau (motivation intrinsèque à la connaissance) ;
- quand ils s'engagent dans les activités scolaires parce qu'ils les considèrent comme quelque chose de cohérent avec leurs valeurs et besoins (motivation extrinsèque intégrée) ;
- quand ils s'engagent dans des tâches scolaires qu'ils jugent importantes pour atteindre des buts personnels (motivation extrinsèque identifiée).

Par contraste, les élèves font preuve d'une motivation non-autodéterminée :

- lorsqu'ils s'engagent dans une activité scolaire parce qu'ils s'y sentent obligés, généralement par des récompenses ou des contraintes matérielles et sociales (régulation externe) ;
- mais également à cause de pressions internes comme la culpabilité (régulation introjectée) ;
- ou lorsqu'ils sont résignés et ne perçoivent aucun lien entre leurs actions et leurs résultats (l'a-motivation).

En résumé, la TAD présume l'existence de six types différents de motivation qui sont plus ou moins autodéterminés. La motivation intrinsèque représente le niveau le plus autodéterminé, alors que l'a-motivation est la forme la moins autodéterminée. Les régulations « intégrée », « identifiée », « introjectée », et « externe » se

Figure 1. – **Résumé des différents types de motivation mis en évidence dans la TAD, et ordonnés en fonction de leur degré d'autodétermination**



situent entre ces deux extrêmes (Ryan & Deci, 2000 ; Vallerand, 1997), la régulation identifiée représentant le « seuil » de l'autodétermination (voir la figure 1).

Selon la TAD, les différents types de motivation ont des conséquences cognitives, affectives et comportementales spécifiques. En contexte scolaire, plusieurs études (pour une revue de littérature voir Reeve, 2002 ; Ryan & Deci, 2000 ; Sarrazin & Trouilloud, 2006) ont montré que les formes de motivation les plus autodéterminées étaient associées à des conséquences éducatives positives (attention, plaisir, persistance dans l'apprentissage, performances élevées), alors que les formes les moins autodéterminées avaient des conséquences négatives (abandon précoce, choix de tâches inadaptées à leur niveau, faibles performances). Dès lors, si la motivation autodéterminée favorise l'implication des élèves en classe, la question critique est de savoir comment catalyser cette forme de motivation.

La théorie apporte des réponses à cette question en octroyant aux facteurs sociaux une influence considérable. Selon la TAD, les facteurs sociaux seraient à même de faciliter une motivation autodéterminée (ou non autodéterminée) en nourrissant ou au contraire entravant l'expression de trois besoins psychologiques fondamentaux : les besoins de compétence (le désir d'interagir efficacement avec l'environnement), d'autonomie (le désir d'être à l'origine de son propre comportement), et de proximité sociale (le désir d'être connecté socialement avec des personnes qui nous sont significatives). Tout environnement social qui permettrait la satisfaction de ces trois besoins, catalyserait en retour une motivation autodéterminée. Par contraste, tout environnement social qui entraverait l'expression de ces trois besoins, augmenterait la probabilité d'apparition d'une motivation non-autodéterminée.

Comportements facilitant *contra* comportements entravant la motivation autodéterminée

Depuis une trentaine d'année, des dizaines d'étude expérimentales ont mis en évidence les conditions environnementales susceptibles de catalyser ou amoindrir la motivation intrinsèque des individus. Par exemple, les tâches qui présentent un défi optimal, les *feedbacks* positifs (Fisher, 1978 ; Ryan, 1982) et formatifs (Koestner *et al.*, 1984) qui promeuvent les sentiments d'efficacité et de compétence, les possibilités de choix et les opportunités de décider soi-même (Zuckerman *et al.*, 1978) nourrissent la motivation intrinsèque.

À l'inverse, les menaces de punitions (Deci & Cascio, 1972), les dates butoirs (Amabile, Dejong & Lepper, 1976), les expressions autoritaires comme « il faut » ou « tu dois » (Ryan *et al.*, 1983), les directives (Koestner *et al.*, 1984), la comparaison sociale (Deci *et al.*, 1981) et la pression compétitive (Reeve & Deci, 1996) diminuent la motivation intrinsèque parce que les gens les ressentent comme des

éléments qui contrôlent leur comportement. De même, et contrairement à l'intuition pédagogique de certains enseignants, les récompenses tangibles peuvent amoindrir la motivation intrinsèque (Lepper, Greene & Nisbett, 1973 ; pour une revue de littérature voir Deci, Koestner & Ryan, 1999) chaque fois que l'élève se sent comme un « pion » dont les comportements sont manipulés ou contrôlés par un levier extérieur, et non comme quelqu'un qui est « à l'origine » de ses propres actions (qui choisit lui-même ce qui est bien pour lui).

Même si la motivation intrinsèque constitue une forme de motivation intéressante à développer dans le cadre scolaire, force est de reconnaître que la plupart des tâches scolaires ne sont pas intrinsèquement intéressantes – les élèves n'éprouvent pas spontanément de plaisir dans le travail scolaire (Ryan & Deci, 2000 ; Reeve, 2002). La question fondamentale est donc de savoir comment faire pour faciliter l'engagement autonome et durable des élèves dans les activités scolaires – c'est-à-dire, comment développer une motivation extrinsèque mais autodéterminée (une régulation identifiée ou intégrée). Quelques études expérimentales apportent des réponses à cette question. Trois types de comportements (pour une revue, voir Deci *et al.*, 1991 ; La Guardia & Ryan, 2002 ; Reeve, 1996 & 2002 ; Ryan & Deci, 2000 ; Reeve, Deci & Ryan, 2004 ; Ryan & La Guardia, 1999) semblent catalyser une motivation autodéterminée : les comportements qui offrent des choix aux élèves, ceux qui justifient l'utilité des activités scolaires et ceux qui reconnaissent les affects et ressentis des élèves.

L'apport de choix permettant l'expression des intérêts, préférences et valeurs des individus, nourrit leur motivation autodéterminée (Zuckerman *et al.*, 1978). À l'école, ces choix peuvent se décliner à plusieurs niveaux. Par exemple, en laissant la possibilité à l'élève de choisir la tâche qui représente pour lui un défi. Cette perception de choix est également favorisée lorsque l'enseignant tient compte de l'opinion des élèves (Deci *et al.*, 1981), dans l'organisation de la classe par exemple, et qu'il s'adresse à eux en leur suggérant différents possibles et en les incitant à prendre des initiatives (Reeve *et al.*, 1999). Cependant pour certains chercheurs (Assor *et al.*, 2002 ; Stephanou *et al.*, 2004), les choix offerts aux élèves ne doivent pas se limiter aux aspects organisationnels (choix de situations d'apprentissage ou d'un camarade de travail) pour favoriser une réelle implication dans les apprentissages. Il s'avère indispensable de faciliter les prises d'initiatives qui relèvent du domaine cognitif, en aidant par exemple les élèves à construire leur propre raisonnement. En d'autres termes, si proposer des choix organisationnels aux élèves est nécessaire, les accompagner dans l'élaboration de raisonnements personnels et complexes est fondamental pour qu'ils deviennent acteurs de leurs apprentissages.

D'autre part, les justifications sur l'importance de faire la tâche favorisent aussi la motivation autodéterminée (Deci *et al.*, 1994 ; Reeve *et al.*, 2002 ; Assor *et al.*, 2002). En soulignant la congruence entre les activités scolaires et les ressources des élèves, elles participent du processus d'internalisation. Néanmoins, le contenu de ces justifications peut induire une variété de comportements plus ou moins positifs. M. Vansteenkiste et ses collaborateurs (Vansteenkiste, Simons, Lens, Soemons, *et al.*, 2004 ; Vansteenkiste, Simons, Lens, Sheldon, *et al.*, 2004 ; Vansteenkiste *et al.*, 2005) ont montré que lorsque les justificatifs mettaient en avant des raisons intrinsèques de faire la tâche (l'épanouissement personnel, la santé, apporter sa contribution à la classe), ils généraient un meilleur apprentissage, plus de persévérance et des performances plus élevées que des justifications de nature plus extrinsèque (la récompense financière, la valorisation de soi, l'apparence physique). Ces auteurs expliquent que réaliser une tâche pour des raisons extrinsèques, focalise l'élève sur des indicateurs de réussite externes ce

qui a tendance à le détourner du processus d'apprentissage. À l'inverse, lorsque la tâche est effectuée pour des raisons intrinsèques, l'apprentissage est vécu comme un moyen de progresser et de satisfaire les besoins psychologiques des élèves ce qui facilite les acquisitions.

Enfin, reconnaître que les élèves puissent avoir des sentiments négatifs (désintérêt, ennui) envers des activités scolaires (Deci *et al.*, 1994 ; Koestner *et al.*, 1984), faire preuve d'empathie à leur égard (Assor, Kaplan & Roth, 2002) et être chaleureux (Ryan & Grolnick, 1986) sont des comportements qui facilitent la motivation autodéterminée. De plus, la recherche semble montrer que les effets de chacun de ces événements s'additionnent pour conduire à davantage de motivation autodéterminée. Dans une étude destinée à tester les effets singuliers et combinés de trois comportements soutenant l'autonomie (une justification sur l'importance de faire la tâche, la reconnaissance des sentiments à propos de l'activité et l'utilisation d'un langage qui véhicule un choix) à l'égard d'une tâche ennuyeuse sur ordinateur, E. L. Deci, *et al.* (1994). Les résultats ont montré que combinés, ces facteurs facilitaient plus la motivation autodéterminée (attestée par le temps consacré à la tâche durant une séquence où le participant était laissé seul) que chacun d'eux pris isolément.

Style « soutenant l'autonomie » *contra* style « contrôlant » de l'enseignant : conséquences sur les comportements scolaires.

Ces premiers travaux sur les comportements susceptibles de nourrir ou d'entraver la motivation autodéterminée, se sont traduits par la mise en évidence de deux styles de supervision : un style « *soutenant l'autonomie* » et un style « *contrôlant* ». Parce qu'il nourrit les besoins des élèves le premier est susceptible de faciliter une motivation autodéterminée, contrairement au second qui tend plutôt à entraver leur expression en conduisant à une motivation non-autodéterminée. J. Reeve (2002) a caractérisé les différences fondamentales qui existent entre ces deux types de styles d'enseignement. Par essence, un enseignant qui soutient l'autonomie est *plus proche* (il respecte les élèves, passe plus de temps à les écouter, à essayer de les comprendre), *plus positif* (il encourage les élèves, et valorise les progrès réalisés), *plus flexible* (il laisse aux élèves des opportunités de faire des choix, leur donne du temps pour résoudre les problèmes par eux-mêmes), et *plus explicatif* (il explique aux élèves l'utilité ou l'importance de faire telle tâche ou telle activité). À l'inverse, un enseignant « contrôlant » a tendance à *prendre tout en charge* (il impose ce qu'il y a à apprendre et comment le faire, utilise un langage directif), à *être plus pressé* (il donne rapidement la solution sans vraiment laisser le temps à l'élève de chercher par lui-même), à *être plus négatif* (il critique davantage), et à *motiver par la pression* (il a recours aux récompenses/punitions, aux exhortations et à la compétition entre élèves).

Plusieurs études de nature corrélationnelle ont fait ressortir les bienfaits du soutien de l'autonomie sur une variété de résultats éducatifs (pour une revue, voir Reeve *et al.*, 2002). Comparé à un climat contrôlant, les élèves dont le professeur soutient l'autonomie réussissent mieux à l'école (Boggiano *et al.*, 1993 ; Flink, Boggiano & Barrett, 1990), ont une compétence perçue plus élevée (Deci *et al.*, 1981 ; Trouilloud *et al.*, 2006), ressentent des émotions plus positives (Patrick, Skinner & Connell, 1993), ont une plus haute estime de soi (Deci *et al.*, 1981), font preuve d'une meilleure compréhension (Boggiano *et al.*, 1993), sont plus actifs dans le traitement des informations (Grolnick & Ryan, 1987), sont plus persévérants dans les études (Vallerand *et al.*, 1997), et sont plus créatifs (Koestner *et al.*, 1984).

Créer un style soutenant l'autonomie des élèves

Malgré les bienfaits d'un style motivationnel soutenant l'autonomie révélés dans la littérature, beaucoup d'enseignants rapportent pourtant que ce concept de *soutien de l'autonomie* leur est peu familier voire étranger (Boggiano *et al.*, 1987). La plupart d'entre eux ont plus spontanément recours à des stratégies contrôlantes comme les récompenses ou les menaces de punitions pour faire travailler les élèves (Boggiano *et al.*, 1987 ; Sarrazin *et al.*, 2006). La question centrale qui se pose alors est de savoir s'il est possible d'aider les enseignants à transformer ce style principalement contrôlant en un style davantage tourné vers le soutien de l'autonomie des élèves. Bien qu'essentielle, très peu d'études se sont attachées à répondre à cette question. Les travaux de J. Reeve (1998 & 2004) sont parmi les seuls à avoir examiné, dans le cadre de la TAD, les effets de programmes de formation sur le style motivationnel d'enseignants. Dans une étude quasi-expérimentale, réalisée auprès de vingt enseignants expérimentés, l'objectif poursuivi par J. Reeve et ses collaborateurs (2004) était de mesurer si le fait de proposer à des enseignants une formation sur les intérêts du soutien de l'autonomie avait une influence sur l'utilisation par ces enseignants de comportements soutenant l'autonomie auprès de leurs élèves et sur l'engagement de leurs élèves en classe. La formation s'organisait autour de la présentation aux enseignants des principes de base de la TAD, et notamment les différents types de motivation des élèves, les différents styles motivationnels des enseignants (soutenant l'autonomie *contra* contrôlant l'autonomie), ainsi que des éléments empiriques montrant les bénéfiques pour les élèves d'un style d'enseignement orienté vers le soutien de l'autonomie (par rapport à un style contrôlant). Les résultats des observations effectuées en classe, une semaine après la formation, révèlent que les enseignants du groupe expérimental ont utilisé davantage de comportements orientés vers le soutien de l'autonomie (par rapport aux enseignants du groupe contrôle). De plus, l'utilisation de ce type de comportements par les enseignants a eu un effet positif sur le niveau d'engagement (cognitif et comportemental) de leurs élèves.

Vers une conception multidimensionnelle du climat motivationnel : le soutien des besoins

Dans la plupart des travaux présentés précédemment, le concept de « soutien de l'autonomie » était entendu comme un terme générique susceptible d'affecter globalement chacun des trois besoins fondamentaux, et conduire en retour à une motivation autodéterminée. L'un des enjeux des travaux actuels est d'affiner le grain d'analyse en tentant de déterminer quels comportements de l'enseignant est de nature à satisfaire/entraver quel besoin spécifique de l'élève (Reeve, Deci & Ryan, 2004) et s'il existe un ordre de priorité dans la satisfaction des besoins. Concernant ce second point, quelques études corrélationnelles récentes réalisées en EPS (Ntoumanis, 2001 ; Standage, Duda & Ntoumanis, 2003a ; Standage, Duda & Ntoumanis, 2006) font par exemple ressortir un primat du besoin de compétence. Dans ces trois études, la satisfaction du besoin de compétence était le plus reliée à la motivation autodéterminée. La satisfaction du besoin d'autonomie et celle du besoin de proximité sociale apportaient une contribution supplémentaire à ce type de motivation, mais avec un poids plus modeste. Cette question du poids et de la complémentarité de chaque besoin en fonction, par exemple, des participants devra assurément être appréhendée dans les recherches futures.

Par ailleurs, le traitement du premier point implique le passage d'une approche unidimensionnelle considérant le soutien de l'autonomie comme une structure générique sous-tendant la perception du climat motivationnel, à une approche multidimensionnelle se donnant comme objet d'étude les relations entre les catégories de comportements de l'enseignant caractérisées par leur impact différencié sur chacun des trois besoins psychologiques des élèves. Quelques chercheurs ont récemment adopté une telle approche (Reeve, Deci & Ryan, 2004 ; Skinner & Belmont, 1993 ; Skinner & Edge, 2002), en dénommant « soutien de l'autonomie », « structure » et « implication », les facteurs sociaux susceptibles de nourrir respectivement les besoins d'autonomie, de compétence et de proximité sociale.

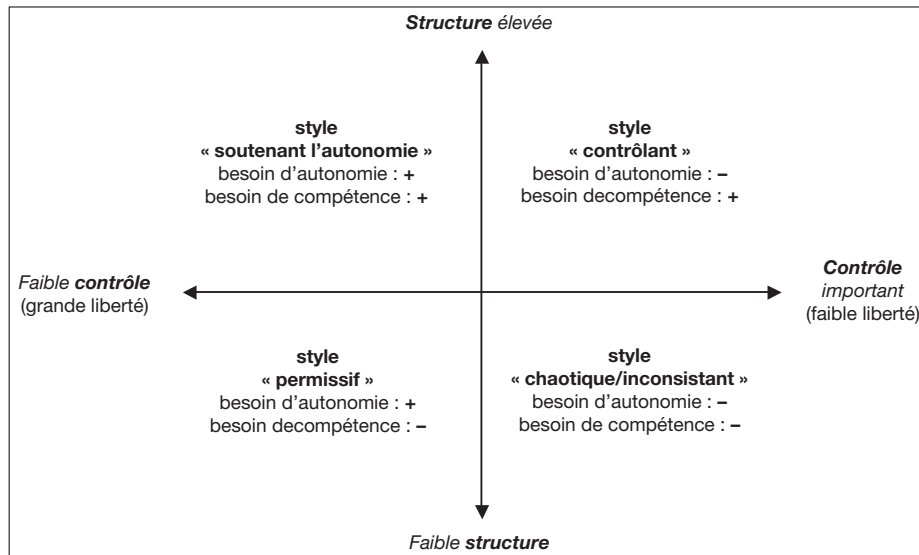
Le soutien de l'autonomie : une marge de liberté dans un environnement structuré

Le concept de soutien de l'autonomie peut paraître confus et être associé à tort à une ambiance permissive de type *laissez faire*. Pour bien comprendre la différence qui existe entre un enseignement soutenant l'autonomie et un environnement permissif, il peut être utile à l'instar de J. Reeve, E. L. Deci et R. M. Ryan (2004), de dissocier les styles d'enseignement en fonction du degré de liberté dont disposent les élèves et de la structuration de la classe. Comme le montre la figure 2, la différence majeure entre ces deux types d'environnement réside dans la notion de structure : élevée dans un climat soutenant l'autonomie et faible dans un environnement permissif. Le style permissif se caractérise par une grande liberté d'action laissée aux élèves. La classe n'est pas structurée, c'est-à-dire qu'il n'y a pas de direction donnée à l'apprentissage des élèves par l'enseignant, ni de limites opposées à leurs actions. Sans limites, les élèves peuvent donc faire ce qu'ils veulent. *A contrario*, un climat soutenant l'autonomie est structuré. Si l'enseignant laisse des choix aux élèves et les encourage à prendre des initiatives, cela s'organise dans un cadre dont il a défini les limites. Plus précisément, un enseignement structuré implique une organisation des activités en classe, la transmission de contenus d'enseignement, la construction de tâches représentant un défi pour chaque élève, la transmission de *feedbacks* adaptés aux problèmes rencontrés par les élèves (pour une revue, voir Reeve, Deci & Ryan, 2004). Le soutien de l'autonomie et la structure ne sont donc pas des dimensions opposées, mais sont conceptualisés comme des dimensions orthogonales du style motivationnel de l'enseignant, chacune étant susceptible de satisfaire ou d'entraver les besoins psychologiques d'autonomie et de compétence (voir la figure 2 ; Connell & Wellborn, 1991 ; Reeve, Deci & Ryan, 2004 ; Skinner & Belmont, 1993).

La figure 2 met également en évidence que les environnements soutenant l'autonomie et contrôlant sont tous deux structurés. La différence essentielle entre les deux tient à la dimension de liberté laissée à l'élève. Contrairement au soutien de l'autonomie, le style contrôlant ne laisse ni choix, ni prise d'initiatives et ne propose que peu, voire aucune marge de liberté aux élèves à l'intérieur de la structure d'enseignement. Dans un environnement contrôlant, l'élève est perçu comme un applicateur des décisions prises par l'enseignant.

Enfin, les environnements scolaires coercitifs faiblement structurés ne constituent pas réellement un style motivationnel référant pour les enseignants (Reeve, Deci & Ryan, 2004). De même qu'une ambiance permissive est une déformation du style soutenant l'autonomie, les environnements scolaires contraignants et peu structurés sont une déformation du style contrôlant. Ce type de climat peut être considéré comme « chaotique », car les contraintes, les directives, et plus globalement les attentes ou exigences de l'enseignant sont floues et les élèves ne savent pas comment y répondre.

Figure 2. – **Un cadre 2 x 2 distinguant les dimensions environnementales relatives à la « structure » et au « contrôle », et les besoins qu’elles satisfont (inspiré de Reeve, Deci & Ryan, 2004)**



L'implication du professeur

Elle constitue le troisième élément d'un climat facilitant l'engagement spontané des élèves. En effet, une des raisons principales de faire des comportements qui ne sont pas intéressants en eux-mêmes (qui ne procurent pas un plaisir spontané aux élèves), c'est de recueillir l'approbation d'autrui significatifs avec lesquels nous nous sentons connectés, que ce soit la famille, un groupe de pairs, ou une société (Deci & Ryan, 2000). Appliqué à la situation éducative, ce principe signifie que pour faciliter l'engagement spontané des élèves dans les tâches proposées, il est fondamental que ces derniers se sentent respectés par, et « connectés » avec leur enseignant (Ryan & Deci, 2000).

Plusieurs auteurs désignent par « implication » la dimension du climat motivationnel qui nourrit le besoin de proximité sociale des élèves (Reeve *et al.*, 2002 ; Reeve, Deci & Ryan, 2004 ; Skinner & Edge, 2002). Pour ces auteurs l'implication renvoie à la relation qui s'établit entre un enseignant et un élève, c'est-à-dire aussi bien à la quantité (temps, énergie) qu'à la qualité (marque d'affection, sourire, note d'humour) des ressources psychologiques allouées par l'enseignant dans cette relation. Il peut paraître *a priori* difficile d'envisager qu'un enseignant puisse soutenir l'autonomie d'un élève (le guider dans son apprentissage par un questionnement approprié, lui livrer des *feedbacks* techniques, lui proposer des tâches adaptées à son niveau) sans s'impliquer. En effet, si l'implication d'un enseignant auprès d'un élève dont il soutient l'autonomie ne fait aucun doute, en revanche la *qualité* de cette implication peut être très diverse. À l'instar des deux autres besoins, le besoin de proximité sociale se distribue sur un axe bipolaire opposant l'implication à l'hostilité (Skinner & Edge, 2002). Un enseignant « hostile » se caractérise par exemple par des comportements froids et distants, par un ton de voix sévère et monocorde, par un refus systématique d'entendre et de prendre en compte le point de vue de l'élève. Par opposition, un enseignant « impliqué » est

chaleureux et disponible pour ses élèves, c'est-à-dire qu'il essaie d'instaurer une relation d'empathie dans laquelle il investit du temps et de l'énergie.

Pour résumer les idées fortes qui découlent de la conception multidimensionnelle, rappelons tout d'abord que selon la TAD, les contextes sociaux qui soutiennent les sentiments de compétence, d'autonomie et de proximité sociale constituent les ingrédients du maintien d'une motivation intrinsèque et du développement d'une motivation autodéterminée. À l'école, pour faciliter l'engagement spontané des élèves dans les tâches d'apprentissage, il convient de réunir les conditions qui permettent la satisfaction de ces trois besoins fondamentaux humains. Les facteurs sociaux susceptibles d'affecter ces derniers peuvent être placés sur trois axes bipolaires aux extrémités desquels chaque besoin est satisfait ou entravé. Ainsi le soutien de l'autonomie (offrir des choix et des options, encourager à la prise d'initiatives, justifier l'utilité et la valeur des tâches) s'oppose à la coercition (diriger par le biais de commandes directives, faire pression sur les élèves) ; la structure (proposer des tâches de défi, donner des *feedback* centrés sur l'apprentissage) s'oppose au chaos (proposer des objectifs flous et peu d'organisation pédagogique), et l'implication enfin (investir du temps et de l'énergie auprès des élèves, manifester des marques d'affection) s'oppose à l'hostilité (être froid et distant, ne pas prendre en compte l'élève). Le tableau II résume les différents éléments du climat susceptible de nourrir chacun des besoins et les stratégies qui devraient être valorisées par l'enseignant pour les mettre en place.

LIMITES THÉORIQUES ET MÉTHODOLOGIQUES À L'APPLICATION EN CLASSE DES RECHERCHES SUR LA MOTIVATION

Bien qu'issus d'éclairages théoriques différents, les travaux qui viennent d'être présentés se rejoignent au niveau de l'intérêt qu'ils portent aux pratiques enseignantes qui créent le climat motivationnel. Les tableaux I et II rassemblent les focalisations et les stratégies issues de la TBA et de la TAD, susceptibles de créer un climat motivationnel propice aux acquisitions. L'objectif de cette dernière partie est d'attirer l'attention sur les nécessaires précautions à prendre et les réserves à avoir concernant l'application en classe des « principes » dégagés de ces travaux mais également sur les difficultés de leur utilisation par les enseignants.

Est-ce que les conclusions des recherches expérimentales peuvent être appliquées en classe ?

La plupart des travaux dans le domaine de la motivation ont été conduits en utilisant la méthode expérimentale. La question est de savoir si le contexte naturel d'enseignement et celui des études expérimentales sont identiques ou tout au moins comparables. Pour ne donner qu'un exemple (mais il n'y a qu'à feuilleter les comptes rendus de recherche pour en trouver d'autres), dans l'étude de J. Reeve *et al.* (2006), il était demandé à des participants de jouer le rôle d'un enseignant, et à d'autres celui d'élève ; l'activité enseignée était la réalisation d'un puzzle, c'est-à-dire une activité de nature cognitive très différente et plus plaisante que la plupart des activités proposées dans le cadre scolaire ; l'interaction se déroulait sur une période de dix minutes seulement et elle confrontait l'« enseignant » à un seul « élève » et non pas à une classe. Il est donc possible d'émettre quelques réserves quant à la validité écologique des résultats trouvés dans ce type d'études expérimentales.

Tableau II. – Résumé des éléments d'un climat soutenant les besoins d'autonomie, de compétence et de proximité sociale et des stratégies susceptibles de les nourrir
(d'après Skinner & Edge, 2002 ; Reeve & Jang, 2006)

Principes éducatifs	Focalisation	Stratégies
Soutenir l'autonomie (besoin d'autonomie)	Impliquer activement les élèves dans le processus d'apprentissage	<ul style="list-style-type: none"> • Expliquer l'utilité des tâches d'apprentissage • Inciter les élèves à prendre des initiatives en offrant des choix et des options • Suggérer différents possibles, différentes pistes de travail à explorer • Éviter l'utilisation des commandes et directives telles que « il faut... », « tu dois... » • Prendre en compte les préférences des élèves • Laisser un temps de travail en autonomie suffisant en fonction des besoins des élèves et éviter les dates butoirs et les pressions temporelles • Encourager les élèves en soutenant leur engagement dans l'apprentissage • Féliciter les élèves pour leurs progrès, efforts, réussite personnelle et non parce qu'ils se conforment à ce que l'enseignant attend d'eux
Structurer l'enseignement (besoin de compétence)	Faire en sorte que l'élève apprenne quelque chose, quelque soit son niveau et ses possibilités	<ul style="list-style-type: none"> • Proposer à tous les élèves des tâches qui représentent un défi • Fixer des objectifs concrets à atteindre • Donner des feedbacks adaptés au niveau de l'élève • Éviter de donner trop vite les solutions, avant que les élèves n'aient eu complètement l'opportunité de les découvrir par eux-mêmes
S'impliquer auprès des élèves (besoin de proximité sociale)	Instaurer une relation pédagogique basée sur l'empathie et la proximité avec les élèves	<ul style="list-style-type: none"> • Écouter attentivement les questions et remarques des élèves • Laisser la possibilité aux élèves d'exprimer librement leurs émotions, ressenties et affects • Dire qu'il est normal de connaître de difficultés quand on apprend, ou d'être mal à l'aise • Investir du temps, et de l'énergie dans l'interaction avec l'élève. • Manifester des marques d'affection et de respect • Éviter les critiques susceptibles de « piquer » l'ego

De plus, les consignes ou stratégies susceptibles de faciliter une motivation particulière sont généralement beaucoup moins claires en classe que dans les situations expérimentales typiques. Par exemple, les manipulations expérimentales destinées à susciter un but d'accomplissement particulier ont pour objectif d'amener les participants à poursuivre un but unique afin d'en tester les effets (faire mieux que les autres élèves *contra* maîtriser la tâche). En situation normale d'enseignement, il y a généralement des messages multiples qui peuvent mettre l'accent sur des buts différents. Par exemple, si l'enseignant propose des tâches de défi à ses élèves et qu'il encourage les efforts et les progrès *et si dans le même temps*, il les évalue sur la base d'un standard social (avec un classement du meilleur au plus faible), on peut s'interroger sur le message qui sera perçu. Selon C. Ames (1992a), il est nécessaire que les structures du climat travaillent de concert pour que les messages de l'enseignant soient orientés vers le même but motivationnel. Dans le cas contraire, cela pourrait générer de la confusion chez les élèves. La contribution positive d'une structure (proposer des tâches de défi dans lesquelles les élèves ont des choix à faire) sur la motivation des élèves pourrait alors être compromise par la mise en place incompatible d'une autre (des pratiques d'évaluation qui favorisent la comparaison sociale). Si les structures motivationnelles sont interdépendantes, se pose alors la question de savoir si elles opèrent de manière additive ou multiplicative ? Si elles sont additives, cela signifie qu'elles se complètent et que l'incongruité de l'une peut être compensée par l'effet positif des autres. Toutefois, si elles sont liées de manière multiplicative, elles ne peuvent pas se compenser les unes par les autres.

Cette limite est également valable pour les structures identifiées dans le cadre de la théorie de l'autodétermination. Savoir si l'une des dimensions a plus de poids que d'autres est une question à laquelle la recherche n'a pour l'instant pas répondu. Par exemple, comment l'élève perçoit-il le climat motivationnel établi par un enseignant qui est chaleureux (comportement d'implication), qui laisse aux élèves quelques opportunités de faire des choix (comportement soutenant l'autonomie), mais qui transmet ses consignes et donne des *feedbacks* sur un ton directif (comportement contrôlant) ?

Cette réflexion aboutit alors à se demander si le climat ne serait pas un construit général qui dépasserait la somme de ses parties (structures). Cette question constitue actuellement une zone d'ombre, peu investie par la recherche (Deci *et al.*, 1994 ; Assor *et al.*, 2002), qui pourrait éventuellement restreindre l'application des « principes » théoriques issus de la recherche aux situations naturelles d'enseignement caractérisées par leur complexité.

Est-ce que les résultats des recherches corrélationnelles transversales sont fiables ?

À côté des recherches expérimentales en laboratoire, les études en classe utilisent massivement le questionnaire. Il s'agit pour la plupart de croiser les réponses des participants à des questionnaires mesurant leur motivation (orientation à l'accomplissement ; autodétermination), et/ou leur perception du climat (variables du TARGET ; style soutenant l'autonomie), et/ou différentes attitudes scolaires (stratégies d'études, auto-handicap, temps consacré au travail à la maison), le plus souvent lors d'une passation unique (plan transversal ou synchronique). Si ce type de plan a des avantages (en particulier la facilité de passation), il comporte également plusieurs limites dont il convient d'être conscient quand on interprète les résultats.

Des problèmes de variance partagée et de biais de positivité/négativité

Mesurer les associations entre des variables de nature différente (le climat de l'enseignant, la motivation de l'élève et différents « résultats » éducatifs), mais en s'appuyant sur une source unique – le point de vue de l'élève – peut produire une surestimation des corrélations entre les différents construits par un *effet de halo* (Thorndike, 1920). Il s'agit d'un phénomène de « contamination » des réponses à l'ensemble du questionnaire (variance partagée) en fonction par exemple de l'attrait ou au contraire de l'aversion qu'éprouve un élève pour la matière ou l'enseignant. On parle de *biais de positivité* quand une caractéristique de l'enseignant ou de la matière, jugée positive par l'élève, a tendance à influencer positivement ses réponses à l'ensemble du questionnaire. Par exemple, appréciant beaucoup le sport, un élève va rapporter une forte motivation pour le cours d'éducation physique, envisager des conséquences positives en terme d'investissement dans cette matière et finalement, considérant son enseignant comme plus sympathique, sera enclin à percevoir ses comportements comme soutenant l'autonomie plutôt que contrôlant. À l'inverse, cet effet de halo peut se traduire par un *biais de négativité* lorsque la perception négative d'un élément particulier affecte l'ensemble des réponses de l'élève au questionnaire. Par exemple, n'aimant pas son professeur de mathématiques, un élève va généraliser ses perceptions négatives à tout ce qui se rapporte à cette discipline. Il sera ainsi enclin à considérer les comportements de son enseignant comme contrôlant, à rapporter une faible motivation et un investissement réduit au strict minimum dans cette matière. Ces risques sont d'autant plus grands que les élèves sont jeunes et qu'ils ont du mal à distinguer des concepts abstraits et parfois proches. En définitive, il est possible que les associations révélées entre les variables dans ce type d'étude, soient en réalité moins importantes.

La question du flux causal entre le climat et la motivation ?

Pour beaucoup de chercheurs, le climat motivationnel est un déterminant de la motivation des élèves. Autrement dit, les recherches suggèrent de manière typique l'existence d'une influence unidirectionnelle qui irait de l'enseignant vers les élèves. Ainsi, une corrélation élevée entre la perception par l'élève, d'un climat soutenant l'autonomie et sa motivation autodéterminée, est interprétée comme la preuve d'une telle influence. Pourtant, la nature corrélationnelle des données n'autorise pas à faire des inférences sur les relations de causalité entre les variables (Pelletier, Boivin & Allain, 2000). Il est d'ailleurs tout à fait possible d'envisager que cette corrélation soit la conséquence d'un flux inverse : l'enseignant est davantage enclin à soutenir l'autonomie des élèves qu'il juge autonomes, et à être plus contrôlant vis-à-vis de ceux qu'il juge peu autonomes (Pelletier, Séguin-Lévesque & Legault, 2002 ; Sarrazin *et al.*, 2006). En adaptant le climat motivationnel au type de motivation des élèves, l'enseignant peut aussi renforcer leur motivation vis-à-vis de l'école (Sarrazin *et al.*, 2006). En définitive, il est difficile d'identifier le flux causal des influences motivationnelles en classe. Les études de terrain qui révèlent une association entre les pratiques enseignantes et la motivation des élèves tendent à grossir une explication possible, alors qu'elles ne permettent pas de dépeindre précisément la direction de cette influence causale.

Est-ce que les réponses des élèves aux enquêtes en classe sont un reflet précis de ce qui se passe en classe ?

Une autre limite réside dans la *subjectivité* des dimensions constitutives du climat motivationnel. L'utilisation d'un questionnaire implique une définition *a priori* des construits et, qui plus est, mesurés par un nombre limité d'items.

Par conséquent, il est probable que les questions proposées par le chercheur au travers du questionnaire ne capturent pas toute l'étendue des informations potentiellement contenues dans l'environnement, associées à un type de climat. De plus, de par sa nature le questionnaire présuppose une interprétation spécifique des différentes pratiques éducatives des professeurs. Les items sont ciblés sur certains comportements de l'enseignant censés refléter un type de climat particulier (maîtrise *contra* compétition, soutien de l'autonomie *contra* contrôle). Selon cette logique, la signification donnée à ces comportements du professeur, est partagée de manière unanime par tous les élèves. Or, les processus cognitifs par le biais desquels l'élève interprète les comportements de l'enseignant sont largement influencés par des facteurs développementaux et culturels (Pintrich, 2003) qui, par essence, rendent ces interprétations singulières.

Par exemple, l'étude de E. A. Skinner et M. J. Belmont (1993) a demandé à la fois aux élèves et aux enseignants d'évaluer le climat de la classe au niveau de la structure, du soutien de l'autonomie et de l'implication. Il s'agissait d'élèves de quatorze classes de CE2 à CM2. Les résultats ont souligné l'absence totale de lien entre les évaluations des enseignants et celles des élèves au niveau des dimensions « structure » et « soutien de l'autonomie » et un lien modeste ($\beta = .25$) pour l'« implication ». Par ailleurs, les comportements d'implication du professeur étaient associés aux réponses des élèves relatives à un climat structuré et soutenant leur autonomie. De cette étude se dégage donc l'idée que les perceptions des élèves relatives au climat de leur enseignant ne sont pas un reflet forcément précis de la réalité. En effet, toutes les réponses de l'élève sont « colorées » par l'implication (chaleur, disponibilité et aide apportées) que ce dernier manifeste à leur égard. Quand celle-ci est élevée, les élèves rapportent des scores élevés à toutes les dimensions du climat, et quand celle-ci est faible, les élèves rapportent des scores plus faibles à ces dimensions. En conclusion, le recueil de données par questionnaire pourrait entraîner une perte d'informations et ne pas être le reflet précis de ce qui se passe dans la classe. Cette limite constitue une autre précaution à prendre pour le praticien qui souhaite s'inspirer des résultats de recherche pour améliorer le climat motivationnel de sa classe.

Est-ce que les enseignants peuvent véritablement utiliser toutes les stratégies d'intervention tirées des recherches ?

Pour contextualiser notre propos, nous traiterons plus précisément des difficultés liées à l'utilisation de deux stratégies d'intervention que nous estimons particulièrement importantes pour favoriser la motivation à travailler des élèves : proposer aux élèves des tâches nouvelles, intéressantes et variées et les encourager à rechercher le défi.

Dans le contexte scolaire actuel, l'utilisation de ces stratégies semble difficile. Les pressions subies par l'enseignant (traiter l'intégralité du programme scolaire, classes chargées, gestion disciplinaire des élèves, etc.) tendent à réduire sa marge de manœuvre en terme d'innovation et de variété pédagogique (Pelletier *et al.*, 2002). Comme le rappelle J. Brophy (1999), les contenus d'enseignement sont sélectionnés essentiellement au regard de ce que la société estime nécessaire pour la formation et l'éducation de tout élève, indépendamment de toute considération sur le caractère plaisant ou non de ces contenus. Par ailleurs, même si l'enseignant disposait d'une plus grande marge de liberté, on peut s'interroger sur la faisabilité de ce principe. En effet, la classe est composée d'au moins vingt-cinq élèves aux besoins et aspirations très variés, et il ne semble pas possible dans ces conditions de proposer des tâches ou activités qui puissent satisfaire l'intérêt intrinsèque de chacun.

D'autre part, s'il peut être opportun de proposer des tâches de défi aux élèves pour éveiller leur motivation intrinsèque ou susciter un but de maîtrise, il est possible qu'en situation normale de classe, ce type de tâche soit plutôt à même d'activer une crainte d'échouer et de se ridiculiser. Comme nous l'avons souligné plus haut, la peur de l'échec est fortement reliée à un but de performance-évitement dont les conséquences sont plutôt néfastes. Dans beaucoup de classes, il est peut-être pire d'essayer un exercice qui comporte un risque d'échec que de ne pas essayer du tout ! Ainsi, même quand les enseignants essaient de proposer des tâches de défi pour les élèves, ces tentatives peuvent ne pas être fécondes.

Est-ce que les enseignants comprennent et valorisent les recommandations faites par les recherches sur la motivation ?

Au-delà de la difficulté pour les enseignants d'accéder au compte rendu des recherches sur la motivation, il n'est pas du tout certain que ces derniers utilisent les recommandations qui en découlent, tout simplement parce qu'ils n'en acceptent pas les principes ou parce qu'ils estiment qu'ils ne fonctionnent pas ! Par exemple, beaucoup d'enseignants croient au puissant levier motivationnel que peut constituer la compétition entre élève (Thorildsen & Nicholls, 1998), alors que les travaux sur la TBA ont montré qu'un accent trop prononcé sur la compétition pouvait produire un sentiment de crainte chez les élèves susceptible d'activer un but performance-évitement.

Par ailleurs, beaucoup d'enseignants ne souscrivent pas au postulat selon lequel les êtres humains ont une inclination naturelle à comprendre et à maîtriser les objets nouveaux, porteurs de défis ou utiles pour fonctionner. Ils considèrent plutôt les élèves comme naturellement non motivés pour les apprentissages scolaires, et peu à même de travailler sans un contrôle externe. Il existe une croyance forte selon laquelle plus la « carotte » ou le « bâton » est gros et plus la motivation sera grande (Boggiano *et al.*, 1987 ; Reeve, 2002). Cette croyance est aussi renforcée par des représentations sociales sur ce qu'est le « bon » enseignant. Dans l'imaginaire collectif, le « bon » enseignant est celui qui fait preuve d'autorité et de sévérité, celui qui adopte une attitude plus démocratique est jugé comme moins compétent (Reeve, 2002).

Enfin, la mise en place réussie d'une ou deux recommandations pratiques préconisées par les recherches sur la motivation nécessiterait des changements importants dans les pratiques habituelles des enseignants. De telles évolutions sont très consommatrices de temps que tous ne sont pas forcément prêts à concéder. Ainsi, simplement dire aux enseignants ce qu'ils devraient faire pour nourrir la motivation de leurs élèves n'est clairement pas suffisant pour que cela arrive. Cela doit être accompagné d'une réflexion plus large sur la manière de modifier les structures éducatives afin d'encourager et soutenir les initiatives innovantes des enseignants.

CONCLUSION ET PERSPECTIVES

« L'augmentation significative de la littérature dans le domaine, montre que l'enseignant joue un rôle actif dans la construction du climat motivationnel, et par conséquent sur la qualité de la motivation des élèves » (Treasure & Roberts, 1995, p. 480).

Dans cette note de synthèse, nous avons voulu dresser un bilan des connaissances sur le thème du climat motivationnel en éducation. Les idées clés peuvent être résumées en trois points :

- l'étude du climat motivationnel en contexte éducatif a démarré avec la théorie de la motivation à l'accomplissement. L'idée essentielle à retenir est que le climat de maîtrise produit des conséquences positives sur les apprentissages scolaires et que le climat de compétition génère des effets controversés que la recherche n'a pour l'instant pas réussi à éclaircir totalement ;
- la théorie de l'autodétermination offre un autre regard sur le climat motivationnel en postulant la nécessité de nourrir trois besoins psychologiques fondamentaux (autonomie, compétence et proximité sociale) des élèves pour favoriser leur motivation et leur engagement scolaire. Si actuellement, les recherches empiriques s'appuient sur une conception bidimensionnelle du climat – le soutien de l'autonomie *contra* le contrôle – des avancées théoriques se sont ouvertes à une vision multidimensionnelle (le soutien de l'autonomie *contra* la coercition, la structure *contra* le chaos, l'implication *contra* l'hostilité) qui offre des perspectives de recherche prometteuses ;
- l'application à la classe des principes éducatifs issus de ces travaux, requiert néanmoins de faire preuve d'une certaine prudence. La complexité du contexte naturel d'enseignement par rapport aux situations étudiées en laboratoire, les contraintes qui s'imposent à l'enseignant, qui réduisent sa liberté pédagogique et qui le poussent à être contrôlant envers ses élèves, ainsi que les limites méthodologiques des études par questionnaire, constituent les principales réserves à avoir quant à la validité écologique de ces recherches.

Pour dépasser ces limites, nous terminerons en proposant quelques directions à valoriser dans des études futures.

Conduire des études avec observation des comportements. Ce type de méthodologie – coûteuse il est vrai – est très peu utilisé par les chercheurs. Pourtant, l'observation des comportements de l'enseignant permet de rendre compte de la réalité de la classe avec plus d'authenticité que le questionnaire et en même temps d'éviter les biais de variance partagée inhérents à l'utilisation exclusive de cette méthodologie.

Mener des études qualitatives. Elles sont à concevoir en complémentarité des études quantitatives (Stefanou *et al.*, 2004 ; Urdan, 2004). Poser des questions ouvertes aux élèves, sous forme d'entretien par exemple, permettrait d'affiner le « grain d'analyse » des échelles de climat et d'identifier les messages de l'enseignant qui « marquent » les élèves.

Encourager les études quasi expérimentales qui visent à modifier le climat de la classe. Ces dernières années la recherche en psychologie de l'éducation a réussi à se doter de cadres de structuration des comportements de l'enseignant autour desquels la mise en place de réformes scolaires semble possible. Mais elles sont encore trop peu nombreuses. Que ce soit sous l'éclairage de la TBA, par le biais du TARGET, ou sur la base de la TAD et de son cadre tridimensionnel, soutien de l'autonomie-structure-implication, ces projets d'une plus grande validité écologique sont à encourager.

Étudier les effets de la manipulation du climat sur une période de temps longue. Savoir si les effets bénéfiques de ces études quasi expérimentales perdurent dans le temps représente aussi une question importante. Quelques recherches semblent montrer que les effets se sont quasiment, voire totalement dissipés dix mois après

l'intervention (Digelidis *et al.*, 2003) et mettent en évidence les multiples influences sociales susceptibles de les atténuer (les parents, les pairs, l'entraîneur sportif ; Weigand *et al.*, 2001). Néanmoins, de plus nombreuses études sont requises pour confirmer cette tendance.

Philippe Sarrazin
philippe.sarrazin@ujf-grenoble.fr
Université Joseph Fourier-Grenoble 1
Laboratoire « Sport et environnement Social »

Damien Tessier
damien.tessier@umb.u-strasbg.fr
Université Marc Bloch-Strasbourg 2
Laboratoire « Sport et environnement Social »

David Trouilloud
david.trouilloud@ujf-grenoble1.fr
Université Joseph Fourier-Grenoble 1
Laboratoire « Sport et environnement Social »

NOTES

- (1) En témoigne le Miroir du Débat qui reflète ce que les Français on dit de leur école. Il contient la synthèse présentant l'éventail des opinions, ainsi qu'un *verbatim* donnant un aperçu des propos tels qu'ils sont parvenus à la commission. Il a été disponible sur Internet à l'adresse suivante : <http://www.debatnational.education.fr>. Voir aussi le numéro d'avril 2004 du *Monde de l'Éducation* présentant un dossier spécial sur ce débat.
- (2) Bien que proche, ce dernier ne se confond pas avec « l'effet-classe » qui intègre d'autres éléments que l'action du professeur, tels que des facteurs contextuels qui s'imposent à l'enseignant (nombre d'élèves, niveau moyen de la classe, tonalité sociale ; Bressoux, 2001).

BIBLIOGRAPHIE

- ALTET M. (1993). « Styles d'enseignement, styles pédagogiques ». In J. Houssaye (éd), *La pédagogie, une encyclopédie pour aujourd'hui*. Paris : ESF, p. 89-101.
- AMES C. (1992a). « Achievement goals and the classroom climate ». In D. H. Schunk & J. L. Meece (éd.), *Student Perceptions in the Classroom*. Hillsdale : L. Erlbaum, p. 327-308.
- AMES C. (1992b). « Classrooms : Goals, structures, and student motivation ». *Journal of Educational Psychology*, vol. 84, p. 261-271.
- AMES C. & ARCHER J. (1988). « Achievement goals in the classroom : Student's learning strategies and motivation processes ». *Journal of Educational Psychology*, vol. 80, p. 260-267.
- AMABILE T. ; DEJONG W. & LEPPER M. (1976). « Effects of externally imposed deadlines on subsequent intrinsic motivation ». *Journal of Personality and Social Psychology*, vol. 34, p. 92-98.
- ANDERMAN L. & ANDERMAN E. M. (1999). « Social predictors of changes in students' achievement goal orientations ». *Contemporary Educational Psychology*, vol. 22, p. 269-298.
- ANDERMAN E. M. ; GRIESINGER T. & WESTERFIELD G. (1998). « Motivation and cheating during early adolescence ». *Journal of Educational Psychology*, vol. 90, p. 84-93.
- ANDERMAN E. M. ; MAEHR M. L. & MIDGLEY C. (1999). « Declining motivation after the transition to middle school : schools can make a difference ». *Journal of Research Development in Education*. vol. 32, p. 131-147.
- ANDERMAN E. M. & MIDGLEY C. (1997). « Changes in achievement goal orientations, perceived academic competence, and grade across the transition to middle level schools ». *Contemporary Educational Psychology*, vol. 22, p. 269-298.
- ANDERMAN E. M. & URDAN, T. (1995). « A multi-level approach to multi-level reform ». *Principal Magazine*, vol. 74, n° 3, p. 26-28.
- ASSOR A. ; KAPLAN H. & ROTH G. (2002). « Choice is good, but relevance is excellent : Autonomy-enhancing and suppressing teacher behaviours in predicting student's engagement in school work ». *British Journal of Educational Psychology*, vol. 72, p. 261-278.
- BENNETT S. I. (1976). *Teaching Styles and Pupil Progress*. London : Open Books.
- BIDDLE S. ; CURY F. ; GOUDAS M. ; SARRAZIN P. ; FAMOSE J.-P. & DURAND M. (1995). « Development of scales to measure perceived physical education class climate : A cross-national project ». *British Journal of Educational Psychology*, vol. 65, p. 341-358.

- BOGGIANO K. M. ; BARRETT M. ; WEIHER A. W. ; MCCLELLAND G. H. & LUSK C. M. (1987). « Use of maximal-operant principle to motivate children's intrinsic interest ». *Journal of Personality and Social Psychology*, vol. 53, p. 866-879.
- BOGGIANO K. M. ; FLINK C. ; SHIELDS A. ; SEELBACH A. & BARRETT M. (1993). « Use of techniques promoting students' self-determination : Effects on students' analytic problem solving skills ». *Motivation and Emotion*, vol. 17, p. 319-336.
- BRESSOUX P. (2001). « Réflexions sur l'effet-maître et l'étude des pratiques enseignantes ». *Les Dossiers des Sciences de l'Éducation*, vol. 5, p. 35-52.
- BROPHY J. (1999). « Research on motivation in education : past, present, and future ». In T. Urdan (éd.), *Advances in Motivation and Achievement. The Role of Context*. New York : JAI Press, p. 1-44.
- CERVELLO E. M. ; JIMENEZ R. ; DEL VILLAR F. ; RAMOS L. & SANTOS-ROSA F. J. (2004). « Goal orientations, motivational climate, equality, and discipline of Spanish physical education students ». *Perceptual and Motor Skill*, vol. 1, p. 271-83.
- CONNELL J. P. & WELLBORN J. G. (1991). « Competence, autonomy, and relatedness : Amotivational analysis of self-system processes ». In M. R. Gunnar & L. A. Sroufe (éd.), *Self processes in development : Minnesota Symposium on Child Psychology*. Hillsdale : L. Erlbaum, Vol. 23, p. 43-77
- CURY F. ; BIDDLE S. ; FAMOSE J.-P. ; GOUDAS M. ; SARRAZIN P. & DURAND M. (1996). « Personal and situational factors influencing intrinsic interest of adolescent girls in school physical education : A structural equation modelling analysis ». *Educational Psychology*, vol. 16, p. 305-315.
- DECI E. L. & CASCIO W. F. (1972). « Changes in intrinsic motivation as a function of negative feedback and threats ». Paper presented at the *Eastern Psychological Association*, Boston.
- DECI E. ; EGHARRI H. ; PATRICK B. & LEONE D. (1994). « Facilitating internalization : The self-determination theory perspective ». *Journal of Personality*, vol. 62, p. 119-142.
- DECI E. L. ; KOESTNER R. & RYAN R. M. (1999). « A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation ». *Psychological Bulletin*, vol. 125, p. 627-668.
- DECI E. L. & RYAN R. M. (2000). « The "what" and "why" of goal pursuits : human needs and the Self-Determination of behaviour ». *Psychology Inquiry*, vol. 11, p. 227-268.
- DECI E. L. & RYAN R. M. [éd.] (2002). *Handbook of Self-determination research*. Rochester : The University of Rochester Press.
- DECI E. L. ; SCHWARTZ A. J. ; SHEINMAN L. & RYAN R. M. (1981). « An instrument to assess adults' orientations toward control versus autonomy with children : Reflections on intrinsic motivation and perceived competence ». *Journal of Educational Psychology*, vol. 73, p. 642-650.
- DECI E. L. ; VALLERAND R. J. ; PELLETIER L. G. & RYAN R. M. (1991). « Motivation in education : The self-determination perspective ». *Educational Psychologist*, vol. 26, p. 325-346.
- DIGELIDIS N. ; PAPAIOUANNOU A. ; LAPARIDIS K. & CHRISTODOULIDIS T. (2003). « A one year intervention in 7th grade physical education aiming to change motivational climate and attitudes towards exercise ». *Psychology of Sport and Exercise*, vol. 4, p. 195-210.
- DROZDA-SENKOWSKA E. & HUGUET P. (2003). « Acquisition et régulation des compétences sociales ». In M. Kail & M. Fayol (éd.), *Les sciences cognitives et l'école : la question des apprentissages*. Paris : PUF, p. 153-212.
- DUDA J. L. (2001). « Achievement Goal research in sport : Pushing the boundaries and clarifying some misunderstandings ». In G. Roberts (éd.), *Advances in motivation in sport and exercise*. Champaign : Human Kinetics, p. 129-182.
- DURU-BELLAT M. & MINGAT A. (1994). « La variété du fonctionnement de l'école : identification et analyse des « effets-maître ». In M. Crahay et A. Lafontaine (éd.), *Évaluation et analyse des établissements de formation*. Bruxelles : De Boeck, p. 131-145.
- DWECK C. S. (1986). « Motivational processes affecting learning ». *American Psychologist*, vol. 41, p. 1040-1048.
- DWECK C. S. (1999). *Self-Theories and Goals : Their Role in Motivation, Personality, and Development*. Philadelphia : Taylor & Francis.
- DWECK C. S. & LEGGETT E. L. (1988). « A social-cognitive approach to motivation and personality ». *Psychological Review*, vol. 95, p. 256-273.
- ELLIOT A. J. (1997). « Integrating the "classic" and "contemporary" approaches to achievement motivation : A hierarchical model of approach and avoidance achievement motivation ». In M. Maehr & P. Pintrich (éd.), *Advances in motivation and achievement*. Greenwich : JAI Press, vol. 10, p. 243-279.
- ELLIOT A. J. & DWECK D. S. (1988). « Goals : An approach to motivation and achievement ». *Journal of Personality and Social Psychology*, vol. 54, n° 1, p. 5-12.
- ELLIOT A. J. & DWECK D. S. (2005). *Handbook of Competence and Motivation*. New York : Guilford Press.
- ELLIOT A. J. & HARACKIEWICZ J. M. (1996). « Approach and avoidance achievement goals and intrinsic motivation : A mediational analysis ». *Journal of Personality and Social Psychology*, vol. 70, n°3, p. 461-475.
- ELLIOT A. J. & MCGREGOR H. A. (2001). « A 2 x 2 achievement goal framework ». *Journal of Personality & Social Psychology*, vol. 80, p. 501-519.
- EPSTEIN J. L. (1988). « Effective schools or effective students : Dealing with diversity ». In R. Haskins & D. MacRae (éd.), *Policies for America's public schools : Teacher equity indicators*. Norwood : Ablex, p. 89-126.
- ESCARTI A. & GUTIEREZ M. (2001). « Influence of the motivational climate in physical education on the intention to practice physical activity or sport ». *European Journal of Sport Science*, vol. 1, n° 4, p. 1-12.
- FISHER C. D. (1978). « The effects of personal control, competence, and extrinsic reward system on performance ». *American Journal of Psychology*, vol. 94, p. 387-398.
- FLANDERS N. A. (1970). *Analyzing teaching behaviours*. Reading : Addison-Wesley.
- FLINK C. ; BOGGIANO A. K. & BARRET M. (1990). « Controlling teaching strategies : Undermining children's self-determination and performance ». *Journal of Personality and Social Psychology*, vol. 59, p. 916-924.
- GALAND B. ; PHILIPPOT P. & FRENAY M. (2006). « Structure de buts, Relations enseignants-élèves et adaptation

- scolaire des élèves : une analyse multi-niveaux ». *Revue française de pédagogie*, n° 155, p. 57-72.
- GOOD T. L. & BROPHY J. E. (2000). *Looking in classrooms*. New York : Longman [8th ed.].
- GOUDAS M. & BIDDLE S. (1994). « Perceived motivational climate and intrinsic motivation in school physical education classes ». *European Journal of Psychology of Education*, vol. 9, p. 241-250.
- GOUDAS M. ; BIDDLE S. ; FOX K. & UNDERWOOD M. (1995). « It ain't what you do, it's the way you do it ! Teaching style affects children's motivation in track and field lessons ». *The Sport Psychologist*, vol. 9, p. 254-264.
- GRANT H. & DWECK C. (2003). « Clarifying achievement goals and their impact ». *Journal of Personality and Social Psychology*, vol. 85, p. 541-553.
- GROLNICK W. S. & RYAN R. M. (1987). « Autonomy in children's learning : An experimental and individual difference investigation ». *Journal of Personality and Social Psychology*, vol. 52, p. 890-998.
- HARACKIEWICZ J. M. ; BARRON K. E. & ELLIOT A. J. (1998). « Rethinking achievement goals : When are they adaptive for college students and why ? » *Educational Psychologist*, vol. 33, p. 1-21.
- HARACKIEWICZ J. M. & ELLIOT A. J. (1998). « The joint effects of target and purpose goals on intrinsic motivation : A mediational analysis ». *Personality and Social Psychology Bulletin*, vol. 24, p. 675-689.
- HIDI S. & HARACKIEWICZ J. M. (2000). « Motivating the academically unmotivated : A critical issue for the 21st century ». *Review of Educational Research*, vol. 70, p. 151-179.
- KARABENICK S.A. (2004). « Perceived Achievement Goal Structure and College Student Help Seeking ». *Journal of Educational Psychology*, vol. 96, n° 3, p. 569-581.
- KAPLAN A. & MAEHR M. (2002). « Adolescent's achievement goals.; Situating motivation in sociocultural contexts ». In F. Pajares & T. Urdan (éd.), *Academic Motivation of Adolescents*. Greenwich : IAP, p. 125-167.
- KOESTNER R. ; RYAN R. ; BERNERI F. & HOLT K. (1984). « Setting limits on children's behaviour : The differential effects of controlling versus informational styles of intrinsic motivation and creativity ». *Journal of Personality*, vol. 52, p. 233-248.
- LA GUARDIA J. & RYAN R. (2002). « What adolescents need : A self-determination theory perspective on development within families, school and society ». In F. Pajares & T. Urdan (éd.) *Academic Motivation of Adolescents*. Greenwich : IAP, p. 193-219.
- LEPPER M. R. ; GREENE D. & NISBETT R. E. (1973). « Undermining children's intrinsic interest with extrinsic rewards : A test of the "over justification" hypothesis ». *Journal of Personality and Social Psychology*, vol. 28, p. 129-137.
- LEWIN K. ; LIPPITT R. & WHITE R. K. (1939). « Patterns of aggressive behavior in experimentally created social climates ». *Journal of Social Psychology*, vol. 10, p. 271-299.
- LINNENBRINK E. A. (2005). « The dilemma of performance-approach goals : The use of multiple goal contexts to promote students' motivation and learning ». *Journal of Educational Psychology*, vol. 97, n° 2, p. 197-213.
- MAEHR M. L. & ANDERMAN E. M. (1993). « Reinventing schools for adolescents : Emphasizing task goals ». *Elementary School Journal*, vol. 93, n° 5, p. 593-610.
- MAEHR M.L. & BRASKAMP L.A. (1986). *The motivation factor : A theory of personal investment*. Lexington : Lexington Books.
- MAEHR M. L. & MIDGLEY C. (1991). « Enhancing student motivation : A schoolwide approach ». *Educational Psychologist*, vol. 26, p. 399-427.
- MAEHR M. L. & MIDGLEY C. (1996). *Transforming School Cultures*. Boulder : Westview
- MAEHR M. L. & NICHOLLS J. G. (1980). « Culture and Achievement Motivation : A second Look ». In N. Warren (éd.), *Studies in Cross-Cultural Psychology*. New York : Academic Press, vol. 3, p. 221-247.
- MEECE J. L (1991). « The classroom context and children's motivational goals ». In M Maehr & P. Pintrich (éd.), *Advances in Achievement Motivation Research*. New York : Academic Press, p. 261-285
- MEECE J. L. ; ANDERMAN E. M. & ANDERMAN H. L. (2006). « Classroom goal structure, student motivation, and academic achievement ». *Annual Review of Psychology*, vol. 57, p. 487-503.
- MEECE J. L. ; HERMAN P. & MC COMBS B. (2003). « Relations of learner-centered teaching practices to adolescents' achievement goals ». *International Journal of Educational Research*, vol. 39, p. 457-475.
- MIDDLETON M. J. & MIDGLEY C. (1997). « Avoiding the demonstration of lack of ability : An unexplored aspect of goal theory ». *Journal of Educational Psychology*, vol. 89, p. 710-718.
- MIDGLEY C. ; ANDERMAN E. M. & HICKS L. (1995). « Differences between elementary and middle school teachers and students : a goal theory approach ». *Journal of Early Adolescence*, vol. 15, n°1, p. 90-113.
- MIDGLEY C. & MAEHR M. (1999). « Using motivation theory to guide school reform ». In A. J. Reynolds, H. J. Walberg, & R. P. Weissberg (éd.), *Promoting Positive Outcomes : Issues in Children's and Families' Lives*. Washington [DC] : Child Welfare League Am, p. 261-285.
- MIDGLEY C. & URDAN T. (1995). « Academic self-handicapping and performance goals : further examination ». *Contemporary Educational Psychology*, vol. 26, p. 61-75.
- MIDGLEY C. & URDAN T. (2001). « Academic self-handicapping and achievement goals : further examination ». *Contemporary Educational Psychology*, vol. 26, p. 61-75.
- MONTEIL J.-C. & HUGUET P. (2002). *Réussir ou échouer à l'école : une question de contexte ?* Grenoble : PUG.
- MURDOCK T. ; HALE N. & WEBER M. (2001). « Predictors of cheating among early adolescent : Academic and social motivations ». *Contemporary Educational Psychology*, vol. 26, p. 96-115.
- NICHOLLS J. G. (1984). « Achievement motivation : Conceptions of ability, subjective experience, task choice, and performance ». *Psychological Review*, vol. 91, p. 328-346.
- NICHOLLS J. G. (1989). *The competitive ethos and democratic education*. Cambridge [Mass.] : Harvard University Press.
- NICHOLLS J. G. ; CHEUNG P. C. ; LAUER J. & PATASHNICK M. (1989). « Individual differences in academic motivation :

- Perceived ability, goals, beliefs, and values ». *Learning and individual differences*, vol. 1, p. 63-84.
- NOLEN S. B. & HALADYNA T. M. (1990). « Personal and environmental influences on student's beliefs about effective study strategies ». *Contemporary Educational Psychology*, vol. 15, p. 166-130.
- NTOUMANIS N. (2001). « A self-determination approach to the understanding of motivation in physical education ». *British Journal of Educational Psychology*, vol. 71, p. 225-242.
- NTOUMANIS N. & BIDDLE S. J. H. (1999). « A review of motivational climate in physical activity ». *Journal of Sports Sciences*, vol. 17, p. 643-665.
- PAPAIOANNOU A. (1994). « Development of a questionnaire to measure achievement orientations in physical education ». *Research Quarterly for Exercise and Sport*, vol. 65, p. 11-20.
- PAPAIOANNOU A. (1995). « Differential perceptual and motivational patterns when different goals are adopted ». *Journal of Sport and Exercise Psychology*, vol. 17, p. 18-34.
- PAPAIOANNOU A. ; MARSH H. W. & THEODORAKIS Y. (2004). « A multilevel approach to motivational climate in physical education and sport settings : An individual or a group level construct ? ». *Journal of Sport & Exercise Psychology*, vol. 26, 1-29.
- PATRICK B. ; SKINNER E. ; & CONNELL J. (1993). « What motivates children's behavior and emotion ? Joint effects of perceived control and autonomy in the academic domain ». *Journal of Personality and Social Psychology*, vol. 65, p. 781-791.
- PELLETIER L. ; BOIVIN M. & ALAIN M. (2000). « Les plans de recherche corrélationnels ». In R. J. Vallerand & E. Hess (éd.), *Méthodes de recherche en psychologie*, New York : Guilford Press, p. 193-238.
- PELLETIER L. ; SÉGUIN-LÉVESQUE C. & LEGAULT L. (2002). « Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors ». *Journal of Educational Psychology*, vol. 94, p. 186-196.
- PINTRICH P. R. (2003). « A motivational science perspective on the role of student motivation in learning and teaching contexts ». *Journal of Educational Psychology*, vol. 95, p. 667-686.
- POSTIC M. & DE KETELE J.-M. (1988). *Observer les situations éducatives*. Paris : PUF.
- PROVENCHER G. (1983). « Les habiletés pour une communication pédagogique efficace chez les maîtres de l'enseignement professionnel ». *Revue des sciences de l'éducation*, vol. IX, n° 3, p. 419-431.
- REEVE J. (1996). *Motivating others : Nurturing inner motivational resources*. Boston : Allyn & Bacon.
- REEVE J. (1998). « Autonomy support as an interpersonal motivating style : Is it teachable ? ». *Contemporary Educational Psychology*, vol. 23, p. 312-330.
- REEVE J. (2002). « Self-determination theory applied to educational setting ». In E. L. Deci & R. M. Ryan (éd.), *Handbook of self-determination research*. Rochester : University of Rochester Press, p. 183-203.
- REEVE J. ; BOLT E. & CAI Y. (1999). « Autonomy-supportive teachers : How they teach and motivate students ». *Journal of Educational Psychology*, vol. 91, p. 537-548.
- REEVE J. ; DECI E. L. & RYAN R. M. (2004). « Self-determination theory : A dialectical framework for understanding socio-cultural influences on student motivation ». In S. Van Etten & M. Pressley (éd.), *Big Theories Revisited*. Greenwich : Information Age Press, p. 31-60.
- REEVE J. ; JANG H. ; HARDRE P. & OMURA M. (2002). « Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting task ». *Motivation and Emotion*, vol. 26, p. 183-207.
- REEVE J. ; JANG H. ; CARRELL D. ; JEON S. & BARSH J. (2004). « Enhancing students' engagement by increasing teachers' autonomy support ». *Motivation and Emotion*, vol. 28, p. 147-169.
- REEVE J. & JANG H. (2006). « What teachers say and do to support students' autonomy during a learning activity ». *Journal of Educational Psychology*, vol. 98, p. 209-218.
- RYAN A. M. ; GHEEN M. H. & MIDGLEY C. (1998). « Why Do Some Students Avoid Asking for Help ? An Examination of the Interplay among Students' Academic Efficacy, Teachers' Social-Emotional Role, and the Classroom Goal structure ». *Journal of Educational Psychology*, vol. 90, n° 3, p. 528-535.
- RYAN R. (1982). « Control and information in the interpersonal sphere : An extension of cognitive evaluation theory ». *Journal of Personality and Social Psychology*, vol. 43, p. 450-461.
- RYAN R. & DECI E. (2000). « Intrinsic and extrinsic motivations : Classic definitions and new directions ». *Contemporary Educational Psychology*, vol. 25, p. 54-67.
- RYAN R. & GROLNICK W. S. (1986). « Origins and pawns in the classroom : Self-report and projective assessments of individual difference in children's perceptions ». *Journal of Personality and Social Psychology*, vol. 50, p. 550-558.
- RYAN R. M. & LA GUARDIA J. G. (1999). « Achievement motivation within a pressured society : Intrinsic and extrinsic motivations to learn and the politics of school reform ». In T. URDAN (éd.) *Advances in motivation and achievement*. Greenwich : JAI Press, vol. 11, p. 45-85.
- RYAN R. M. ; MIMS V. & KOESTNER R. (1983). « Relation of reward contingency and interpersonal context to intrinsic motivation : A review and test using cognitive evaluation theory ». *Journal of Personality and Social Psychology*, vol. 45, p. 736-750.
- SARRAZIN P. ; FAMOSE J.-P. & CURY F. (1995). « But motivationnel, habileté perçue et sélection d'un niveau de difficulté d'une voie en escalade ». *STAPS*, vol. 38, p. 49-61.
- SARRAZIN P. ; ROBERTS G. ; CURY F. ; BIDDLE S. & FAMOSE J.-P. (2002). « Exerted effort and performance in climbing among boys : The influence of achievement goals, perceived ability, and task difficulty ». *Research Quarterly for Exercise and Sport*, vol. 73, p. 425-436.
- SARRAZIN P. ; TESSIER D. ; PELLETIER L. ; TROUILLOU D. & CHANAL C. (2006). « The Effects of Teachers' Expectations about Students' Motivation on Teachers' Autonomy-Supportive and Controlling Behaviour ». *International Journal of Sport and Exercise Psychology*, vol. 4, p. 283-301.
- SARRAZIN P. & TROUILLOU D. (2006). « Comment motiver les élèves à apprendre ? Les apports de la théorie de l'autodétermination ». In P. Dessus & E. Gentaz (éd.), *Comprendre les apprentissages, sciences cognitives et éducation*, Paris : Dunod, t. 2, p. 123-141.

- SKAALVIK E. M. (1997). « Self-enhancing and self-defeating ego orientation : Relations with task and avoidance orientation, achievement, self-protections, anxiety ». *Journal of Education Psychology*, vol. 89, n° 1, p. 71-81.
- SKINNER E. A. & BELMONT M. J. (1993). « Motivation in the classroom : Reciprocal effects of teacher behavior and student engagement across the school year ». *Journal of Educational Psychology*, vol. 85, p. 571-581.
- SKINNER E. & EDGE K. (2002). « Parenting, motivation, and the development of children's coping ». In L. J. Crockett (éd.), *Agency, Motivation, and the Life Course : The Nebraska Symposium on Motivation*. Lincoln : University of Nebraska Press, vol. 48, p. 77-143.
- STANDAGE M. ; DUDA J. & NTOUMANIS N. (2003). « A model of contextual motivation in physical education : using constructs from self-determination and achievement goal theories to predict physical activity intentions ». *Journal of Educational Psychology*, vol. 95, p. 97-110.
- STANDAGE M. ; DUDA J. L. & NTOUMANIS N. (2003). « Predicting motivational regulations in physical education : the interplay between dispositional goal orientations, motivational climate and perceived competence ». *Journal of Sports Sciences*, vol. 21, p. 631-647.
- STANDAGE M. ; DUDA J. L. & NTOUMANIS N. (2006). « Students' motivational process and their relationship to teacher ratings in school physical education : A self-determination theory approach ». *Research Quarterly for Exercise and Sport*. Vol. 77, p. 100-110.
- STEFANOUCI C. R. ; PERENCEVICH K. C. ; DI CINTIO M. & TURNER J.C. (2004). « Supporting autonomy in the classroom : Ways teachers encourage student decision making and ownership ». *Educational Psychology*, vol. 39, n° 2, p. 97-110.
- THEEBOOM M. ; DE KNOP P. & WEISS M. R. (1995). « Motivational climate, psychological response, and motor skill development in children's sports : A field based-intervention study ». *Journal of Sport and Exercise Psychology*, vol. 17, p. 294-311.
- THORNDIKE E. L. (1920). « A constant error in psychological ratings ». *Journal of Applied Psychology*, vol. 4, p. 469-477.
- THORKILDSEN T. A. & NICHOLLS J. G. (1998). « Fifth graders' achievement orientations and beliefs : Individual and classroom differences ». *Journal of Educational Psychology*, vol. 90, p. 179-201.
- TREASURE D. C. (1997). « Perceptions of the motivational climate and elementary school children's cognitive and affective response ». *Journal of Sports & Exercise Psychology*, vol. 19, p. 278- 290.
- TREASURE D. C. & ROBERTS G. C. (1995). « Application of achievement goal theory to physical education : Implications for enhancing motivation ». *Quest*, vol. 47, p. 45-489.
- TREASURE D. C. & ROBERTS G. C. (2001). « Student's perceptions of the motivational climate, achievement beliefs and satisfaction in physical education ». *Research Quarterly for Exercise and Sport*, vol. 72, p. 165-175.
- TROUILLOUD D. ; SARRAZIN P. ; BRESSOUX P. & BOIS J. (2006). « Relation Between Teachers' Early Expectations and Students' Later Perceived Competence in Physical Education Classes : Autonomy-Supportive Climate as a Moderator ». *Journal of Educational Psychology*, vol. 98, n° 1, p. 75-86.
- TRUONG N. (2003). « De l'enfant roi à l'élève client ». *Le monde de l'éducation*, n° 337, p. 30-47.
- TURNER J. C. ; MIDGLEY C. ; MEYER D. K. ; GHEEN M. ; ANDERMAN E. M. ; KANG Y. & PATRICK H. (2002). « The Classroom Environment and Students' Reports of Avoidance Strategies in Mathematics : A Multimethod Study ». *Journal of Educational Psychology*, vol. 94, n° 1, p. 88-106.
- URDAN T. (2004). « Using multiple methods to assess students' perceptions of classroom goal structures ». *European Psychologist*, vol. 9, n° 4, p. 222-231
- URDAN T. & MIDGLEY C. (2003). « Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence ». *Contemporary Educational Psychology*, vol. 28, p. 524-551.
- URDAN T. ; KNEISEL L. & MASON V. (1999). « Interpreting messages about motivation in the classroom, examining the effects of achievement goal structures ». In T. Urdan (éd.), *Advances in motivation and achievement*. Stamford, CT : JAI Press, vol. 11, p. 123-158.
- VALLERAND R. J. (1997). « Toward a hierarchical model of intrinsic and extrinsic motivation ». In M. P. Zanna (éd.), *Advances in Experimental Social Psychology*. New York : Academic Press, n° 29, p. 271-360.
- VALLERAND R. J. ; FORTIER M. S. & GUAY F. (1997). « Self-determination and persistence in a real life setting : Toward a motivational model of high school dropout ». *Journal of Personality and Social Psychology*, vol. 72, p. 1161-1176.
- VANSTEENKISTE M. ; SIMONS J. ; LENS W. ; SHELDON K. & DECI E. (2004). « Motivating learning, performance, and persistence : The synergistic effects of intrinsic goal contents and autonomy-supportive contexts ». *Journal of personality and social psychology*, vol. 87, n° 2, p. 246-260.
- VANSTEENKISTE M. ; SIMONS J. ; LENS W. ; SOEMONS B. & MOTOS L. (2005). « Examining the motivational impact of intrinsic versus extrinsic goal framing and autonomy-supportive versus internally controlling communication style on early adolescents' academic achievement ». *Child development*, vol. 76, n° 2, p. 483-501.
- VANSTEENKISTE M. ; SIMONS J. ; LENS W. ; SOEMONS B. ; MOTOS L. & LACANTE M. (2004) « "Less is sometimes more" : Goal-content matters ». *Journal of Educational Psychology*, vol. 96, p. 755-764.
- WEIGAND D.A. & BURTON S. (2002). « Manipulation Achievement motivation in physical education by manipulating the motivational climate ». *European Journal of Sports Science*, vol. 2, n° 1, p. 1-14.
- WEIGAND D. A. ; CARR S. ; PETHERICK C. & TAYLOR A. (2001). « Motivational climate in sport and physical education : The role of the significant others ». *European Journal of Sport Science*, vol. 1, n° 4, p. 1-13.
- WOLTERS C. (2004). « Advancing achievement goal theory : Using goal structure and goal orientation to predict students' motivation, cognition, and achievement ». *Journal of Educational Psychology*, vol. 96, p. 236-250.
- ZUCKERMAN M. ; PORAC J. ; LATHIN D. ; SMITH R. & DECI E. L. (1978). « On the importance of self-determination for intrinsically motivated behaviour ». *Personality and Social Psychology Bulletin*, vol. 4, p. 443-446.

NOTES CRITIQUES

BAUTIER Élisabeth (dir.). *Apprendre à l'école, apprendre l'école : des risques de construction d'inégalités dès la maternelle*. Lyon : Éd. de la Chronique sociale, 2006. – 256 p. (Pédagogie/formation).

Cet ouvrage collectif dirigé par Élisabeth Bautier est composé de neuf chapitres répartis dans trois parties distinctes. La première partie situe l'école maternelle dans une perspective historique en retraçant les principales missions qui lui étaient assignées : la protection, l'éducation, l'apprentissage. Aujourd'hui, l'école maternelle tend à devenir une école, c'est-à-dire un espace d'apprentissage. Cette transformation justifie logiquement l'objectif que s'assignent ensuite les auteurs dans les deuxième et troisième parties de l'ouvrage et qui consiste à identifier empiriquement et à problématiser théoriquement les obstacles rencontrés tant par les enseignants que par les élèves dans la fabrication d'un rapport au savoir à l'école maternelle. Si certains chapitres auraient pu mûrir davantage avant publication, l'ouvrage constitue cependant un ensemble intégré dont la cohérence se noue autour d'une distinction centrale entre « le faire » et « l'apprendre » à l'école maternelle. Cette distinction actualise, au niveau de l'école maternelle, une réflexion dont on trouve des traces dans la plupart des travaux de l'équipe ESCOL depuis plusieurs années et qui s'enracine, notamment, dans les travaux de Basil Bernstein : le fait pour l'élève d'être en train de *faire* quelque chose en classe (qu'il s'agisse d'un exercice systématique ou d'une activité globale) ne signifie pas nécessairement que ce dernier est en train d'*apprendre* quelque chose, et ne suffit en tout cas pas à définir (ni pour lui, ni pour son professeur) ce qu'il est éventuellement en train d'apprendre. L'école maternelle au sein de laquelle le « faire » (colorier, dessiner, jouer, se mouvoir) tient une place centrale, mais qui aujourd'hui poursuit en même temps des objectifs d'apprentissage, constitue à l'évidence un terrain particulièrement favorable à l'examen empirique d'une telle tension.

Le titre de l'ouvrage – *Apprendre à l'école, apprendre l'école* – exprime assez bien sa thèse centrale. Il souligne la nécessité, pour apprendre à l'école, de comprendre la logique, souvent implicite, de l'école. Ainsi, par exemple, les apprentissages scolaires présupposent que l'on appréhende le langage en tant qu'objet d'analyse (« l'apprendre » qui nécessite distanciation et objecti-

vation des savoirs), et pas, ou pas uniquement, en tant qu'outil de communication (« le faire » qui peut se réaliser dans un usage du savoir immédiat et non distancié). Or, dans certaines familles, souvent défavorisées, la constitution de la langue (ou d'autres savoirs) comme objet d'analyse et d'apprentissage ne va pas de soi. Dans d'autres milieux sociaux, à l'inverse, les enfants sont souvent sollicités par leurs parents pour réfléchir sur les mots qu'ils utilisent, sur leur agencement, sur la logique de la langue comme ensemble structuré de règles formelles. La distinction (et le risque de confusion) entre « le faire » et « l'apprendre » est ainsi mise en relation étroite avec la problématique des inégalités sociales de réussite scolaire, comme l'indique le sous-titre de l'ouvrage.

Au fil des chapitres, à partir des divers matériaux empiriques collectés, les auteurs mettent en évidence des séquences pédagogiques au cours desquelles la confusion entre le faire et l'apprendre se réalise. Les auteurs montrent que la confusion peut dans bien des cas être produite par les enseignants eux-mêmes, par exemple lorsque les consignes énoncées par l'enseignant lors de certains cours d'éducation physique, portent exclusivement sur l'exécution des mouvements attendus sans explicitation de l'apprentissage visé, encore moins de sa raison d'être ou de ses liens avec d'autres apprentissages. L'ambiguïté est également alimentée par les enseignants lorsque ceux-ci rappellent sans cesse aux élèves l'importance du soin et de la qualité de l'exécution (un cahier soigné) : dans l'esprit de certains élèves, le soin peut alors devenir l'objectif et tenir lieu de critère de réussite, alors même que l'apprentissage effectivement poursuivi est autre (p. 142). Plus généralement, les difficultés de construction d'un rapport au savoir pour les élèves (mais aussi pour les enseignants) peuvent se révéler à partir d'une variété d'indications empiriques que les auteurs mettent bien en évidence, comme par exemple la manière dont les enfants sont parfois conduits à confondre la fin temporelle d'une activité et sa finalité cognitive (p. 120). Il arrive aussi qu'une activité en cache une autre, comme dans le cas rapporté en introduction (p. 10) où un exercice présenté oralement par l'enseignant en termes de coloriage et de découpage, poursuit en réalité un objectif de reconnaissance des mots à découper dans une phrase : certains élèves consacrent alors toute leur attention à découper

et colorier avec une grande application mais échouent à identifier l'objectif effectivement poursuivi (reconnaître des lettres, des mots et constituer le langage en objet de connaissance). Les élèves s'appliquent alors « à faire la tâche demandée, mais restent éloignés du savoir supposé être enseigné » (p. 126).

En soulignant que les « vulgates constructivistes actuelles » (p. 12) insistent avant tout sur la mise en activité des élèves et peuvent donner à penser que seul compte le fait d'être en train de *faire* quelque chose, les auteurs ne veulent cependant pas reconstituer un débat entre méthodes transmissives et méthodes actives (p. 250). Leur préoccupation nous semble surtout centrée sur l'identification des risques, plus ou moins invisibles, qu'encourent toujours les méthodes actives. En mettant empiriquement en évidence de tels risques, ils ne cherchent pas à disqualifier les méthodes en question mais bien à mettre en évidence les conditions de leur efficacité (et de leur équité) pédagogique.

On peut toutefois regretter le peu d'indications dont les auteurs font état pour caractériser les établissements étudiés. En effet, rien n'indique quelles sont les caractéristiques sociales du public des établissements scolaires au sein desquels les enquêtes ont été menées. Il ne s'agit pas là d'une simple remarque de méthode. Dans un contexte où il est de plus en plus évident qu'il existe une hiérarchie sociale des établissements, l'absence d'indication et surtout de problématisation à propos des établissements dans lesquels les enquêtes ont été menées, empêche d'articuler le questionnement des auteurs avec la problématique des inégalités entre établissements scolaires : si la critique, plus ou moins implicite, des auteurs à l'égard des enseignants (ou de l'école), réside dans le déficit de stratégies pédagogiques visant à constituer un « rapport au savoir » chez les élèves (qui ne le possèdent éventuellement pas initialement), la question qui consiste à savoir si de telles stratégies sont davantage présentes dans certaines écoles plutôt que dans d'autres, n'est pas traitée par les auteurs. L'hypothèse selon laquelle les enseignants auraient tendance à favoriser un rapport distancié et objectivé au savoir dans les écoles fréquentées par des classes supérieures et moyennes, et à y renoncer (pour privilégier un rapport plus exclusif et direct au « faire ») dans les écoles accueillant un public issu de milieux défavorisés, n'est pas posée. Or, une telle hypothèse, infirmée ou confirmée par la confrontation aux matériaux empiriques, aurait probablement pu nourrir la réflexion des auteurs sur cette problématique et consolider la portée critique de leur travail.

Éric Mangez

Université catholique de Louvain
Cerisis & Girsef

CASPARD Pierre ; LUC Jean-Noël & SAVOIE Philippe. *Lycées, lycéens, lycéennes : deux siècles d'histoire*. Lyon : INRP, 2005. – 501 p.

Issu d'un colloque organisé en 2002 par le centre Roland Mousnier de l'université Paris IV-Sorbonne et le Service d'histoire de l'éducation de l'INRP, à l'occasion du bicentenaire de la loi du 11 floréal an X (1^{er} mai 1802) instituant les lycées, ce volume, discret sur le versant de la commémoration, rassemble de nombreuses communications originales, dont certaines annoncent même des perspectives nouvelles dans le domaine de recherche historique considéré.

Une première partie décrit les enjeux politiques, sociaux et culturels de l'enseignement secondaire. Elle propose des articles de Philippe Savoie, sur les grandes étapes de la création des lycées et de l'enseignement secondaire ; de Marie-Madeleine Compère, qui nous fait découvrir les boursiers nationaux depuis l'Ancien Régime jusqu'à l'Empire ; de Julien Vasquez, qui traite le cas des lycées aquitains au début du XIX^e siècle ; de Jean-Luc Marais, qui interroge l'action des aumôniers des lycées ; de Patrick Cabanel, sur la présence des protestants et des juifs dans les lycées de la III^e République ; de Frédéric Abécassis, à propos de la mission laïque française en Égypte au XX^e siècle ; de Jean-Michel Chapoulie, sur les promoteurs des grandes réformes du Front Populaire et de la IV^e République, moments clefs de la démocratisation de l'enseignement secondaire ; et de Youenn Michel, sur l'admission des langues régionales au lycée depuis cinquante ans.

La seconde partie pénètre dans l'univers des contenus d'enseignement et nous fait comprendre leur structuration progressive, suivant des thématiques avec lesquelles les historiens du Service d'histoire de l'éducation nous ont familiarisés. On y trouve les études d'Annie Bruter, sur la création et l'autonomisation de l'enseignement de l'histoire pendant le premier XIX^e siècle ; de Martine Jey, sur l'enseignement littéraire au tournant des XIX^e et XX^e siècles ; de Philippe Marchand, sur les sujets d'histoire du baccalauréat entre 1880 et 1914 ; de Jean-Yves Dupont, sur l'enseignement de la mécanique dans les classes préparatoires scientifiques ; de Renaud D'Enfert, qui confronte et rapproche l'enseignement des mathématiques dans le secondaire et dans le primaire au début du XX^e siècle ; d'Evelyne Héry, qui retrace l'application de la réforme de 1902 par les professeurs de l'époque ; et de Ludivine Bantigny, sur les courants de modernisation dans les lycées de la seconde moitié du XX^e siècle.

La troisième partie aborde les sociabilités, les représentations et les pratiques collectives des acteurs de l'enseignement secondaire, ce qui situe cette fois la réflexion à

la croisée des investigations historiques et des analyses sociologiques. Bruno Poucet relate la conquête de l'identité professionnelle des professeurs de philosophie. Yves Verneuil présente une enquête sur la Société des agrégés. André D. Robert rappelle le rôle d'une catégorie certes connue mais trop peu étudiée, les professeurs des classes élémentaires des lycées entre la III^e et la V^e République. Enfin, deux études prennent pour objet les élèves de nos établissements modernes : Anne-Marie Sohn s'est penchée sur les lycéens et lycéennes des années 1960, et Robi Morder a examiné leurs formes de conscience et d'action collectives.

La quatrième et dernière partie complète ce riche panorama en traitant de l'ordre et du changement dans l'administration des lycées. Deux articles s'intéressent à l'architecture des établissements, ceux de Marc Le Cœur sur la période 1802-1940 et d'Antoine Prost sur les collèges et lycées de 1960 à 1985. Jean-François Condette expose le point de vue des recteurs d'académie face à ce qui se révèle comme une crise des lycées de garçon à la fin du XIX^e siècle ; tandis que Stéphanie Dauphin évoque le rôle d'Octave Gréard à Paris à la fin du XIX^e siècle. Cette partie contient aussi deux études sur les modes de vie scolaires : dans la première Jean-Claude Caron, à partir d'une enquête de 1853-1854, analyse les régimes disciplinaires des lycées du second Empire, et dans la seconde Patrick Clastres envisage les problèmes posés par (et autour de) l'internat public.

Il faut s'arrêter avant cela à l'introduction de J.-N. Luc, « À la recherche du "tout puissant Empire du milieu" », texte de synthèse d'un grand intérêt puisque, après un résumé éclairant des grandes lignes de l'histoire de l'institution, il parcourt de façon sinon exhaustive du moins systématique la bibliographie existante sur le sujet depuis un siècle (et souvent l'auteur a soin de préciser la provenance institutionnelle donc le statut social et culturel des auteurs cités). Il y a jusqu'à présent peu de travaux divulgués au-delà du cercle des spécialistes : la plupart du temps, on s'arrête aux études de la regrettée Françoise Mayeur, d'Antoine Prost, ou plus avant de Paul Gerbod ou de Maurice Gontard. Mais grâce à J.-N. Luc nous disposons d'un bilan détaillé de la recherche et de ses résultats anciens ou plus récents, ce qui pourrait faire de ce texte, même s'il n'est pas seul de son espèce, un précieux outil pour tout chercheur débutant ou confirmé en ce domaine. À cela fait écho en fin de volume la conclusion de P. Caspard. Celui-ci, après avoir noté le regain d'intérêt pour l'histoire de l'éducation en général, apprécie de façon contrastée les progrès réalisés sur le terrain de l'enseignement secondaire, et signale les nouveaux objets, révélant aussi, d'une part les zones investies et celles encore en friche, et d'autre part les élaborations

aussi bien historiographiques que conceptuelles les plus prometteuses – sur les disciplines scolaires, sur l'établissement scolaire, sur les élèves, entre autres.

Il faut en outre préciser que le réel dynamisme de l'histoire de l'éducation et de l'enseignement dont témoigne ce volume ne se comprend que si l'on connaît l'évolution typique par laquelle ce secteur a intégré l'essentiel des préoccupations et de l'épistémologie de la « nouvelle histoire ». Comme le note J.-N. Luc : « l'histoire des lycées [...] a su profiter de l'essor de l'histoire culturelle et de l'apport de l'histoire des représentations, de la fécondité de l'histoire sociale, du regain d'intérêt pour la micro-histoire et du développement de l'histoire des femmes et de la jeunesse ».

Cet ouvrage a d'abord une singularité qui mérite d'être remarquée avant toute autre. Il nous révèle que cette institution, le lycée, et ce segment, l'enseignement secondaire – selon une dénomination qui a cours depuis le début du XIX^e siècle –, ne se réduisent pas aux seuls caractères de stabilité et de permanence auxquels on s'attendrait, par exemple si l'on pensait simplement, du côté des élèves, à l'examen du baccalauréat, ou du côté des professeurs, au rôle de l'agrégation. L'histoire qu'on nous dessine ici ne suit pas le cours d'un fleuve tranquille. C'est une histoire foisonnante et sinueuse, parfois même agitée ; et ce fait est d'autant plus surprenant que l'enseignement secondaire accueille pendant longtemps un tout petit nombre d'élèves.

On a en mémoire les grandes ruptures politiques : l'invention des écoles centrales sous la Révolution puis leur suppression en 1802, la création du monopole napoléonien (avec certains espaces d'initiative laissés à l'Église et aux instances privées, il est vrai), puis la vigoureuse contestation catholique des années 1840, la liberté d'enseignement accordée par la loi Falloux en 1850 ; et au XX^e siècle la grande réforme de 1902, puis, il y a quarante ans, l'arrivée de ce qu'on peut appeler le collège démocratique. Mais ce n'est pas de cela, de cette histoire politique et institutionnelle, qu'il s'agit. Si le regard se déplace vers les phénomènes sociaux, on découvre les acteurs et leurs « stratégies » (les corporations à la conquête de leur « spécialité » ou de leur identité, les publics et l'affirmation de leurs intérêts, les administrations et la poursuite de leurs idéaux) ; vers les phénomènes culturels, on observe la production interne de la culture scolaire, donc la structuration des « matières » et la définition de catégories savantes appropriées, sans ignorer bien sûr à l'externe les contextes de formation et de circulation des savoirs, etc. Or dans tous les cas, apparaissent toutes sortes d'hésitations, de décalages et de divergences, voire de contradictions. Divergences entre des héritages et des innovations,

entre des emprunts et des créations, entre des projets de l'État central et des réalités locales résistantes ou réticentes, mais aussi et surtout entre des logiques fonctionnelles de tous ordres, toujours difficiles à unifier... On en a l'exemple avec la tension dont Ph. Savoie explique qu'elle confronte, au début du XIX^e siècle, une logique de l'établissement scolaire (et sa plus ou moins bonne réputation), et une logique de la chaire professorale (et sa plus ou moins haute dignité). D'autant que cette dernière suscite la spécialisation donc l'éclatement des disciplines : l'histoire se sépare des lettres, comme les sciences physiques des mathématiques, tandis que les langues vivantes se renforcent. C'est ainsi également qu'A. Bruter suit la migration de la discipline historique de la périphérie vers le centre du curriculum scolaire, dans le contexte de la restauration des humanités classiques ; ou que M. Jey, complétant les travaux d'A. Chervel, suit l'évolution du français au sein des Belles Lettres à la fin du XX^e siècle, en montrant que cette discipline s'éloigne cependant des humanités, s'autonomise, et engage une « transformation des pratiques rhétoriques en savoirs sur les textes littéraires » où s'imposent bientôt la dissertation littéraire et de l'explication française. On pourrait bien sûr transporter le schéma sur d'autres terrains, notamment celui de la question religieuse abordée par les études de P. Cabanel et J. L. Marais. Nous percevons donc une histoire jamais tracée par des réformateurs prophétiques, mais traversée par les nombreux conflits qui tiennent au rôle alors joué par l'enseignement secondaire dans la société de l'époque.

Une seconde originalité de l'ouvrage, certes conforme à ce qu'on connaît depuis une ou deux décennies dans la recherche historique, mais qui donne ici toute sa mesure, c'est l'exigence de se tenir au plus près des réalités pratiques. Ceci explique les choix de méthode effectués, en particulier les sources utilisées par les auteurs de ces études. On nous présente en effet, disons pêle-mêle : des travaux d'élèves, des rapports d'inspection, des rapports de conférences pédagogiques, des sujets d'examen, des extraits de correspondances administratives, des dossiers de personnels, des enquêtes, des fragments de la presse spécialisée ou générale, des romans, des autobiographies, etc. Mais n'en déduisons pas que ces orientations aient pour simple motif la recherche du concret et du qualitatif pour eux-mêmes. Le but est d'atteindre le grain de réalité où s'imposent les problèmes qui, aux différentes époques, et dans la durée, ont pu déclencher une action, formuler une pensée, bref initier une expérience collective. On le constatera dans tous les domaines explorés – les modalités et les contenus de l'enseignement, le fonctionnement des institutions, les procédures administratives, les stratégies des acteurs, etc.

Qui plus est, ceci permet de revisiter des moments dont nous n'avions peut-être qu'une vision générale et abstraite, comme la réforme de 1902, qui a introduit les sections « modernes », scientifiques, qui a décidé la classe d'une heure au lieu de deux, qui a incité au cours parlé plutôt que dicté, etc. Dans une passionnante étude sur les évolutions de tous ordres entraînés par cette grande rupture, E. Héry décrit ainsi les réactions des professeurs appelés à entrer dans le nouveau système par les injonctions ministérielles. R. D'Enfert adopte une démarche semblable, en exposant, de façon aussi originale, la formation d'une culture mathématique commune au secondaire et au primaire dans la même période, pour identifier à la suite un ensemble de normes d'action très modernes, en particulier l'exigence d'adapter les contenus aux possibilités intellectuelles des élèves, ce dont on attribue à tort l'invention à des courants plus marginaux (les courants de l'« Éducation nouvelle »). Cette démarche caractérise également les études d'Annie Bruter ou de Ph. Marchand sur l'enseignement de l'histoire, comme celles de J.-C. Caron ou de P. Clastres sur le régime disciplinaire et les punitions au XIX^e siècle. Il y a là au total un grand réalisme historique, dont les bénéfiques et les résultats sont fort sensibles, et dont il reste à espérer qu'il suscite l'intérêt des sciences de l'éducation en général, en plus de celui des historiens...

Au-delà de son contenu factuel, et quand on se déplace plus avant dans le XX^e siècle, l'histoire des lycées et des lycéens retracée dans ce volume a enfin l'avantage de saisir la provenance donc la consistance de nos idéaux de modernité – de notre volonté de « modernisation » faudrait-il dire – ce qui recouvre autant d'interrogations très prégnantes dans le présent des débats et des polémiques relatives au système éducatif. « Enseignement moderne », « méthodes modernes », « culture moderne » (sans oublier « démocratie », « égalité des chances », etc.), toutes ces expressions sont à la fois présentes et assez anciennes. Ce n'est donc pas le moindre mérite des études rassemblées ici que de nous mettre en garde contre la vraie innocence ou la feinte naïveté de ceux qui voudraient expliquer les difficultés actuelles de l'enseignement secondaire par les turbulences – funestes selon eux –, des années 1960 et la « massification » du public scolaire qui en a résulté. En l'occurrence, l'histoire des lycées et des lycéens désigne au moins deux grands registres sur lesquels se joue, de façon problématique insistons-y, une grande part de l'histoire de la modernité éducative.

Sur le plan pédagogique tout d'abord, où s'affirment les normes pratiques ou « techniques » des activités d'enseignement et d'apprentissage, on voit d'abord s'étendre le souci des méthodes, et plus encore de ces méthodes dites « modernes », « nouvelles », « actives », etc. Constat

d'autant plus intéressant que nous pourrions réserver ces questions au seul niveau de l'enseignement primaire. En réalité, elles n'ont pas cessé de se poser dans l'enseignement secondaire à partir du début du siècle. Dans les études d'E. Héry ou de R. D'Enfert notamment, nous voyons comment elles se forment au cœur de la grande rupture de 1902. Avec l'article de L. Bantigny sur « la modernité dans le lycée des années 1950 », comme avec celui de J.-M. Chapoulie sur la pensée et l'action des deux grands réformateurs que furent Gustave Monod et Paul Langevin, le premier depuis le Front populaire et le second à la Libération, nous commençons de même à entrevoir ce qui est sans doute un itinéraire inattendu des « méthodes actives ».

Certes, nous savions que ces conceptions furent portées par des groupes d'influence qui les justifièrent dans une doctrine globale de la rénovation pédagogique ; ce fut le cas du Groupe français d'Éducation nouvelle (GFEN), créé en 1921, à la tête duquel on trouve alors Paul Langevin, Henri Wallon ou Henri Piéron, et qui était de longue date lié à des instances et des personnes comme Ferrière en Suisse, Decroly en Belgique, etc. On connaît de même les expériences à grande échelle que furent les lycées pilotes et les sept cents « classes nouvelles » ouvertes en 1946 – avec des effectifs réduits, le développement des arts, l'introduction des travaux manuels et de l'« étude du milieu de vie » – classes qui seront supprimées en 1952 (surtout pour des raisons budgétaires et par manque d'enseignants). Mais, chose moins connue et plus significative encore, se constitue en dehors de ces réseaux militants une vulgate dont les progrès sont peu à peu sensibles dans tous les domaines de la vie des établissements. On pourrait évoquer en ce sens ne serait-ce que l'éducation civique, qui ne cesse de réviser ses fins et ses moyens ; mais aussi le désir réitéré de modifier les pratiques de la discipline, d'améliorer les relations avec les parents, ou encore, nous apprend A. Prost, d'agir sur les réalisations architecturales.

Le second plan de modernisation est celui sur lequel s'effectue le choix des contenus de la culture scolaire mais aussi et surtout, à travers ces choix, la définition des fins de l'enseignement et l'appel aux idéaux éducatifs correspondants. L'article de J.-N. Luc nous a d'abord présenté la césure survenue depuis le début du XIX^e siècle au moins dans la culture classique, lorsque les sciences émergent à côté puis en face du continent littéraire (sans qu'on doive confondre l'enseignement de la rhétorique et celui, ultérieur, de la littérature). En 1852, Fortoul crée la « bifurcation », qui permet de suivre une section scientifique après la quatrième ; quelques années plus tard est institué le secondaire « spécial », qui aura son baccalauréat, dit « moderne », en 1891 ; et c'est la réforme de 1902, on

l'a déjà constaté, qui marque l'étape décisive du processus de disjonction, en ce qu'elle offre, à partir de la classe de seconde, trois sections classiques dont l'une comporte des sciences, et la fameuse section « moderne », sans latin, le tout conduisant à un bac unique avec quatre options. Or, d'après ces repères, nous pouvons entrevoir plusieurs phénomènes distincts mais convergents au XX^e siècle.

Premier phénomène, une diversification, qui semble sans limites, de la configuration globale de la culture scolaire. De là découle l'essor des sections scientifiques à partir des années 1950 : le bac « sciences expérimentales » est créé en 1946, le bac « mathématiques et technique » en 1947 ; puis arrivent les sections techniques et économiques (en 1954), ce qui donne les bacs A, B, C, D et E ; à quoi s'ajoutent en 1965 les autres sections techniques que sont les F, G et H, si bien que s'efface la catégorie des lycées classiques et modernes au profit de la dualité des lycées d'enseignement général et des lycées d'enseignement technologique. Ne manque plus que la création en 1985 des baccalauréats professionnels, pour que se mette en place la dualité actuelle des lycées d'enseignement général et technologique et des lycées d'enseignement technique et professionnel.

Second phénomène important, aux conséquences tellement sensibles aujourd'hui, le renversement de la hiérarchie entre les lettres et les sciences dans cette configuration. On se souvient que la filière scientifique, dite « moderne », est restée longtemps de moindre prestige (donc plutôt féminine)... C'est bien l'inverse qui est vrai aujourd'hui. De ce point de vue, on ne saurait trop souligner le rôle des réformes et des réformateurs de l'après-guerre dont traite ici J.-M. Chapoulie. Les plans de réforme liés aux noms de G. Monod et de P. Langevin ont ainsi affirmé, certes le retrait du latin, mais aussi le crédit des nouvelles matières citées plus haut, les travaux manuels et l'« étude du milieu » (vieux thème des pédagogies nouvelles et alors lié aux « classes nouvelles »), en se réclamant des sciences expérimentales par différence et même en opposition à la culture gréco-latine.

Un troisième phénomène complique sans doute le précédent, c'est la montée d'une autre exigence, lisible par exemple dans un rapport de 1949 cité par J.-M. Chapoulie, qui manifeste la volonté de « préparer [l'enfant] à comprendre son temps et à y jouer un rôle actif ». Cette tendance, que décrit aussi l'article de L. Bantigny à propos du lycée des années 1950 en termes d'« adaptation de la culture lycéenne au “monde moderne” », apporte un nouveau mode de définition et de valorisation de la culture scolaire (en rapport étroit avec l'importation des méthodes actives), en ce qu'il propose d'abolir la frontière entre la culture scolaire et le monde extérieur à l'école.

À cela fait d'ailleurs écho le fameux slogan des pédagogies nouvelles, qui a eu cours surtout dans l'enseignement primaire cette fois : « ouvrir l'école sur la vie »...

On pourrait, en bonne logique historique et sociologique, mais prudemment, mettre ces phénomènes en rapport avec la pression exercée sur les élèves depuis la fin des années 1950 par la culture de masse spécialement destinée à cette nouvelle classe d'âge, qui en devient du coup une véritable figure sociale : la « jeunesse » – même s'il faut bien admettre comme le souligne A.-M. Sohn que c'est autant la scolarité et le lycée qui façonnent en retour, dans la société, l'image de l'adolescence – d'où la prégnance de la vie lycéenne et des rituels scolaires dans les médias.

Pour finir, insistons donc sur le grand intérêt extra scientifique de ces données : elles rendent possible un ré-examen du lien noué, à l'origine de la période de démocratisation, entre d'une part l'accès généralisé des enfants à l'enseignement secondaire par le truchement du collège puis du lycée dans sa version contemporaine, et d'autre part cette double évolution, pédagogique et culturelle, qui réoriente la transmission scolaire dans la forme et dans le fond. Une interrogation d'autant plus urgente qu'elle est probablement le point aveugle des polémiques d'aujourd'hui. Car cette réorientation de la transmission scolaire n'est pas une conséquence mécanique et incontrôlée de la « massification », elle n'est pas une sorte d'ajustement passif, effectué par des réformateurs sans idéaux, aux capacités et à la mentalité des « nouveaux publics » – même si cette idée d'adaptation apparaît effectivement dans l'après-guerre, notamment sous la plume de G. Monod. L'histoire des lycées nous donne une version différente, en un mot, de ce rapport entre la démocratisation, avec ses conséquences de massification (y compris associée à la critique de l'élitisme) et les évolutions pédagogiques et culturelles de l'école. S'il y a bien une articulation entre les deux termes du rapport, celle-ci reste par conséquent à examiner, à élucider, sans oublier que chaque terme a une logique qu'il lui appartient, une temporalité propre, en un mot une histoire singulière.

François Jacquet-Francillon
Université Lille 3-Charles de Gaulle & INRP

CHALMEL Loïc. *Oberlin : le pasteur des Lumières*. Strasbourg : La Nuée bleue, 2006. – 237 p.

Professeur à l'université de Rouen, Loïc Chalmel est aussi vice-président et coordonnateur du Conseil scientifique du musée Oberlin. L'ouvrage qu'il nous propose ici semble davantage l'œuvre du « conservateur » que

celle de l'universitaire. Loïc Chalmel a fait paraître en 1999 un ouvrage à vocation scientifique sur le pasteur Oberlin (*Le pasteur Oberlin*, Paris : PUF, 1999). Le présent ouvrage s'en distingue sans pour autant se réduire à une visite de musée. C'est à un voyage et à un partage que l'auteur nous invite, à travers un ouvrage qui se présente comme une « biographie thématique », un bel objet, de surcroît, à l'iconographie riche et étonnante. Le voyage dans le temps et l'espace, le partage s'adressent aux éducateurs d'aujourd'hui, non dans un esprit nostalgique, mais dans le but de faire vivre l'actualité des valeurs, des idéaux et des pratiques du passé. S'agissant de ce projet, Loïc Chalmel n'hésite pas à se considérer comme dépositaire d'un « devoir de mémoire » : « Le musée et le biographe s'associent en effet à un devoir de mémoire, en remodelant les matériaux du passé sans les reproduire à l'identique » (p. 202). L'auteur est bien ainsi « conservateur », mais « des choix de valeurs effectués par des individus singuliers, dans des contextes socio-historiques particuliers » (*ibid.*) ; il se fait, par là, « entremetteur d'une culture patrimoniale tout à la fois spécifique et universelle » (*ibid.*), l'idée étant que le rapport que les éducateurs entretiennent avec le passé doit être vivant et vitalisant : le passé ne saurait être considéré « de manière linéaire », « comme si tout était irréversiblement écrit », mais comme un « espace d'expériences, source d'inspirations, [qui] permet de glaner des sens perdus, devient susceptible de faire renaître des possibles inachevés, ou d'ensemencer des utopies restées en friche... ». Autant dire que Loïc Chalmel, ici comme dans l'ensemble de ses travaux scientifiques, illustre et défend avec toute l'exigence requise le domaine de l'histoire de la pédagogie si négligé dans le champ des sciences de l'éducation.

Il est vrai que le pasteur Oberlin occupe une place peu commune dans l'histoire de la pédagogie. Germanique de naissance et latin d'adoption, piétiste morave, il a exercé durant cinquante-neuf ans son ministère au Ban de la Roche, dans cette terre déshéritée des Vosges alsaciennes dont les habitants sont si pauvres qu'on appelle « riches » ceux dont la misère est moindre. Son prédécesseur et maître, dans le petit village de Waldersbach, est Jean-Georges Stuber (1722-1797), « précurseur autant visionnaire que méconnu de l'éducation populaire, par l'apprentissage du français langue étrangère, la pratique du chant choral, le développement des premières écoles pour adultes, de la première bibliothèque de prêt au monde, de la formation des maîtres, etc. » (p. 8). Jean-Frédéric Oberlin inscrit ses pas dans ceux de son prédécesseur et, dans un premier temps, parachève son œuvre : sous son ministère, la bibliothèque de prêt prend une extension importante. Accroissement du fonds, tant quantitatif que qualitatif. Les ouvrages mis à la disposition du public traitent d'une grande diversité de sujets : religion, philosophie, biologie,

astronomie, géographie, littérature. Ainsi les éducateurs et régents du Ban de la Roche disposent-ils d'une documentation pédagogique très fournie grâce aux cahiers thématiques que le pasteur rédige à partir de ses propres lectures. Peu à peu, la documentation se complète d'une collection d'histoire naturelle. Bien sûr, la construction de ces supports pédagogiques s'organisent autour de l'idée centrale selon laquelle la contemplation et la compréhension de la nature rapprochent l'homme du divin Créateur. Mais le projet n'est pas pensé dans un seul but d'édification. Oberlin était non seulement pasteur, ayant charge d'âmes, mais pasteur « des Lumières », authentiquement curieux des développements de la pensée scientifique rationnelle, ce qui nous engage dans l'appréciation d'un double itinéraire, théologique et scientifique.

Au plan théologique, Oberlin fait sien le mot de Stuber affirmant qu'« autrefois », c'était le « Dieu des ténèbres », tandis qu'à présent « nous tâchons d'avoir le Dieu de lumière ». Mais chez Oberlin, la formule se soutient d'un incessant et pressant questionnement théologique, un « parcours initiatique », selon la formule de Loïc Chalmel, entre mysticisme et piétisme. Ses tendances mystiques, voire gnostiques, les formes ésotériques que prendront certaines de ses pratiques mystiques seront à l'origine d'une première rupture, symbolique en tout cas, avec son maître Stuber. Oberlin s'émancipe en s'inspirant de plus en plus du référentiel théologique de Zinzendorf représentant de la deuxième génération piétiste, dont la thèse de *ecclesiola in ecclesia* sera traduite au plan pédagogique par la mise en œuvre de la « petite école dans l'école ». Oberlin caractérisait volontiers sa profession de foi par l'appellation de « catholique évangélique », néologisme à comprendre comme le trait d'un esprit indépendant qui saura recueillir le nouvel idéal de tolérance et d'ouverture de la Révolution française. Si les petites écoles étaient à l'origine liées aux fondements religieux piétistes moraves, elles s'enrichissent progressivement des valeurs de l'humanisme républicain sans pour autant se séculariser. L'œuvre pédagogique d'Oberlin peut ainsi se lire comme un trait d'union entre piétisme morave et humanisme laïc.

Mais Oberlin est aussi pasteur « des Lumières » par son intérêt et sa curiosité exceptionnelles pour les sciences et toutes les découvertes scientifiques. Il n'y a pas pour lui contradiction entre science et théologie, conformément aux idées de l'*Aufklärer* Christian Wolff. Il y a une cohérence entre nature, raison et théologie. Lire, trier, classer, ordonner, comprendre, la quête de sens d'Oberlin n'a rien de la « recherche pure », ni du « dilettantisme scientiste ». Elle est ordonnée à deux fins : un projet évangélique qui place l'homme au cœur de la création, une volonté de vulgarisation au service de l'éducation des enfants, de la formation des adultes, d'une œuvre sociale.

Au plan pédagogique, Oberlin ne se contente pas plus de parachever l'œuvre de Stuber. Il est à l'origine du concept de « maison-école », véritable lieu de vie qui se distingue des autres habitations par l'existence d'une salle spécialisée pour rassembler les enfants (le « poêle d'école ») ainsi que de chambres pour les pensionnaires. À partir de 1778, la manière de légiférer pour fixer la méthode et les programmes des écoles du Ban de la Roche s'inspire des pratiques allemandes qu'Oberlin a eu l'occasion d'observer au cours de voyages en 1778 et 1780. La formation des adultes est assurée par des cours du soir. Toutefois, la grande originalité pédagogique d'Oberlin réside dans la création des « poêles à tricoter », destinés au jeune enfant, premier dispositif de préscolarisation au monde. Créés en 1770, ils sont le fruit d'une préoccupation et d'une rencontre : la préoccupation touche le sort réservé au jeune enfant ; la rencontre est celle d'une jeune villageoise, Sara Banzet qui, « de sa propre initiative, accorde en sa présence l'idée de l'enseignement du tricotage avec celle de l'éducation de la petite enfance » (p. 109). Enthousiasmé par l'initiative, Jean-Frédéric Oberlin la fait entrer comme institutrice à son service. Très vite, il recrute des conductrices à qui il confère un véritable statut social par contrat. Les « poêles à tricoter » ne sont pas des écoles de charité. Les conductrices sont des jeunes filles célibataires et non émancipées. Parmi toutes celles qui consacreront leur vie à la petite enfance, trois fortes personnalités ont marqué l'histoire : Sara Banzet, Anne-Catherine Gagnière et Louise Scheppler. Fait significatif si l'on songe à l'histoire de la préscolarisation aboutissant bien plus tardivement à la création des écoles maternelles, le destin de l'éducation préscolaire au Ban de la Roche a vu celle-ci devenir l'affaire exclusive de femmes protestantes.

La création des « poêles à tricoter » a suscité résistances et protestations de toutes origines. De fait, nous dit Loïc Chalmel, cette institution « se construit en rupture avec les conceptions théologiques, culturelles, politiques et sociales de l'époque ». On y voit la transgression de tabous dont la place de la femme n'est pas le moindre. Au plan théologique, c'est cette institution qui traduit le mieux le projet de *ecclesiola in ecclesia* qui consiste à concevoir une assemblée d'élus, élément moteur, source de régénération pour toute la communauté évangélique. Cela se traduit au plan pédagogique par la mise en œuvre des principes pédagogiques de Comenius : le « poêle à tricoter » est l'institution par excellence qui permet de mettre en œuvre une véritable pédagogie d'éveil dont le principe est d'occuper les mains pour ouvrir l'esprit. La leçon de tricot a rapidement fait place à toute une gamme d'activités, de jeux visant l'exercice physique en vue de développer l'activité intellectuelle.

L'œuvre sociale d'Oberlin se situe en cohérence avec sa vocation de pédagogue. Dans cette vallée déshéritée où les habitants doivent tout tirer d'eux-mêmes, Oberlin cherche à promouvoir une agriculture raisonnée plus efficace. Le projet repose sur quatre institutions : la bibliothèque de prêt, le magasin, une pépinière et la Société d'agriculture. Entraînement communautaire, mutualisation des savoirs, coéducation inspirent une dynamique dont la raison d'être est à comprendre comme un refus de la résignation et du fatalisme face aux caprices de la nature.

« Et plus bas, et plus haut ! », la formule quelque peu schématique en laquelle se résume la vocation du pasteur Oberlin est l'expression métaphorique de ses deux voies : « visionnaire du ciel et ouvrier de la terre ».

Laurence Loeffel
Université de Picardie-Jules Verne
CURAPP

GRANDIÈRE Marcel. *La formation des maîtres en France : 1792-1914*. Lyon : INRP, 2006. – 221 p. (Éducation, histoire, mémoire).

Les discours sur la formation des maîtres opposent volontiers deux traditions. Depuis la création des diverses agrégations au XIX^e siècle, on devient professeur en suivant un *curriculum* universitaire spécialisé, qui s'achève dans un concours vérifiant des savoirs universitaires, sous une forme académique très normée (dissertation, exposé, leçon magistrale). Dans la tradition primaire, à l'inverse, les futurs instituteurs sont recrutés par concours pour suivre une formation conjuguant les savoirs intellectuels et professionnels (les multiples « matières » au programme et l'entraînement à faire la classe). À la dimension pratique de la formation, caractéristique du primaire, la III^e République aurait ajouté la morale professionnelle laïque et l'esprit de corps qui ont fait la renommée des « hussards noirs » formés par l'école normale. Que penser de cette vulgate, régulièrement réactivée ? Les IUFM ont-ils « primarisé » la formation des professeurs secondaires – pour le meilleur (le stage en établissement) ou pour le pire (le « verbiage pédagogue ») – ou imposé à tous le modèle secondaire (*curriculum* universitaire et préparation au concours avant une année trop courte de « prise de fonction ») ? Quelle dynamique résultera de l'intégration aux universités ? S'installant dans l'histoire longue, une équipe de l'IUFM des Pays-de-la-Loire a collecté et édité les 227 textes officiels qui réglaient la formation des maîtres du primaire de la Révolution à la première guerre mondiale (INRP, à paraître). Marcel

Grandière qui a dirigé cette recherche publiée aujourd'hui un livre synthétique qui éclaire d'un nouveau jour les débats actuels sur la professionnalisation.

Les trois grandes étapes qu'il distingue enjambent les régimes politiques. La première étape commence par un échec, celui de l'école normale de l'An III, destinée à instruire à Paris ceux qui diffuseraient les « savoirs savants » dans les institutions des départements (ce seront les écoles centrales, destinées à remplacer les collèges d'Ancien Régime). C'est le premier stage national de formation de formateurs, en quelque sorte. Les plus grands savants du temps (Monge, Berthollet, Laplace, Lagrange, Daubenton...) font des conférences ou des cours, parfois brillants, mais les auditeurs, dépassés ou attendant autre chose, désertent en nombre. Au modèle descendant, centralisé, centré sur les savoirs, est opposé un modèle local, une formation pratique sur le terrain. Les débats de la Restauration tournent donc autour de deux grands modèles : les cours normaux forment les novices dans les classes des experts, l'école normale (qui reprend en partie le projet des écoles centrales) vise prioritairement à élever le niveau d'instruction des maîtres. Dans son rapport sur la Hollande, Cuvier décrit comment on a chargé les maîtres des écoles (publiques ou non) de recruter les bons élèves sur place, en leur offrant une scolarité gratuite prolongée, en échange d'un travail de sous-maître, sanctionné par divers certificats de capacité. Les Hollandais ont généralisé la formule employée en France par tous les ordres religieux enseignants, qu'ils s'adressent aux élites comme les jésuites ou les oratoriens, ou au peuple comme les Frères des écoles chrétiennes. On forme les novices en leur demandant d'assister un maître confirmé dans sa classe, tout en complétant leur formation intellectuelle. Pour mettre l'instruction populaire au service du progrès, il suffirait de garder ce dispositif de tutorat en changeant les outils (les manuels choisis par les autorités pour propager les bons contenus par de « bonnes méthodes »). En Hollande, les avantages liés à la fonction (revenus et statut) font affluer les candidats. En France, s'il y a assez de candidats instruits en ville, ce n'est pas le cas pour les campagnes où vivent la majorité des élèves. C'est finalement le modèle de l'école normale, inspiré de l'Allemagne, que Guizot impose en 1833. Chaque conseil général doit créer et gérer une école normale de garçons, dont le programme d'étude est défini par le ministère. En 1840, quand sortent les premiers bilans, le choix est si violemment contesté que certains demandent son abandon : entre 1840 et 1879, la formation par l'école normale demeurera l'objet de controverse.

Trois questions récurrentes traversent la période qui va de 1840 à 1879, à travers quatre régimes de gouvernement (la Monarchie de Juillet, la Révolution de 1848, le

Second Empire et la période de l'Ordre moral). Question politique des partages entre pouvoir local et pouvoir central : les écoles normales apparaissent aux autorités locales comme une intrusion étatique sur leur territoire et les congrégations enseignantes qui voient croître leur clientèle dénoncent la mainmise de l'État sur les choix éducatifs des familles. Question sociale de l'ordre établi : conçues pour former les maîtres à instruire et moraliser le peuple, afin d'éviter les désastres d'une nouvelle révolution, les écoles normales sont ressenties comme des foyers d'agitation politique, de mœurs relâchées, si bien que les normaliens mettraient en danger la paix sociale. Question pédagogique de la formation, enfin : les programmes des écoles normales sont jugés encyclopédiques et formels. Ils accouchent de pseudo-savants, ces « incapables prétentieux » qui au lieu de montrer l'exemple du dévouement au métier, de l'obéissance aux autorités et de la solidarité de corps, résistent aux tutelles locales aussi bien qu'aux injonctions du ministre, se sentent supérieurs aux deux autres tiers de maîtres, formés par les congrégations (un tiers) ou sur le tas (un tiers). Les écoles normales ne forment donc qu'un maître laïque sur deux et les conditions de travail sont si peu attractives qu'après leur sortie, beaucoup ne prennent pas de poste (plus du tiers des promotions dans l'Eure, d'après le rapport de 1847, au moins un maître sur huit, à l'échelle de la France).

Dans ces conditions, pourquoi garder de coûteuses écoles normales ? L'explosion de l'édition scolaire et de la presse pédagogique permet un foisonnement d'initiatives à tous les niveaux. Nombre d'enseignants qui ont passé le brevet en étudiant seuls, ou après avoir été sous-maître chez leur ancien instituteur en milieu rural, peuvent trouver dans cette diffusion massive de l'imprimé, des sources d'information qui nourrissent leur réflexion, permettent aux talents de se révéler, même sans école normale. En 1860, quand le ministre Rouland lance le concours « Quels sont les besoins de l'instruction primaire, dans une commune rurale, du triple point de vue de l'école, des élèves et du maître ? », 5 040 mémoires sont rédigés, 172 sont sélectionnés. Il y a alors 36 700 instituteurs publics : c'est dire qu'un sur sept a répondu. Si les écoles normales sont néanmoins maintenues, c'est donc bien pour des motifs plus politiques que fonctionnels. L'État ne peut pas abandonner le pouvoir conquis de haute lutte contre les autorités locales, en matière de recrutement et de contrôle des maîtres. Le corps des inspecteurs et des directeurs d'école normale constitue un relais direct des politiques ministérielles, dont il s'agit de renforcer la légitimité.

De 1879 à 1914, on pourrait croire que la question est réglée : la loi Ferry de 1881 confirme le pouvoir de gestion des autorités locales (chaque conseil général doit

créer une école normale de jeunes filles, si ce n'est déjà fait), et leur retire toute autorité sur les élèves-maîtres, mis sous tutelle du seul directeur d'école normale, puis de l'inspection qui gère les nominations et les changements de postes. Les écoles normales sont le cœur du dispositif d'enseignement public, d'autant que la création des deux écoles normales supérieures de Fontenay et de Saint-Cloud permet de recruter des professeurs pour les écoles normales parmi les élèves d'élite des écoles normales, sans faire appel aux professeurs des établissements secondaires avoisinants. La question des contenus jugés scientifiquement nécessaires et/ou pédagogiquement utiles reste cependant un point crucial, malgré la rhétorique optimiste des discours officiels. La disparition des aumôniers chargés de l'enseignement religieux et moral libère du temps pour les « sciences de l'éducation » (la psychologie de l'enfant, la sociologie de l'éducation, la pédagogie générale). Les recteurs et inspecteurs généraux (issus de l'université), la vieille génération d'inspecteurs ou d'instituteurs dénoncent aussitôt l'envahissement d'un jargon inutile, alors que l'emploi du temps peine à faire sa place aux matières secondaires (la musique, l'écriture, l'agriculture, l'éducation physique). Toute l'énergie des normaliens est absorbée par la préparation au brevet élémentaire en première année, au brevet supérieur les deux suivantes. Il est bien prévu qu'ils aillent vingt jours par an dans les classes, mais ces journées de stage ne comptent pas pour les examens. Rien d'étonnant si les maîtres d'application constatent un faible investissement (comme aujourd'hui, lorsque des professeurs des écoles en première année sont « parachutés » dans des classes au moment des concours blancs).

Dès 1886, la loi Goblet réforme : le concours d'entrée ne sera ouvert qu'aux élèves déjà titulaires du brevet, les écoles normales auront en charge la seule préparation du brevet supérieur. Cette décision est fonctionnelle (elle perdura jusqu'en 1940), mais elle ouvre une crise de recrutement immédiate : dès 1887, les effectifs de candidats baissent. Les bons élèves des villages ne peuvent plus préparer le brevet élémentaire dans l'internat gratuit de l'école normale, et puisque les brevetés peuvent directement être recrutés par les inspecteurs d'académie pour être mis devant les élèves sans autre préparation, pourquoi se fatiguer à passer le concours ? En 1911, comme en 1840, les écoles normales forment à peine plus d'un instituteur laïque sur deux et cette crise des vocations, jugée alarmante, fait baisser la qualité des lauréats au concours. C'est que les avantages symboliques proclamés par les républicains n'ont pas été suivis d'avantages matériels suffisants (nomination, traitement, retraite) au moment où d'autres débouchés s'ouvrent aux bons élèves des milieux modestes (écoles industrielles, commerciales ou agricoles). La suppression de l'enseignement congréganiste en

juillet 1904 oblige à des mesures drastiques pour accroître et améliorer le recrutement.

La réforme de 1905 cherche à répondre aux critiques contre des programmes, « une science hâtive de manuel, où se heurtent effroyablement les niaiseries de l'ancienne scolastique, les mensonges de la philosophie officielle et d'informes données scientifiques sans coordination », écrit Clémenceau en 1896. On croirait lire un réquisitoire « républicain » contre le « pédagogisme » des IUFM, dans ce portrait au vitriol des écoles normales de Jules Ferry et leur dogmatique philosophie d'État. La réforme de 1905 choisit de répartir en deux temps la formation intellectuelle (le brevet supérieur est placé en fin de 2^e année) et la formation pratique (avec au moins deux mois de stage en 3^e année, ce qui oblige à trouver des écoles en nombre). Mais que faire de ceux qui ont échoué au brevet supérieur ? Et comment mobiliser ceux qui viennent d'y réussir après un intense bachotage, avec trente heures de cours encadrés de temps d'études, lorsqu'ils passent en 3^e année avec cinq heures de cours par semaine, assortis de travaux libres et de lectures personnelles ? Ce temps libéré pour apprendre à « penser par soi-même », construire une culture professionnelle intégratrice est satisfaisant sur le papier, mais risque d'être vécue par les normaliens comme une simple pause entre les temps de stage. Un article de 1912 sur la réforme des écoles normales (Gasquet) conduit la logique à son terme : « C'est vers la libération entière des années d'école normale qu'il faut tendre. La préparation professionnelle ne doit pas être subordonnée à une épreuve qui n'a rien de professionnel ». D'autres plaident déjà pour l'intégration des normaliens aux classes de lycée (le député Massé) en vue du baccalauréat : pourquoi maintenir les barrières culturelles et sociales entre formation primaire et secondaire ? Les normaliens pourraient être boursiers et partager le même cursus intellectuel que les futures élites avant d'être formés au métier (projet que reprendront les Compagnons de l'Université après-guerre). Élie Reynier, militant syndical et professeur à l'école normale de Privas, dit très brutalement les dangers politiques d'un projet où il voit « la fin de l'école laïque » : « Le lycée, milieu bourgeois, imprégné de conservatisme social, où tant de jeunes agrégés crient haut leur haine du primaire, leur catholicisme, leur nationalisme [...], on en peut tout redouter pour la formation des futurs maîtres primaires ».

Conclusion : paradoxalement, ce qui caractérise la formation primaire tout au long du XIX^e siècle, c'est que le niveau d'instruction des maîtres ne cesse de s'élever, du fait des institutions et examens mis en place. À la veille de la Grande Guerre, s'il est pensable de former les maîtres du primaire à travers un curriculum secondaire, ce projet bute sur une question idéologique (la culture bourgeoise

et le conservatisme politique des élites) plutôt qu'intellectuelle (niveau trop élevé, savoirs inutiles). À l'inverse, la vulgate a retenu que la réussite de l'institution était du côté de la formation pratique. Le maître modèle était un « simple primaire », certes, mais on évoque avec nostalgie sa maîtrise des « bonnes méthodes » et sa morale professionnelle irréprochable. Or, à suivre Marcel Grandière, on voit que la « professionnalisation » des études, qui était la mission spécifique des écoles normales, n'a cessé d'être, tout au long du siècle, le talon d'Achille de l'institution. C'est que derrière les débats théoriques sur les façons les plus pertinentes de séparer/relier « les matières et les méthodes », de distinguer/intégrer la maîtrise des savoirs scolaires et celle de gestes professionnels, deux facteurs pèsent prioritairement.

Premier facteur, l'attractivité économique et sociale du métier. Le même dispositif aura une fortune différente selon la conjoncture. Les crises économiques sont des périodes bénies pour recruter des fonctionnaires, mais les périodes de croissance, des moments de reflux où le recours à des personnels temporaires, non formés, revient en force (sous la Monarchie de Juillet, la Belle Époque, comme pendant les Trente Glorieuses). L'autoformation fonctionne d'autant mieux que les salaires attendus stimulent la concurrence : dans la France du XIX^e, les avantages symboliques et juridiques (la tutelle lointaine de l'État plutôt la tutelle proche des mairies) s'accordent plus aisément que les avantages matériels. Ceci est d'autant plus sensible que les coûts en personnels sont progressivement transférés des communes (salaires, logements, avantages en nature) et des familles (écolage) au département, puis à l'État, pour assurer l'égalité des traitements. Tout avantage accordé par l'État se répercute sur l'ensemble national, mais on sait qu'il y a d'autres façons si on est une commune riche, d'avantager localement une école comme le montrent les écarts actuels entre écoles maternelles, ou d'avantager certains maîtres (avec des compléments de salaire comme les services d'études ou de cantine). Bien des établissements (écoles primaires supérieures, écoles de l'industrie ou du commerce, écoles de dessin) sous tutelle municipale, ont constitué une marge non négligeable de pouvoir, expliquant l'attachement des maires à des établissements confessionnels. Or ces rivalités entre pouvoir local et pouvoir central ont des incidences indirectes mais majeures, sur les politiques de recrutement et de formation.

Car le second facteur est clairement politique : au XIX^e siècle, les luttes autour du pilotage de la formation voient la montée en puissance du pouvoir central contre tous les pouvoirs locaux plus ou moins coalisés, où se retrouvent les municipalités, les familles et les Églises. Or, si la formation intellectuelle peut être aisément instituée

de façon centralisée et évaluée dans des examens classiques, la formation pratique relève bien plus des ressources locales, exige d'autres types de « professionnels » et d'autres modes d'évaluation. On n'apprend pas à faire la classe de la même façon dans une classe de ville ou de village (en centre-ville ou en ZEP, dirions-nous aujourd'hui). Pour enlever à l'Église son pouvoir sur l'école et la formation des maîtres, il fallait faire prévaloir une « formation d'État », qui puisse qualifier intellectuellement les maîtres à l'égal des curés et leur permette de devenir des secrétaires de mairie appréciés des collectivités locales. Pour ce faire, la question des diplômes (brevets élémentaire, supérieur) est autrement sensible que celle des méthodes de lecture qui, à l'époque, ne soucient aucun parent d'élève. Qu'en est-il aujourd'hui au temps de la décentralisation/déconcentration ? Comment gérer la concurrence potentielle entre plusieurs tutelles de formation, l'État, les universités et les régions ? La concurrence existe déjà entre la collation des grades qui est un monopole universitaire relevant de conventions européennes (le LMD) et les concours de recrutement qui relèvent de l'employeur (le ministre en France, la région ou la ville dans d'autres pays d'Europe). Prochain épisode à suivre, donc, avec l'intégration des IUFM aux universités et la création des « mastères professionnels aux métiers de l'enseignement » que cette manne d'étudiants ne va pas manquer de susciter.

Anne-Marie Chartier
INRP

MALET Régis & BRISARD Estelle (dir.). *Modernisation de l'école et contextes culturels : des politiques aux pratiques en France et en Grande-Bretagne* / Préface d'Agnès van Zanten. Paris : L'Harmattan, 2005. – 277 p. (Éducation comparée).

En France, les comparaisons internationales demeurent un exercice marginal alors qu'elles constituent un véritable champ de recherche outre-Manche et qu'elles s'appuient sur des revues scientifiques de qualité. C'est le mérite du livre de Régis Malet et d'Estelle Brisard de nous rappeler que nul n'est prophète en son pays et que la méthode comparative est aussi, comme l'écrivait Émile Durkheim, au fondement de l'administration de la preuve. L'ouvrage prend comme fil conducteur les transformations du métier d'enseignant pour interroger les effets de convergence des politiques éducatives en France et en Angleterre et les spécificités nationales ou locales, notamment les contextes culturels, qui résistent à l'emprise de la globalisation. Si les systèmes éducatifs français et anglais sont soumis à un même impératif de rationalisation, selon

des exigences renforcées d'efficacité et de rentabilité, les expériences sociales des enseignants y sont très différentes. Ces différences, minutieusement analysées, montrent aussi que l'adaptation de la profession enseignante à la diversification des publics scolaires et à l'enseignement de masse présente des configurations spécifiques dont une approche macrosociologique ne peut à elle seule rendre compte. D'où l'intérêt d'une démarche comparative dont les principes, décrits par Marie-Pierre Moreau, relèvent de choix méthodologiques et théoriques qui accordent une place essentielle aux contextes locaux et à la dynamique des acteurs mais respectent aussi un principe de précaution en termes d'équivalence linguistique ou conceptuelle.

Une première série d'articles entreprend de saisir les dynamiques transnationales qui, en dépit du développement d'un champ de références internationales et des transformations de l'État éducateur, interagissent avec des traditions locales, des formes d'administration scolaire, et des modalités de formation des enseignants, en opérant une médiation sociale et culturelle propre à chaque pays. Ainsi, comme l'affirme Jenny Ozga, si des politiques « nomades » tendent à définir un certain nombre d'objectifs au niveau supranational en termes de modernisation des politiques d'éducation, selon des rhétoriques et des stratégies cohérentes visant à promouvoir un agenda néo-libéral, les logiques d'appropriation de ces référentiels par les acteurs locaux témoignent de politiques « enracinées », comme l'illustre l'exemple écossais, dans lesquelles se maintiennent des récits collectifs attentifs au maintien d'une identité nationale et à la préservation d'un bien commun. La mise en parallèle par Régis Malet de l'histoire des systèmes d'enseignement fournit une autre direction pour penser les évolutions de l'administration scolaire en Angleterre et en France. Si les deux pays semblent avoir vécu des trajectoires opposées, centralisation du curriculum et de l'évaluation dans un pays à forte tradition locale pour l'Angleterre, lois de décentralisation déléguant les compétences de l'État aux collectivités territoriales dans le cas de la France, des traits similaires peuvent être identifiés : autonomisation des établissements, mise en œuvre de régulations intermédiaires, développements de l'évaluation. Toutefois, cette convergence formelle des systèmes éducatifs masque des différences en matière de construction des savoirs scientifiques. Alors que la recherche anglaise hésitait entre un accompagnement du management scolaire proche de l'expertise et un travail critique des effets négatifs des réformes néo-libérales, l'étude du travail et de la professionnalisation des enseignants en France accompagnait la découverte du local et l'ouverture de la boîte noire de l'établissement scolaire.

C'est de cette tradition que se réclament indirectement les articles d'Anne Barrère et de Aziz Jellab qui livrent ici une analyse comparée des expériences du travail enseignant dans le contexte de l'enseignement secondaire français. Pour Anne Barrère, c'est davantage la façon dont les enseignants traversent les épreuves propres à leur activité professionnelle qui expliquent les différences entre les situations vécues plutôt que les caractéristiques des établissements scolaires selon qu'ils accueillent des élèves difficiles ou non. C'est pourquoi, selon le deuil qu'ils font de leur discipline, la nature de la confrontation quotidienne avec les élèves, leur accommodement aux formes de l'évaluation, et le degré de reconnaissance dont ils bénéficient, les enseignants ressentiront avec plus ou moins de force les contraintes de leur environnement. De même, comme le décrit Aziz Jellab pour les lycées professionnels, les éléments constitutifs du travail enseignant varient selon le rapport au savoir que privilégient les professeurs et le type de relations qu'ils adoptent avec leurs élèves. Alors que le profil social et professionnel des enseignants de lycées professionnels a évolué, et que les élèves entretiennent des rapports de plus en plus distancés avec la culture ouvrière, différentes postures pédagogiques se juxtaposent selon la place accordée à l'accompagnement des élèves et à la transmission des savoirs. Toutefois, ces postures individuelles s'inscrivent de plus en plus dans des configurations plurielles du travail collectif qui résultent de l'émergence d'une forme plus libérale de l'investissement pédagogique autour du projet et d'une nouvelle distribution des tâches dans l'établissement scolaire. Marie-Christine Lefloch décrit ainsi la multiplicité des configurations qui caractérisent cette coordination de l'action à plusieurs, depuis les couples professionnels ou les triades d'enseignants rassemblés autour d'un projet jusqu'aux réseaux plus institutionnels pour lesquels les liens de proximité sont plus distendus.

Les formes d'organisation du travail des enseignants sont malgré tout assez différentes en France et en Angleterre. Côté français, le repli de l'enseignant sur la salle de classe et l'absence d'injonction au travail collectif témoignent d'une norme professionnelle où la revendication de l'autonomie pédagogique est corollaire de l'appartenance à une communauté sociale limitée entre pairs. Dans le système éducatif anglais, une prescription managériale beaucoup plus forte s'appuie sur différentes formes de leadership et un collectif de travail beaucoup plus structurant dans les pratiques pédagogiques. L'inscription du travail enseignant dans une éthique communautaire s'entrecroise avec une culture organisationnelle propice à la mise en œuvre d'un professionnalisme collectif soucieux d'une mise en commun à des fins d'efficacité. Ces divergences permettent de mieux comprendre la manière dont les enseignants de chaque

pays s'approprient les réformes scolaires et définissent leur propre travail. À partir de plusieurs comparaisons internationales entre la France et l'Angleterre, Marilyn Osborn et Elizabeth Mc Ness ont pu démontrer que la mise en œuvre des politiques éducatives faisaient l'objet de médiations sociales et culturelles importantes et qu'il était nécessaire de prendre en compte les attitudes complexes des enseignants confrontés au changement. La conception de l'étendue des responsabilités éducatives varie sensiblement, les enseignants anglais oscillant entre acceptation professionnelle du changement et sentiment de démoralisation face aux pressions du management, alors que les enseignants français témoignent d'un profond scepticisme et d'une lassitude beaucoup plus grande vis-à-vis des réformes. Ces différences se font ressentir jusque dans l'enseignement des mathématiques, qu'elles concernent la gestion du groupe-classe, les routines pédagogiques, l'usage du manuel scolaire, comme en témoigne l'article de Birgit Pepin. Au travers d'une ethnographie très précise inspirée du travail de Marcel Mauss, Maroussia Raveau montre que ces variations sont aussi perceptibles dans la mise en ordre des corps et les différentes postures adoptées par les élèves à l'école primaire.

L'ensemble de ces contributions, fort bien documentées et riche d'exemples, offre au lecteur l'occasion d'un décentrement et d'une réflexion approfondie sur les changements à l'œuvre dans la profession enseignante à l'aune des réformes éducatives propres à chaque pays. Elles montrent bien l'intrication et l'hybridation de processus complexes qui, du local au global, contribuent à un déplacement des modes d'organisation du travail et de la gestion quotidienne des établissements scolaires. Sans doute aurait-on pu attendre de la part de Régis Malet et d'Estelle Brisard une réflexion plus distanciée sur les notions de modernisation, d'organisation apprenante, de gouvernance qui sont autant d'éléments d'une rhétorique managériale que des concepts proprement sociologiques. De même, dans un souci de comparaison, il aurait été opportun de souligner davantage côté anglais le contexte et les effets des politiques de « *shame and blame* » (honte et blâme) de la profession enseignante soumise depuis deux décennies à la pression d'un État managérial très différent d'un modèle bureaucratique contrairement à ce qu'il est écrit dans le livre. Enfin, même si l'ouvrage ne pouvait pas tout traiter, il est dommage qu'aucun article ne décrive la transformation du rôle des corps d'inspection en Angleterre qui ont pourtant largement contribué à la mise en œuvre et à l'accompagnement des réformes dans un sens néo-libéral, tandis que leurs fonctions demeuraient plus traditionnelles en France. Mais, au-delà de ses faiblesses toute relatives, le livre constitue un exercice réussi d'analyse comparée qui, par l'étendue de sa

bibliographie et de ses études de cas, offre une mise en perspective intéressante et un cadre stimulant de réflexion pour l'expert comme pour le néophyte.

Romuald Normand
INRP
UMR Éducation & Politiques

RONDAL Jean-Adolphe. *Expliquer l'acquisition du langage : caveats et perspectives*. Sprimont [Belgique] : P. Mardaga, 2006. – 234 p. (Psychologie et sciences humaines).

Le livre de Rondal se compose de trois chapitres : le premier est consacré à un examen critique des apports de la linguistique générative d'inspiration chomskyenne à la possibilité d'une psycholinguistique. Le second présente une revue des perspectives développementales (seulement en ce qui concerne le langage oral) et le troisième est une synthèse sur le thème de « fonctionnement langagier et ontogenèse ».

L'ouvrage se caractérise, me semble-t-il, d'abord par des exigences expérimentales fortes en ce qui concerne ce qui pourrait être une explication : « le terme explication est à prendre au sens strict en sciences inductives, c'est-à-dire la définition d'un ensemble de variables nécessaires et suffisantes pour rendre compte du phénomène » (p. 9), même si, de fait, on reste souvent loin d'un tel ensemble nécessaire et suffisant. Ce sont donc avant tout des données de type expérimental qui sont prises en compte, avec une riche connaissance de la littérature mondiale, ce qu'atteste la bibliographie qui fait de ce livre un ouvrage de référence.

Par ailleurs, dans tout l'ouvrage, Rondal adopte une perspective modulaire, c'est-à-dire fondée sur l'interaction de capacités relativement indépendantes, indépendance qui est attestée surtout par des exemples de dissociations dues à la pathologie, dissociations qui illustrent en particulier la séparation éventuelle des compétences linguistiques au sens étroit (phonétique ou phonologie, syntaxe et lexique) et des capacités langagières à mettre en œuvre dans des conduites complexes de communication de ces savoirs linguistiques spécifiques.

Il est difficile de « compter » les modules qui doivent être ainsi supposés. Au moins, nous dit Rondal, (p. 8) un niveau de régulations pragmatiques, une organisation discursive « des informations à échanger », les étages morphosyntaxiques, morpholexicaux et phonologique.

Commencer par un examen de la théorie chomskyenne vise à mettre en tête de ses préoccupations le problème de

l'« apprenabilité » du langage ainsi que d'une base neurologique rendant compte de cette apprenabilité : le point de vue développé par Rondal par opposition au « chomskysme simple » (ou « dogmatique » ?) de Pinker vise à reconnaître des bases neurologiques multiples rendant possible l'acquisition et non l'innéité de représentations et en particulier de représentations de type syntaxique considérées comme le noyau central de la langue.

On ne revient pas ici sur les évolutions de la théorie chomskyenne que Rondal retrace.

On peut rappeler avec lui d'abord que l'universalité ne signifie pas innéité.

D'autre part, la difficulté qu'il y a à décrire des relations syntaxiques en dehors de relations de type sémantique et pragmatique (par exemple posé-ajouté) particulièrement nettes dans le dialogue. Rondal semble plutôt, dans des perspectives proches de celle de Schlesinger ou de Langacker (quelles que soient leurs différences) s'interroger sur ce que pourrait être une centralité de la « sémantique » entre le cognitif, le pragmatique et l'expression linguistique dans ses formes attestées. Il s'agit en tout cas de mettre en relation des niveaux (modules) indépendants et partiellement liés, non de supposer d'abord l'autonomie et la prévalence d'un niveau syntaxique. Comme on l'a dit plus haut, Rondal insiste sur le fait que les diverses pathologies mettent en évidence une relative indépendance des modules en question (en particulier entre capacité à utiliser des formes grammaticales correctes, ce qui peut être disjoint de la capacité à mettre en place une communication pratique ou narrative intelligible). Par ailleurs, on peut noter que les relations sont complexes entre production, compréhension, paraphrase, articulation avec d'autres systèmes de signes (ce dernier problème est peu abordé dans le texte). De même qu'entre genres de discours, dialogue et monologue. Tout cela fait que la question de la caractérisation des modules en question est plus devant nous que « derrière nous ». En tout cas, cette perspective modulaire suppose des zones cérébrales relativement séparées conditionnant telle acquisition et non un innéisme des représentations syntaxiques en particulier.

Le chapitre II (en dimensions le plus important : une centaine de pages) apporte beaucoup de données empiriques. Ainsi en ce qui concerne la relative indépendance des acquisitions de la langue au sens strict et des capacités cognitives au sens large ou sur la facilitation de l'acquisition de la langue par les modifications de la langue des adultes (ou des autres jeunes enfants) adressée aux plus jeunes.

En passant, Rondal discute les assertions de Bruner concernant les relations entre LAD (*Language Acquisition*

Device) et LASS (*Language Acquisition Support System*) puis LAS, et en fonction de la complexité de leurs relations et aussi du fait, qu'au moins temporairement, le modèle de Bruner semblait supposer la validité, locale, de celui de Chomsky.

Sur la question difficile de l'interaction entre les modules, l'exposé est complexe et reste ouvert. En tout cas, comme souvent, les argumentations de Pinker semblent trop rapides (p. 126). Ainsi, dire que les enfants atteints d'un syndrome de Williams ont un langage sans troubles de l'intelligence est inexact ; ils ont plutôt le langage (évalué grammaticalement) d'enfants de 5-6 ans d'âge mental). Leur langage est moins atteint que d'autres capacités cognitives, ce qui peut s'expliquer en termes d'atteintes neurologiques partielles et non d'une « capacité langagière innée » qui serait gardée.

La question des grammaires effectives par rapport aux grammaires normées reste à l'horizon, de même que celle de l'articulation possible des traits universels et de ceux qui caractérisent telle langue ou famille de langues. Ici aussi la problématique est ouverte. En tout cas, il semble bien que la « grammaire universelle » proposée par Chomsky ressemble un peu trop à celle de l'anglais. Mais cette question aussi est plutôt évoquée que traitée.

Rondal introduit à la fin du chapitre II une longue discussion sur ce que peut être la mise en évidence de facteurs héréditaires et en particulier la différence entre l'innéité de capacités à apprendre qui ne saurait être identique à l'innéité des capacités de représentations innées au sens chomskien. Il semble préférable de supposer un cerveau précocement apte à traiter les matériaux linguistiques, une capacité mise en place dans la vie intra utérine à reconnaître les schémas de la langue maternelle, en fonction de propriétés rythmiques, non une « grammaire universelle innée ».

Le dernier chapitre apporte quelques éclaircissements sur la question compliquée de l'articulation entre l'utilisation des structures et la prise de conscience de ces structures dans une grammaire explicite. Peu de données certaines ici. Il semble assuré que l'enfant (comme l'adulte) peut s'exprimer « correctement » sans conscience claire des structures grammaticales ni des procédures de segmentation (nécessaires à l'écrit). Quant à la nature d'une grammaire explicite facilitant l'acquisition et/ou le passage à l'écrit, elle reste objet de discussion (sans oublier la question de l'intelligibilité même des notions grammaticales en question). Tout comme la question de cette grammaire explicite comme projetant plus ou moins sur l'oral des organisations de certains types de messages écrits. Quant à la question de savoir si un objet « grammaire » est séparable de l'ensemble des conditions d'utilisation du

langage ou si l'on doit enseigner une grammaire de telle langue ou de tels groupes de langue, elle n'est pas résolue. En tout cas, les supports cérébraux des manèges implicites et explicites ne se recouvrent pas et on ne peut supposer une transformation « naturelle » des connaissances implicites en connaissances explicites.

Comme il est inévitable, un « caveat » pour reprendre le mot de l'auteur (« mise en garde » pour ceux qui ignorent le latin) s'impose : aucun ouvrage ne peut traiter de l'ensemble de ce que serait « Expliquer l'acquisition du langage ». La question des relations du langage aux autres sémiologies, au langage du corps est pratiquement absente. L'auteur parle beaucoup plus de la capacité à produire des énoncés qu'à celle de les comprendre. De même la question des genres de discours n'est pas abordée, pas plus que celle de l'articulation du thème du discours, de la nature des interlocuteurs et des finalités du discours. Que la pathologie puisse mettre en évidence des dissociations est un fait. Cela n'éclaire pas pour autant l'interaction des « modules » dans l'acquisition ordinaire.

L'ouvrage concerne plus les premières acquisitions que la suite. On peut dire que dans la recherche portant sur ce que peut être qu'« expliquer » le langage, Rondal met surtout l'accent sur ce qui rend neurologiquement possible cette acquisition. S'il ajoute des considérations sur ce qui la rend réelle dans le dialogue ou dans la simplification du langage adulte adressé, bien d'autres aspects de ces relations factuelles, forcément variables, comme le lien du langage aux autres pratiques sémiotiques, les variations effectives du recours au langage, la multiplicité des modes de dialogue, comme du langage solitaire ou du dire à plusieurs (par exemple dans le chant) ne sont pas développés.

Ou encore, l'ouvrage n'éclaire pas beaucoup les questions de l'articulation entre l'acquisition en milieu familial et celle dans les différents milieux scolaires (ce n'était pas son objet). Tout ce qu'on peut dire est que vers 2-3 ans la fréquence dans le discours parental est le meilleur prédicteur du développement de l'enfant.

Par rapport aux problèmes proprement pédagogiques, on pourrait dire que l'apport de ce livre est triple.

D'abord la richesse des informations que nous donne cet ouvrage sur les approches expérimentales du premier développement du langage.

Ensuite, les « caveats ». Rondal nous aide à nous méfier d'un modèle de la « langue déjà dans le sujet » comme préalable absolu à la représentation du langage entre les humains. En particulier il met en évidence les failles du « syntaxocentrisme » chomskien et la lourdeur inutile des hypothèses innéistes.

Enfin, le développement d'un modèle fondé sur l'interaction de « modules indépendants » est bien venu pour étayer une réflexion pédagogique sur l'interaction des niveaux et non la référence à l'image simpliste de la simple hiérarchie des étages, qui commencerait par la phonologie du mot, pour aller à la syntaxe, puis enfin de là au texte. On peut proposer que l'approche modulaire (quel que soit le contenu exact que l'on donne aux modules) correspond bien aux problèmes scolaires : ainsi il y a intrication, non lien direct entre la capacité à développer une argumentation ou à raconter et celle à manier « correctement » la syntaxe orale. Ces remarques peuvent s'étayer indirectement sur la prise en compte de la diversité des zones cérébrales mobilisées dans la mise en œuvre du langage, en particulier d'une beaucoup plus grande par-

ticipation des deux hémisphères qu'on ne le pensait il y a quelques années.

Dans la même perspective, Rondal éclaire la complexité du problème de la variété des types de « prises de conscience » utiles ou nécessaires au développement langagier. Beaucoup de questions restent devant nous sur ce sujet comme en ce qui concerne l'articulation entre logique du langage et logique des systèmes mathématiques.

Qui s'en plaindrait ? En tout cas, sur beaucoup de points, Rondal déblaye la voie. Avec compétence et clarté. Qu'il en soit remercié.

Frédéric François
Université Paris 5-René Descartes

La revue a reçu...

- ARANGO Diana Soto & JARDILINO José Rubens Lima (dir.). *Políticas universitarias en latinoamérica : historia y perspectiva*. São Paulo : Emblema, 2006. – 242 p. (Historia de la universidad latinoamericana ; n° 3).
- Agora : débats/jeunesse*, 2006, n° 41 : « Jeunes, genre et société », dossier coordonné par Sylvia Faure & Chantal de Linarès.
- Aster : recherches en didactique des sciences expérimentales*, 2006, n° 42 : « Le corps humain dans l'éducation scientifique », dossier coordonné par Patricia Marzin.
- BARRÉ-DE MINAC Catherine & REUTER Yves (dir.). *Apprendre à écrire au collège dans les différentes disciplines*. Lyon : INRP, 2006. – 268 p. (Documents et travaux de recherche en éducation ; 54).
- BARUS-MICHEL Jacqueline ; ENRIQUEZ Eugène & LÉVY André (dir.) ; avec la collab. de Jean-Marc Huguët. *Vocabulaire de psychosociologie : positions et références*. Ramonville-Saint-Agne : Érès, 2006. – 590 p.
- BERTHELOT Jocelyn. *Une école pour le monde, une école pour tout le monde : l'éducation québécoise dans le contexte de la mondialisation*. Montréal : VLB. – 219 p.
- BIDART Claire (dir.). *Devenir adulte aujourd'hui : perspective internationales*. Paris : L'Harmattan, 2006. – 231 p. (Débats-jeunesse).
- BISSONNETTE Steve ; MARIO Richard & GAUTHIER Clermont. *Comment enseigne t-on dans les écoles efficaces ? Efficacité des écoles et des réformes*. Laval [Québec] : Presses de l'université Laval, 2006. – 163 p. (Formation et profession).
- BODÉ Gérard & LE BUHAN Pierre. *Les établissements d'enseignement technique en France, 1789-1940 : t. 3, les Côtes-d'Armor*. Lyon : INRP, 2006. – 439 p. (Service d'histoire de l'éducation).
- BOYER Myriam (dir.) & LEGRAND Jérôme (collab.). *Le goût des sciences/préface de Yves Gaułupeau ; introduction de Fabien Locher*. Rouen : INRP : Musée national de l'Éducation, 2006. – 121 p. ; illus. en n. & b. et en coul.
- CAUMEIL Jean-Guy (dir.). *Fonder les savoirs scolaires : épistémo-anthropologie et médiation pour un regard nouveau sur l'éducation*. Lyon : SCÉREN-CRDP de l'académie de Lyon ; IUFM de l'académie de Lyon, 2006. – 169 p. (Savoirs en pratiques).
- CHARMASSON Thérèse (dir.). *Histoire de l'enseignement XIX^e-XX^e siècles : guide du chercheur*. Paris : CTHS ; INRP : SHE. – 728 p. (Orientations et méthodes ; n° 9).
- CHARLIER Jean-Émile & MOENS Frédéric (dir.). *Observer, décrire, interpréter : un état des méthodes en sciences sociales*. Lyon : INRP, 2006. – 239 p. (Éducation, politiques, sociétés).
- CHERVEL André. *Histoire de l'enseignement du français du XVII^e au XX^e siècle*. Paris : Retz, 2006. – 831 p. (Les usuels Retz).
- COLOMB Jacques (dir.). *Articulation école/collège en mathématiques : ruptures et continuités*. Lyon : INRP, 2006. – 232 p. (Documents et travaux de recherche en éducation ; 55).
- DELAHAYE Jean-Paul. *Le collège unique, pour quoi faire ? Les élèves en difficulté au cœur de la question*. Paris : Retz, 2006. – 157 p. (Défis d'éducation).
- DENIS Daniel & KAHN Pierre (éd.). *L'école de la Troisième République en questions : débats et controverses dans le Dictionnaire de pédagogie de Ferdinand Buisson*. Bern : P. Lang, 2006. – VIII-283 p. (Exploration, éducation : Histoire et pensée).
- DESMEUZES-BALLAND Sylvette. *Mères-fils : une relation malmenée/préface de Samuel Lepastier*. Paris : Éd. de l'Archipel, 2007. – 301 p.
- DESSUS Philippe & GENTAZ Édouard (dir.). *Apprentis-sages et enseignement : sciences cognitives et éducation*. Paris : Dunod, 2006. – 247 p. (Psycho sup).

- DUBUS Alain. *La notation des élèves : comment utiliser la docimologie pour une évaluation raisonnée*. Paris : A. Colin, 2006. – 270 p. (E : élève-école-enseignement).
- DUCANCEL Gilbert (coord.). *Sens et code au cycle 2 : apprentissages progressifs de l'écrit*. Paris : Hachette, 2006. – 303 p. (Profession enseignant)
- DUCOING Patricia (dir.). *La recherche en éducation au Mexique : état des travaux*. Paris : L'Harmattan, 2006. – 264 p. (Association francophone internationale de recherche scientifique en éducation [AFIRSE]).
- Échanger : valorisation de pratiques innovantes*, 2006, n° 77 : « L'action culturelle en LP »
- European Journal of Psychology of Education : A Journal of Education and Development*, 2006, vol. XXI, n° 3 : « Inclusive Education Ten Years After Salamanca », coordonné par Margarida César (Lisbonne) & Mel Ainscow (Manchester, UK).
- FIGARI Gérard & MOTTIER-LOPEZ Lucie (éd.). *Recherche sur l'évaluation en éducation : problématiques, méthodologie, épistémologie*. Paris : L'Harmattan, 2006. – 257 p. (Évaluer).
- GALAND Benoît & BOURGEOIS Étienne (dir.). *(Se) Motiver à apprendre*. Paris : PUF, 2006. – 234 p. (Apprendre).
- GANASCIA Jean-Gabriel. *Les Sciences cognitives*. Paris : Éd. du Pommier, 2006. – 186 p. (Poche-Le Pommier).
- GAUTHIER Roger-François. *Les contenus de l'enseignement secondaire dans le monde : état des lieux et choix stratégiques*. Paris : UNESCO, 2006. – 140 p.
- GELIN Rodolphe. *Le robot, ami ou ennemi ?* Paris : Éd. du Pommier, 2006. – 62 p. (Les petites pommes du savoir).
- HASNI Abdelkrim ; LENOIR Yves & LEBEAUME Joël (dir.). *La formation à l'enseignement des sciences et des technologies au secondaire dans le contexte des réformes par compétences*. Québec : Presses de l'université du Québec, 2006. – 263 p. (Éducation-recherche).
- HOFFMANS-GOSSET Marie-Agnès. *Apprendre l'autonomie, apprendre la socialisation*/préface de Louis Porcher. Lyon : Éd. de la Chronique sociale, 2006. – 163 p. (Pédagogie/formation).
- INSTITUT NATIONAL DE RECHERCHE PÉDAGOGIQUE : CENTRE ALAIN SAVARY [CENTRE NATIONAL DE RESSOURCES SUR LES PRATIQUES ÉDUCATIVES ET SOCIALES EN « MILIEUX DIFFICILES »]. *Apprendre et enseigner en « milieux difficiles » : sélection d'articles du bulletin XYZep*. Lyon : INRP, 2006. – 191 p.
- International Review of Education*, 2006, vol. 52, n° 5 : « Varia ».
- JACQUET Denis ; ZABALIA Marc & LE HALLE Henri (dir.). *Adolescences d'aujourd'hui : textes issus du XXIX^e symposium de l'Association de psychologie scientifique de langue française (APSLF)*. Rennes : Presses universitaires de Rennes, 2006. – 163 p. (Psychologie).
- L'Orientation scolaire et professionnelle*, 2006, vol. 35, n° 2 (numéro spécial) : « Adolescences-2 », coordonné par B. Dumora & F. Bariaud.
- L'Orientation scolaire et professionnelle*, 2006, vol. 35, n° 3.
- LANI-BAYLE Martine & MALLET Marie-Anne (coord.). *Événements et formation de la personne : écarts internationaux et intergénérationnels : t. 1. 2003-2005 ; t. 2 2005-2006*. Paris : L'Harmattan, 2006. – vol. 1, 256 p. ; vol. 2, 367 p. (Histoire de vie et formation).
- LANOUE Éric. *L'école catholique en Côte d'Ivoire 1945-2005 : politique, religion et fait scolaire en Afrique*/préface de Claudine Vidal. Pessac [Gironde] : Éd. de la Maison des sciences de l'homme d'Aquitaine, 2006. – 401 p.
- LE GOAZIOU Véronique. *Lecteurs précaires : des jeunes exclus de la lecture ?* Paris : L'Harmattan, 2006. – 198 p. (Débats-jeunesse).
- LENOIR Yves ; XYPAS Constantin & JAMET Christian (dir.). *École et citoyenneté : un défi multiculturel*/préface de Dominique Schnapper. Paris : A. Colin, 2006. – 264 p. (Sociétales).
- LETHIERRY Hugues. *Des conflits à l'école : des rixes du métier*. Lyon : Éd. de la Chronique sociale, 2006. – 175 p. (Pédagogie/formation).
- LETHIERRY Hugues. *Écrire la correspondance : (é)loge de la lettre*/avant-propos de Michel Ménaché. Lyon : Éd. de la Chronique sociale, 2006. – 203 p. (Savoir communiquer).
- LUC Jean-Noël & NICOLAS Gilbert. *Le temps de l'école de la maternelle au lycée 1880-1960*. Paris : Éd.

- du Chêne, 2006. – 310 p., illus. en noir & blanc (Gens de France).
- MAROY Christian & le réseau « REGULEDOC ». *École, régulation et marché : une comparaison de six espaces scolaires locaux en Europe*. Paris : PUF, 2006. – 390 p. (Éducation & société).
- MIALARET Gaston. *Sciences de l'éducation : aspects historiques, problèmes épistémologiques*. Paris : PUF, 2006. – 287 p. (Quadrige ; grands textes).
- MILMEISTER Marianne & WILLIAMSON Howard (dir.). *Dialogues et réseaux : organiser les échanges entre les acteurs du secteur jeunesse [version française]*. Esch-sur-Alzette [Luxembourg] : Éd. Phi ; Centre d'études sur la situation des jeunes en Europe (CESIJE) ; université du Luxembourg, 2006. – 240 p. (Monographies de la recherche jeunesse ; série Scientiphic, n° 2).
- NARCY-COMBES Marie-Françoise. *La communication interculturelle en anglais des affaires à l'usage des étudiants francophones*. Paris : PUR, 2006. – 171 p. (Didact ; Anglais).
- NORDMANN Charlotte. *Bourdieu/Rancière : la politique entre sociologie et philosophie*. Paris : Éd. Amsterdam, 2006. – 179 p.
- PAGET Denis (dir.). *L'école : aventure commune et savoirs partagés*. Paris : Institut de recherches de la FSU ; Syllepse, 2006. – 185 p. (Nouveaux regards).
- PASA Laurence ; RAGANO Serge & FIJALKOW Jacques (éd.). *Entrer dans l'écrit avec la littérature de jeunesse*. Issy-les-Moulineaux : ESF, 2006. – 174 p. (Pédagogies ; recherche).
- POCHON Luc-Olivier ; BRUILLARD Éric & MARÉCHAL Anne (éd.). *Apprendre (avec) les progiciels : entre apprentissages scolaires et pratiques professionnelles*. Neuchâtel : IRDP ; Lyon : INRP, 2006. – 377 p.
- Quaderni degli annali dell'istruzione*, 2005, n° 112 : « Progettot ESaT [European Science and Technology] : Progetto pilota europeo per il miglioramento dell'offerta formativa nel Settore Scientifico e Tecnologico ».
- RAVEAUD Maroussia. *De l'enfant au citoyen : la construction de la citoyenneté à l'école en France et en Angleterre*. Paris : PUF, 2006. – 209 p. (Éducation & société).
- Recherche et formation pour les professions de l'éducation*, 2006, n° 52 : « De l'éthique professionnelle en formation et en recherche », dossier coordonné par Laurence Gavarini & Dominique Ottavi.
- Repères : recherches en didactique du français langue maternelle*, 2006, n° 34 : « L'écriture de soi et l'école », dossier coordonné par Marie-France Bishop & Marie-Claude Penloup.
- Revista iberoamericana de Educación/de Educação*, 2006, n° 40 : « Educación para el desarrollo sostenible 1/Educação para o desenvolvimento sustentável 1. »
- Revue du MAUSS (Mouvement anti-utilitariste dans les sciences sociales)*, 2006, n° 28 : « Penser la crise de l'école : perspectives anti-utilitaristes ».
- Revue de psychologie de la motivation*, 2006, n° 41 : « Éducation et société ».
- ROBERT André D. *Miroirs du syndicalisme enseignant/ postface de Louis Weber*. Paris : Institut de recherches de la FSU ; Syllepse, 2006. – 187 p. (Nouveaux regards).
- RONDAL Jean-Adolphe. *Expliquer l'acquisition du langage : caveats et perspectives*. Sprimont [Belgique] : P. Mardaga, 2006. – 234 p. (Psychologie et sciences humaines).
- RUS Eric de. *Intériorité de la personne et éducation chez Edith Stein*. Paris : Éd. du Cerf, 2006. – 318 p. (La nuit surveillée).
- VASCONCELLOS Maria. *L'enseignement supérieur en France*. Paris : La Découverte, 2006. – 121 p. (Repères ; sociologie n° 429).
- VIDBERG Martin. *Le journal d'un remplaçant*. Paris : Delcourt, 2007. – 125 p., illus. (Shampooing).

PISA: secondary analyses, theoretical and methodological questions and debates

Rolf V. Olsen and Svein Lie – Large-scale international comparative achievement studies in education: their primary purposes and beyond

p. 11

The present article presents a systematic overview of primary and secondary research perspectives related to the large-scale international studies like PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study). Firstly, the primary purpose for these studies is discussed, and it is described how the main purpose have gradually been changed from educational research to a means for providing educational indicators for control and effective national policy development. This change has largely left the area open for rich possibilities for a variety of secondary analyses of the data, which has been released for public use. The main part of the article goes on to describe various perspectives for relevant secondary analyses. For each of these perspectives a number of secondary studies are briefly discussed, and further research is suggested. Our main aim is to demonstrate the rich potential for further research based on the international database for these studies. There are certainly a variety of research questions, both at national and international level, waiting to be effectively addressed by these freely available data.

Pierre Vriinaud – Measuring literacy in PISA. Methodology is the answer, but what was the question?

p. 27

The purpose of the article is to show how methodological aspects (measurement models and algorithmic estimates) and the way literacy is conceptualised and defined in the PISA survey are interacting. To introduce the matter, We first present the conceptual changes in assessing the knowledge acquired by students in the international surveys from the first IEA surveys to the PISA survey. Bearing in mind that any measurement is a construct, we then explore the psychometric theories founding the measuring models used for assessing the knowledge acquired by students. By presenting classical approaches, we can then introduce the concepts of difficulty and item discrimination as well as the concept of global consistency in item tests. We will insist on the conditions for validating psychometric models (unidimensionality and conditional independence of items) that create strong constraints on the way the measured variables are set up. Next we present the PISA measurement model. It belongs to the item response model group (RMI). PISA singled out a model known as the Rasch model. We will see the consequences of this choice on the assessed skill and test format. Next we will tackle a technical issue, not often tackled: algorithm estimate. The RMI settings (item difficulty, subject-matter competencies) are not calculated but estimated. Numerous studies have been done on estimation algorithms. We here re-state that estimation algorithms come partly from thinking on missing values. It leads us to mention the typology of missing values in PISA. We underline the fact that procedures lead not to estimate one value per subject but to construct the distribution of competencies in each subject. We insist on the consequences of such an approach in the interpretation of the results and the way secondary analyses

are conducted. Finally, we discuss the truly unidimensional aspect of variables in terms of contents and methodology. We ask ourselves whether there could be an apparent contradiction when considering simultaneously underscales and full scales. Then we wonder on the plurality of acceptable interpretations of literacy – as it is measured in PISA – when related to other surveys aiming at assessing the same type of competencies.

Jean Moreau, Christian Nidegger, Anne Soussi – Definition of competencies, methodological choices and repercussions on school policy in Switzerland

p. 43

In this paper we will use examples to highlight the methodological and conceptual choices made in the PISA survey. To what extent do the defined and operationalised competencies allow comparisons between different education systems while not introducing cultural and linguistic biases? Do the sophisticated statistical methods used allow for transparency and the appropriation of the data by the results recipients: politicians, the school community? In the second part, we will show the effects of the PISA results on school policy in Switzerland and more particularly on the creation of a national project for school harmonization (HarmoS). This project uses methods based on those used in PISA, notably a competence based approach, in spite of different objectives such as the management and regulation of the education systems.

Manfred Prenzel, Karin Zimmer – National extensions of PISA 2003 in Germany: Key results and benefits

p. 55

In Germany, the international PISA survey was considerably extended in 2003. The sample size was substantially increased to provide detailed results for the sixteen German *Länder* [Districts] on all issues addressed on a worldwide scale. Moreover, the national design includes an oversampling of students with a migration background. Newly developed national tests for science and problem-solving investigated facets not covered internationally, thus yielding a more comprehensive picture of these domains. Additional school and teacher questionnaires focused on specific challenges schools face in Germany, and identified the actions schools take to address these challenges. Finally, to monitor progress in class, the measurement of mathematics and science competencies was repeated after one school year for a subset of students. The present paper paradigmatically illustrates the benefits of enlarging the international core design of PISA, and presents the key results thus obtained.

Martine Rémond – Focusing on PIRLS and PISA international assessment of French students

p. 71

In this article, the results of two international surveys on literacy are analysed and compared to those of the last French assessment surveys. Whether they are in fourth grade for PIRLS or tested at the age of 15 for PISA, the French students achieve a rather average score in reading literacy. Compared to the other thirty participating countries. Our cross-analyses of assessment devices allow to trace the evolution of heterogeneity as revealed by PISA in 2003 and propose recommendations for curriculum contents and teaching strategies.

Élisabeth Bautier, Jacques Crinon, Patrick Rayou, Jean-Yves Rochex – literacy results, students' ways and means: secondary analyses of the PISA 2000 study

p. 85

This article reports a study of secondary and further analyses of the PISA 2000 survey on literacy. It combines work done before PISA 2000 analysing its tests, grading and reporting methods – as defined by PISA designers –, a second processing of the answers of more than 800 French students based on different theoretical assumptions, and a collection of interviews aiming at better understanding the ways and means of the students working on those tests. This set of analyses leads us to question both our concepts of literacy and methodological and psychometric presuppositions upon which PISA 2000 is based. It also makes us object to the logic of "ranking" very often found in comments or publications about those results.

Virginie Albe – Students' discursive processes and social roles in discussion groups discussing a socio-scientific controversy

p. 103

The study of socio-scientific issues in class was proposed as a contribution to the democratization of science. For this purpose, some situations encouraging discursive practices between students were developed. We studied how 11th-graders dealt with a current socio-scientific controversy in discussion groups. Using a microethnographic approach, we identified different processes of collective discussion in the groups. The results show that epistemological, contextual and social elements influence the construction of knowledge in those situations.

Monique Loquet – Analysing professional practices: example of experience-based assessments by a corporal expression teacher

p. 119

The goal of this article is to identify the nature of the knowledge relevant to the teaching of physical education by thoroughly studying a situation in continuing education for teachers. This study is part of didactics research work. It aims at clarifying the way the choice of knowledge to be taught and the choice of educational practices. The analysis clearly shows the instructor's inventiveness in regulating the trainees' artistic physical activities. It also shows the benefits that teachers can get from a programme teaching professional practices applied to the changing reality of classroom teaching.

Marie-Paule Poggi, Nathalie Wallian, Mathilde Musard – The social construction of school inspector judgment: definition of the teacher's job through the analysis of inspector reports in physical education

p. 131

The main hypothesis is that the inspector judgments does not answer only to pedagogical and didactical norms but they are also built from social constraints such as the type of school students and their sociocultural characteristics of the teachers. The content analysis of 111 school inspector reports helps to better understand the educational model valorized by the school institution. The inspector discourse tends towards to standardize the institutional expectations. The general educational principles concern the use of "active methods" and the student differentiation. Meanwhile, these requirements are more important for disfavored schools, female PE teachers, and less experienced teachers. In addition, nobody knows what pupils learn during the lesson. In disregarding the actual student practices, does not the inspector deprive of a dimension, which is essential to assess the effectiveness of teaching?

Damien Tessier, Philippe Sarrazin, David Trouilloud – Motivating environments set by teachers and students' learning activities: what research says

p. 147

The purpose of this article is to report current research work on teachers' behaviours in class and their effects on the students' learning activities. The gathering of those studies was done mainly within the framework of two theoretical paradigms: the theory of motivation through a sense of accomplishment and the theory of students' self-determination. It shows first the positive consequences of completely mastered environments and the debated effects of competitive environments, and then the benefits of the styles supporting the students' autonomy; contrary to assessment-oriented styles. Nevertheless, applying the teaching principles derived from those studies to class teaching requires the teachers to be careful. The complexity of true teaching environments compared to situations developed in laboratory, the constraints on teachers that reduce their teaching freedom and make them be more assessment-oriented, and the methodological limits of questionnaire studies, give no guarantee as to the ecological validity of such research work.

PISA: análisis secundarios, cuestiones y debates teóricos y metodológicos

Rolf V. Olsen, Svein Lie – Las valuaciones internacionales y la investigación en educación: principales objetivos y perspectivas

p. 11

Este artículo propone una vista estructurada del conjunto de las perspectivas de investigaciones primarias y secundarias en relación con estudios internacionales de gran amplitud (PISA et TIMSS). Se examina primero el objetivo de estos estudios antes de describir de qué manera su objetivo principal ha ido evolucionando poco a poco de la investigación en educación a la constitución de recursos de indicadores pedagógicos de control y de desarrollo de las políticas nacionales. Esta evolución ha dado paso a abundantes posibilidades de análisis secundarios de los datos con vistas de su difusión al público. En su mayor parte, el artículo se dedica a una descripción de las perspectivas variadas de análisis secundarios pertinentes y para cada una de estas perspectivas, se presenta brevemente cierto número de estudios secundarios y se proponen prolongamientos de investigación. Nuestra intención es demostrar el fuerte potencial de investigaciones complementarias contenido en esas bases de datos internacionales.

Pierre Vriгдаud – La medida de la literacia en PISA. La metodología es la respuesta, pero ¿cuál era la pregunta?

p. 27

El objetivo de este artículo es mostrar la interacción entre los aspectos metodológicos (el modelo de medida y los algoritmos de valoración) y la manera en que se conceptualiza y se define la literacia en la encuesta PISA. Para introducir el tema, se evocan primero los cambios conceptuales cuya valoración de las experiencias de los alumnos ha sido tratada en las encuestas internacionales desde las primeras encuestas del IEA hasta la encuesta PISA. Después de recordar que cualquier medida es una construcción, se exponen las teorías psicométricas que fundan los modelos de medida utilizados para la valoración de las experiencias de los alumnos. La presentación de los enfoques clásicos permite introducir los conceptos de dificultad y de discriminación de los elementos, así como el concepto de consistencia global de una prueba. Se insiste en las condiciones de validez de los modelos psicométricos (unidimensionalidad e independencia condicional de los elementos) que imponen coacciones fuertes sobre la manera en que se construye la variable medida. Se presenta después el modelo de medida utilizado en PISA que pertenece a la familia de los modelos de respuesta al elemento (MRI). PISA ha privilegiado el modelo llamado de Rasch, se discute acerca de las consecuencias de esta elección sobre la competencia medida y el formato de la prueba. Se trata después un punto muy técnico muy poco tocado: el del algoritmo de valoración. Los parámetros de los MRI (dificultad de los elementos, competencias de los sujetos) no son calculados sino apreciados. Los algoritmos de valoración han dado lugar a numerosos trabajos. Se recuerda que el algoritmo de valoración procede en parte de una reflexión sobre los valores que faltan. Esto induce a evocar la tipología de los valores que faltan en PISA. Se subraya que el procedimiento induce a valorar no un valor por sujeto sino a construir la distribución de las competencias de cada sujeto, se insiste sobre las consecuencias de este enfoque sobre la

interpretación de los resultados y sobre la manera de realizar análisis secundarios. Por fin, se discute el carácter realmente unidimensional de la variable desde el punto de vista de los contenidos y desde el punto de vista metodológico. Nos interrogamos sobre la contradicción aparente que puede haber en considerar simultáneamente subescalas y una escala global. Después, nos interrogamos sobre la pluralidad de las interpretaciones que puede soportar la literacia tal y como se mide en PISA cuando se pone en relación con otras encuestas que tienden a evaluar el mismo tipo de competencia.

Christian Niddeger, Anne Soussi, Jean Moreau – Definición de las competencias, elecciones metodológicas y repercusiones sobre la política escolar en Suiza

p. 43

En este artículo, trataremos de poner de manifiesto las elecciones metodológicas y conceptuales tomadas en cuenta en la encuesta PISA, ilustrando con algunos ejemplos. ¿En qué medida las competencias definidas y operacionalizadas permiten la comparación de los sistemas educativos? o ¿no introducirán rodeos culturales y lingüísticos? ¿Los métodos estadísticos sofisticados utilizados permiten la transparencia y la apropiación de los datos por los destinatarios de los resultados: responsables políticos y actores de la escuela? En una segunda parte, mostraremos los efectos de los resultados de PISA sobre la política escolar en Suiza y más particularmente sobre la aplicación de un proyecto de armonización escolar nacional (HarmoS) que se inspira ampliamente de gestiones puestas en obra en la encuesta PISA, especialmente enfoque por competencias, a pesar de objetivos diferentes tales como el pilotaje y la regulación de los sistemas educativos.

Manfred Prenzel, Karin Zimmer – PISA 2003 en Alemania: Principales resultados y utilidad de las pruebas nacionales suplementarias

p. 55

El estudio PISA 2003 se ha realizado en Alemania enriquecido por numerosas extensiones nacionales. A fin de poder proceder a comparaciones entre los 16 *Länder*, el tamaño de la muestra se ha aumentado de manera sensible. También se ha integrado una muestra suplementaria de alumnos procedentes de la inmigración. Durante un segundo día de pruebas, los alumnos han sido sometidos a pruebas suplementarias en matemáticas y en ciencias. Los equipos de dirección, pero también los docentes han sido interrogados en cada establecimiento sobre sus cargas de trabajo así como sobre su participación a la mejora de la calidad de la docencia. Por fin, un año más tarde se han vuelto a someter a pruebas de matemáticas y ciencias a los alumnos o clases que tomaron parte en PISA 2003, a fin de medir sus progresos en el aprendizaje. En nuestro artículo, según los resultados de nuestra encuesta, mostramos el provecho que se puede sacar de una ampliación del estudio PISA.

Martine Rémond – Enfoques de las valuaciones internacionales PIRLS y PISA sobre los alumnos franceses

p. 71

En este artículo, los resultados de las dos encuestas internacionales sobre la *reading literacy* se analizan y se comparan con los de las últimas valuaciones-balances francesas. Estén en CM1 para PIRLS o evaluados a los 15 años en PISA, entre la treintena de países participantes los alumnos franceses dan pruebas de un nivel bastante medio en lo que concierne *reading literacy*. Nuestros análisis cruzados de los dispositivos de evaluación permiten percibir una especie de

génesis de la heterogeneidad revelada por PISA en 2003 y formular recomendaciones sobre los contenidos y sobre las estrategias de enseñanza.

Élisabeth Bautier, Jacques Crinon, Patrick Rayou, Jean-Yves Rochex – Resultados en literacia, modos de hacer y universos utilizados por los alumnos: análisis secundarios de la encuesta PISA 2000

p. 85

Este artículo da cuenta de un trabajo de análisis secundario y complementario de las pruebas de la encuesta PISA 2000 que trata de la literacia. Se han conjugado un análisis *a priori* de esas pruebas, de sus modalidades de corrección y de valoración – tales como las concibieron los creadores de PISA –, un nuevo tratamiento estadístico de las respuestas de más de 800 alumnos franceses, construido a partir de hipótesis teóricas diferentes, y una recopilación de conversaciones cuyo objeto es entender mejor los modos de hacer de los alumnos frente a las pruebas consideradas. El conjunto de los análisis así producidos mueve a interrogar a la vez las concepciones de la literacia y los presupuestos metodológicos y psicométricos sobre los que se apoya la encuesta PISA 2000, y a discutir la lógica de “lista de resultados” en la que a menudo se han publicado y comentado sus resultados.

Virginie Albe – Procedimientos discursivos y papeles sociales de alumnos en grupos de discusión sobre una controversia socio-científica

p. 103

A fin de contribuir a la democratización de las ciencias, se ha propuesto en clase el estudio de cuestiones socio-científicas. Con este fin, se han desarrollado situaciones que favorecen las prácticas discursivas entre alumnos. Hemos estudiado cómo alumnos de primero enfocan una controversia socio-científica actual en grupos de discusión. En un enfoque micro-etnográfico, hemos identificado diferentes procedimientos de discusión colectiva en los grupos. Los resultados muestran que factores epistemológicos, contextuales, y sociales influyen en la elaboración de conocimientos en esas situaciones.

Monique Loquet – Análisis de los gestos profesionales: ilustración de “el ojo del tratante de caballos” en un formadora en expresión corporal

p. 119

A través del estudio preciso de una situación de formación continua destinada a los profesores, este artículo se propone identificar la naturaleza de los saberes pertinentes que enseñar en Educación física y deportiva. Este estudio, incluido en el campo de las investigaciones didácticas, pretende, además, dilucidar la manera en que se entremezclan la elección de los saberes que transmitir y la elección de los gestos de transmisión. El análisis evidencia la ingeniosidad de la formadora para controlar la actividad física artística de los cursillistas. Muestra la importancia para los profesores de una formación a los gestos profesionales aplicados a la realidad inestable de la clase.

Marie-Paule Poggi, Nathalie Wallian, Mathilde Musard – La construcción social del juicio inspectorial: la definición del oficio de docente vista a través de un análisis de los informes de inspección en EPS

p. 131

Hacemos la hipótesis que el juicio inspectorial no sólo responde a normas pedagógicas y didácticas sino que también se construye a

partir de obligaciones sociales tales como el tipo de público escolar y las características socio-culturales de los docentes. El análisis del contenido de ciento once informes de inspección nos permite examinar la naturaleza de los modelos de intervención valorizados por la institución. El discurso inspectorial tiende a uniformizar sus esperas invocando principios generales que preconizan principalmente el valerse de métodos activos y de un tratamiento diferenciado de los alumnos. Sin embargo, estas exigencias se hacen más apremiantes en establecimientos desfavorecidos, dirigiéndose a mujeres y a jóvenes docentes. Por otra parte, no se sabe nada de lo que los alumnos aprenden durante una lección. Al ignorar las prácticas reales de los alumnos, ¿no le falta al inspector una dimensión esencial para evaluar la eficacia de la docencia?

Damien Tessier, Philippe Sarrazin, David Trouilloud – Clima motivacional instaurado por el docente y actividades de aprendizaje de los alumnos: el estado de las investigaciones

p. 147

El objetivo de este artículo es presentar el estado de las investigaciones sobre los comportamientos del docente en clase y sus efectos sobre las actividades de aprendizaje de los alumnos. Realizadas principalmente en el marco de dos paradigmas teóricos– la teoría de la motivación a la realización y la teoría de la autodeterminación– la comparación de estos trabajos muestra por una parte las consecuencias positivas de un clima de dominio y los efectos controvertidos de un clima de competición, y por otra parte los beneficios de un estilo que sostiene la autonomía de los alumnos a la inversa de un estilo que controla. Sin embargo, la aplicación a la clase de los principios educativos procedentes de esos estudios, requiere cierta prudencia por parte de los docentes. La complejidad del contexto natural de docencia en comparación con las situaciones estudiadas en laboratorio, las coacciones que se imponen al docente y reducen su libertad pedagógica y lo inducen a ser censor para con sus alumnos, así como los límites metodológicos de los estudios por cuestionario, constituyen las principales reservas que hay que tener en cuanto a la validez ecológica de estas investigaciones.

PISA: Sekundaranalysen, theoretische und methodologische Fragen und Debatten

Rolf V. Olsen, Svein Lie – Internationale Vergleichende Analysen und Forschung in der Erziehung: Hauptziele und Perspektiven

p. 11

Dieser Artikel bietet einen strukturierten Überblick über die gesamten Perspektiven in der Primär- und Sekundärforschungen in Bezug auf die breiten internationalen vergleichenden Analysen (PISA und TIMSS). Das Ziel dieser Studien wird vorerst untersucht, dann beschreiben wir die Art und Weise, wie ihr Hauptanliegen sich nach und nach von der Forschung in der Erziehung bis zum Zusammenbau von Ressourcen über pädagogische Indikatoren zur Kontrolle und Entwicklung der nationalen Erziehungspolitik entwickelt hat. Dieser Werdegang hat den Weg geöffnet für zahlreiche Möglichkeiten der Sekundäranalysen im Hinblick auf ihre öffentliche Verbreitung. Der Artikel nimmt sich hauptsächlich eine Beschreibung der verschiedenen Perspektiven für relevante Sekundäranalysen vor. Und für jede Perspektive wird eine gewisse Zahl Sekundäranalysen kurz vorgestellt und weiterführende Forschungsansätze werden angeboten. Unsere Absicht besteht darin, das starke Potential an Zusatzforschungen zu zeigen, das in diesen internationalen Datenbanken steckt.

Pierre Vriinaud – Die Bewertung der *literacy* bei PISA: Methodologie ist die Antwort, aber was war bitte noch mal die Frage?

p. 27

Das Ziel dieses Artikels ist, die Wechselwirkungen zwischen methodologischen Aspekten (das Meßverfahren und die Schätzungsalgorithmen) und der Art und Weise zu zeigen, wie *literacy* bei der PISA Studie konzipiert und definiert wird. Um das Thema einzuführen, erwähnen wir zuerst die konzeptuellen Veränderungen, die bei der Evaluierung der Schüler in den internationalen Studien seit den ersten IEA-Untersuchungen bis hin zur PISA-Studie stattgefunden haben. Nachdem wir daran erinnert haben, dass jede Messung ein Konstrukt ist, stellen wir die psychometrischen Theorien vor, auf denen die Meßverfahren beruhen, die zur Evaluierung der Schüler benutzt werden. Die Vorstellung der klassischen Herangehensweisen ermöglicht es, die Begriffe der Schwierigkeit und der Diskriminierung der Items einzuführen, so wie den Begriff der globalen Konsistenz einer Aufgabe. Wir unterstreichen die Geltungsbedingungen der psychometrischen Modelle (Eindimensionalität und bedingte Unabhängigkeit der Items), die starken Druck über die Art und Weise ausüben, wie die gemessene Variable gebildet ist. Dann stellen wir das Meßverfahren vor, das bei PISA benutzt wird und das der Gattung der Modelle der Antwort auf das Item (MRI) angehört. PISA hat das sogenannte Rasch-Modell bevorzugt: Wir erörtern die Konsequenzen dieser Wahl über die gemessene Kompetenz und über das Format der Prüfung. Dann besprechen wir einen höchst technischen Punkt, der selten angegangen wird : das Schätzungsalgorithmus. Die Parameter der MRI (Schwierigkeit der Items, Kompetenzen der Subjekte) werden nicht ausgerechnet sondern eingeschätzt. Die Schätzungsalgorithmen sind das Thema zahlreicher Arbeiten. Wir erinnern daran, dass der Schätzungsalgorithmus sich zum Teil aus einer

Überlegung über die fehlenden Werten ergibt. Dies führt dazu, die Typologie der fehlenden Werte bei PISA zu erwähnen. Wir betonen, dass die Prozedur darauf hinausläuft, nicht etwa einen Wert für jeden Subjekt, sondern die Verteilung der Kompetenzen jedes Subjekts zu schätzen. Wir betonen die Konsequenzen dieser Herangehensweise über die Interpretation der Ergebnisse und über die Art und Weise, wie man Sekundäranalysen durchführt. Zum Schluss befragen wir das eindimensionale Charakter der Variable, was die Inhalte und die Methodologie angeht. Wir stellen uns Fragen über den anscheinenden Widerspruch in der gleichzeitigen Betrachtung der Unterskalen und einer globalen Skala. Dann stellen wir uns Fragen über die Vielfalt der Interpretationen, die die *literacy* ertragen kann, wenn sie wie bei PISA geschätzt wird und wenn man sie mit anderen Studien vergleicht, die die gleichen Kompetenzen evaluieren.

Jean Moreau, Christian Nidegger, Annse Soussi – Definition des Konzeptbegriffs, methodologisches Vorgehen und Auswirkungen auf die Schweizer Schulpolitik

p. 43

In diesem Artikel versuchen wir zuerst, anhand einiger Beispiele, die methodologische und konzeptuelle Ausgangstellung der Pisa-Studie darzustellen. Erlauben Ausformulierung und Operationalisierung des Kompetenzbegriffs den Vergleich von Bildungssystemen oder führen sie nicht vielmehr kulturelle und sprachlichen Begriffsverschiebungen ein? Erlauben die hochentwickelten statistischen Methoden, die eingesetzt werden, die Lesbarkeit und die Aneignung der Resultate durch Bildungspolitiker und Akteure in den Schulen? Im zweiten Teil des Artikels zeigen wir die Auswirkungen der Pisa-Resultate auf die Bildungspolitik in der Schweiz, insbesondere auf das Projekt der Harmonisierung der obligatorischen Schule (HarmoS). Dieses Projekt orientiert das sich weitgehend an der Pisa-Studie, insbesondere was das die Kompetenzanforderungen betrifft, auf die trotz unterschiedlicher Zielsetzungen und Steuerung der Bildungssysteme zurückgegriffen wird.

Manfred Prenzel, Karin Zimmer – PISA 2003 in Deutschland: Zentrale Befunde und der Nutzen nationaler Erweiterungen

p. 55

Die PISA-Erhebung 2003 wurde in Deutschland mit zahlreichen nationalen Erweiterungen durchgeführt. Um Vergleiche zwischen den 16 Bundesländern durchführen zu können, wurde die Stichprobengröße um ein mehrfaches vergrößert. Außerdem wurde eine Zusatzstichprobe von Schülerinnen und Schülern mit Migrationshintergrund gezogen. An einem zweiten Testtag erhielten die Schülerinnen und Schüler zusätzliche nationale Mathematik- und Naturwissenschaftstests. Neben den Schulleitungen wurden an jeder Schule Lehrkräfte zu Belastungen und Aktivitäten zur Qualitätsverbesserung befragt. Schließlich wurden die Schülerinnen und Schüler bzw. Klassen, die an PISA 2003 teilnahmen, ein Jahr später noch einmal in Mathematik und in den Naturwissenschaften getestet, um die Lernfortschritte zu erfassen. In unserem Beitrag zeigen wir beispielhaft an Ergebnissen unserer Studie, was es nützen kann, den Ansatz von PISA zu erweitern.

Martine Rémond – Erläuterung der internationalen Studien PIRLS und PISA, was französische Schüler angeht

p. 71

In diesem Artikel werden die Ergebnisse bei den beiden internationalen Studien in Bezug auf *reading literacy* analysiert und mit denen

der letzten französischen Bilanz-Bewertungen verglichen. In der 4. Klasse (CM1) bei PIRLS oder auch in der 9. Klasse bei PISA haben französische Schüler ein eher mittelmäßiges Niveau in *reading literacy*, wenn man sie mit den ungefähr dreißig anderen Ländern vergleicht. Unsere vergleichende Analysen der Erhebungsmodalitäten ermöglicht uns, eine Art Genesis der Heterogenität zu beobachten, die von PISA 2003 offenbart wurde, und somit Anregungen über Lehrinhalte und Unterrichtsstrategien zu empfehlen.

Élisabeth Bautier, Jacques Crinon, Patrick Rayou, Jean-Yves Rochex – Leistungen in der *literacy*: Herangehensweisen und von Schülern mobilisierte Gedankenwelten: Sekundäranalysen der Studie PISA 2000

p. 85

Dieser Artikel berichtet über Sekundär- und Zusatzuntersuchungen der Aufgaben aus der Studie PISA 2000, was *literacy* angeht. Es wurden dreifache Vorgänge miteinander verknüpft: zuerst eine im Voraus geführte Analyse dieser Aufgaben und derer Verbesserungs- und Bewertungsweisen – so wie sie von den Entwerfern der PISA-Studie vorgeschrieben waren; dann eine statistische auf unterschiedlichen theoretischen Hypothesen beruhende Neubearbeitung der Antworten von mehr als 800 französischen Schülern, und drittens eine Gesprächensammlung, die darauf hinzielte, die Herangehensweisen der Schüler vor den jeweiligen Aufgaben besser zu verstehen. Sämtliche sich aus dieser Phase ergebende Analysen führen dazu, sowohl die Auffassungen der *literacy* als auch die methodologischen und psychometrischen Voraussetzungen in Frage zu stellen, auf denen die PISA-Studie 2000 basiert. Gleichermaßen kann man folglich die „Preisträger“-Logik bestreiten, nach der ihre Ergebnisse häufig veröffentlicht und kommentiert wurden.

Virginie Albe – Diskursive Vorgänge und soziale Rollen der Schüler in Diskussionsgruppen über eine gesellschaftlich-wissenschaftliche Auseinandersetzung

p. 103

Das Besprechen gesellschaftlich-wissenschaftlicher Fragen in der Klasse wurde empfohlen, um die Demokratisierung der Wissenschaften zu fördern. Dazu wurden Situationen entwickelt, die diskursive Praktiken bei Schülern fördern. Wir haben darüber geforscht, wie in Diskussionsgruppen Schüler der 11. Klasse (*1ère*) an eine zeitgenössische gesellschaftlich-wissenschaftliche Auseinandersetzung herangehen. Einer mikro-ethnographischen Vorgehensweise zufolge haben wir verschiedene Vorgänge der kollektiven Diskussion in den jeweiligen Gruppen identifiziert. Die Ergebnisse zeigen, dass epistemologische, kontextbedingte und soziale Faktoren die Entwicklung des Wissens in solchen Situationen beeinflussen.

Monique Loquet – Analyse beruflicher Gesten: Das „Auge des Kenners“ bei einer Ausbilderin im Fach körperlicher Ausdruck

p. 119

An Hand der genauen Beobachtung einer Situation in einem Fortbildungsseminar ist das Ziel dieses Artikels, die relevanten Kenntnisse zu identifizieren, die im Sportunterricht herüberzubringen sind. Diese Forschungsarbeit, die der didaktischen Forschung zugeordnet ist, zielt zudem darauf ab, die Art und Weise zu erklären, wie sich die Wahl der zu vermittelnden Kenntnisse und die Wahl der Gesten der Vermittlung verbinden. Die Analyse betont den Einfallsreichtum der Ausbilderin bei der Regulierung der künstlerischen

physischen Aktivität der Teilnehmer. Sie zeigt den Gewinn, den die Lehrkräfte aus einer Ausbildung über berufliche Gesten ziehen können, die der sich stets verändernden Realität in der Klasse gerecht werden.

Marie-Paule Poggi, Nathalie Wallian, Mathilde Musard – Die soziale Herstellung der Beurteilung der Inspektoren: die Definition des Lehrerberufs durch eine Analyse von Inspektionsberichten im Unterrichtsfach Sport

p. 131

Wir gehen von der Hypothese aus, dass die Beurteilung des Inspektoren nicht nur pädagogischen und didaktischen Normen entspricht, sondern dass sie sich auf Grund sozialer Zwänge bildet, wie etwa dem Typus der Schülerschaft und den sozio-kulturellen Merkmalen der jeweiligen Lehrer. Die Analyse über den Inhalt von 111 Inspektionsberichten erlaubt uns, die Unterrichtsformen näher zu betrachten, die von der Institution Schule positiv bewertet werden. Die Inspektoren neigen dazu, ihre Erwartungen zu uniformisieren, indem sie sich auf Hauptprinzipien berufen, die hauptsächlich aktive Methoden und eine differenzierte Schülerbehandlung fördern. Allerdings werden diese Forderungen an problembelasteten Schulen, bei Frauen und bei jungen Lehrern stärker und deutlicher ausgedrückt. Übrigens erfährt man in diesen Berichten nichts über das, was die Schüler im Unterricht lernen. Wenn er die tatsächlichen Praktiken der Schüler ignoriert, stellen wir die Frage, ob der Inspektor sich nicht einer wesentlichen Dimension zur Beurteilung der Unterrichtseffizienz entgehen lässt.

Damien Tessier, Philippe Sarrazin, David Trouilloud – Vom Lehrer erzeugtes Motivationsklima und Lernaktivitäten der Schüler: der Stand der Forschung

p. 147

Ziel dieses Artikels ist es, den Stand der Forschung über das Verhalten des Lehrers in der Klasse und deren Auswirkungen über die Lernaktivitäten der Schüler vorzustellen. Im Rahmen zweierlei theoretischer Paradigmen durchgeführt, – die Theorie der Motivierung zum Handeln und die Theorie der Selbstbestimmung – zeigt die Rezension dieser Forschungsarbeiten zwei Aspekte: einerseits die positiven Folgen eines Klimas der Beherrschung und die bestrittenen Folgen eines Klimas des Wettkampfes; und andererseits die positiven Auswirkungen einer Unterrichtsmethode, die die Autonomie der Schüler fördert (im Gegensatz zu einer kontrollierenden Methode). Jedoch verlangt die Anwendung in der Klasse der aus diesen Forschungsarbeiten hervorgehenden erzieherischen Prinzipien eine gewisse Vorsicht von den Lehrern. Erstens die Komplexität des natürlichen Unterrichtskontextes im Vergleich mit den im Labor bearbeiteten Situationen, zweitens die sich dem Lehrer aufbürdenden Zwänge, die die pädagogische Freiheit einschränken und die ihn dazu führen, eine kontrollierende Unterrichtsmethode gegenüber seinen Schülern zu benutzen, und drittens die methodologischen Grenzen der über Fragebögen geführten Forschungsarbeiten sind die Hauptbedenken, die man gegenüber der ökologischen Gültigkeit dieser Forschungen äußern kann.