



Voyage dans un espace multidimensionnel

Textes réunis en l'honneur de
Daniel Bain



Voyage dans un espace multidimensionnel

Textes réunis en l'honneur de
Daniel Bain

rédigés par :

Jean-Paul Bronckart

Jean Cardinet

Marcel Crahay

André Giordan

Armin Gretler

Aletta Grisay

Philippe Perrenoud

Préface de Jacques Weiss



Table des matières

Avant-propos5

Norberto Bottani

Préface7

Jacques Weiss

**Les enseignants, leurs croyances, leurs pratiques
d'évaluation et l'échec scolaire**15

Marcel Crahay

**Réflexions sur l'effet-école
(et sur ses manifestations dans un archipel du bout du monde)**35

Aletta Grisay

**Les systèmes éducatifs face aux inégalités et à l'échec scolaire:
une impuissance teintée de lassitude**53

Philippe Perrenoud



De la didactique de la langue à la didactique de la littérature71

Jean-Paul Bronckart

Didactique des sciences: faut-il dépasser le constructivisme? 91

André Giordan

Faut-il généraliser?107

Jean Cardinet

**Zur Bedeutung international vergleichender
Leistungsmessungen im Bildungswesen125**

Armin Gretler

Adresses et contacts des auteurs147



Daniel Bain



Avant-propos

Ce recueil de textes est un hommage que la communauté des chercheuses et chercheurs du Département de l’instruction publique du Canton de Genève, et notamment du Service de la recherche en éducation (SRED), a souhaité rendre à Daniel Bain à l’occasion de son départ à la retraite.

La dernière année d’un long engagement entièrement dédié au service de l’école publique genevoise, Daniel Bain l’a vécue dans un environnement nouveau, à certains égards aseptisé, loin des bruits, des odeurs, des cris, des couleurs, de l’agitation et des tensions qui font le charme unique d’une école et, dans son cas particulier, du Collège des Coudriers où il a passé plusieurs années au sein du Centre de recherches psychopédagogiques du Cycle d’orientation. Au Quai du Rhône, le nouveau siège du SRED, les couloirs sont vides et silencieux. Il y règne une atmosphère feutrée et studieuse, certainement appréciable mais qui, je le crains, doit avoir été difficile à supporter pour un homme de terrain comme Daniel Bain: il n’avait pas besoin d’une distanciation physique pour analyser avec une acuité extraordinaire et une finesse remarquable la scène scolaire et ses acteurs. Cependant, Daniel Bain n’a rien laissé entrevoir de ses sentiments. Avec la distinction et la gentillesse qui le caractérisent, il s’est adapté au nouveau milieu avec un sens de la collégialité exemplaire dont je lui suis reconnaissant. Il a aussi retroussé ses manches et s’est lancé dans de nouveaux travaux, bien qu’il eût demandé de pouvoir profiter de cette période pour terminer les recherches qu’il avait en chantier.

En réunissant les contributions de scientifiques qui ont eu le privilège d’accompagner Daniel Bain pendant sa carrière, les chercheuses et chercheurs du SRED veulent témoigner à leur collègue, qu’ils considèrent comme l’un des chercheurs en éducation parmi les plus brillants et les plus fertiles de son époque en Suisse romande, toute leur estime et leur reconnaissance. Tous ceux qui ont eu la chance de travailler avec lui peuvent lui savoir gré de ses leçons de rigueur intellectuelle dans l’analyse et le contrôle des données, de sa générosité et surtout de sa passion pour la recherche, constamment alimentée par un grand attachement pour l’école, objet de nos soucis et de nos études.

Norberto Bottani
Directeur du SRED



Préface

Jacques Weiss

Introduction

Le présent ouvrage rassemble des contributions d'éminents chercheurs, amis et par ailleurs compagnons du parcours professionnel de Daniel Bain.

Le lecteur pourrait être frappé au premier abord par l'hétérogénéité des profils des chercheurs sollicités: psycholinguiste, sociologue, psychologue et pédagogue, et par la diversité des thèmes traités. Ceux-ci vont en effet de la didactique du français à celle des sciences, de l'étude de l'efficacité des organisations scolaires au développement de la théorie de la généralisabilité.

L'amplitude des domaines traités dans cet ouvrage, tous en lien avec les travaux de Daniel Bain, atteste en effet l'éclectisme des questions scientifiques que celui-ci s'est posées tout au long de sa carrière, en même temps que son ouverture à la diversité des problématiques propres à l'école.

La cohérence de l'œuvre de Daniel Bain – et de cet ouvrage – est moins à chercher dans l'objet traité que dans les démarches adoptées; des démarches constamment empreintes de rigueur, d'exigence et de doute scientifique. Daniel Bain les a appliquées dans ce que sont ses champs de préoccupations et ses domaines d'excellence. Parmi ceux-ci, nous en retiendrons trois, étroitement associés aux articles présentés dans cette publication. Le premier, celui qui a certainement fait l'objet du plus grand nombre d'études de sa part, porte sur l'étude du fonctionnement des systèmes de formation, sur l'efficacité de son organisation, sur l'échec scolaire. Ces études concernent les conditions fonctionnelles de l'apprentissage. Trois textes y sont consacrés. Aletta Grisay, Marcel Crahay et Philippe Perrenoud ont, chacun de leur point de vue de psychopédagogue ou de sociologue, abordé cette question. Le deuxième domaine se rapporte à l'apprentissage lui-même et à l'intervention didactique. Jean-Paul Bronckart, avec qui Daniel Bain a travaillé étroitement, mais aussi André Giordan, traitent de ces questions. Troisièmement, et ceci atteste son souci constant de rigueur et de précision, Daniel Bain s'est engagé, avec Jean Cardinet, à perfectionner les outils de mesures statistiques, outils dont l'usage pourrait être encouragé lors du traitement des données des *surveys* internationaux. Jean Cardinet, en toute logique, expose donc les derniers développements de la théorie de la généralisabilité, alors qu'Armin Gretler montre l'importance pour la Suisse de participer aux enquêtes internationales.



Nous allons reprendre chacun de ces trois grands thèmes en entrant plus précisément dans l'exposé des auteurs.

Le fonctionnement du système de formation

Marcel Crahay, Aletta Grisay et Philippe Perrenoud abordent ensemble une problématique similaire, celle de l'échec scolaire, des effets des mesures prises pour l'enrayer, et au travers de ces études, analysent les représentations et les détours qui freinent les initiatives les plus prometteuses. Sont notamment étudiés *l'effet-enseignant*, *l'effet-école* et *l'effet-système* sur la réussite scolaire.

L'effet-enseignant

Marcel Crahay et Daniel Bain ont tous deux cherché à dégager les effets du redoublement sur la réussite scolaire. Tous deux, ils ont démontré l'inefficacité de cette mesure et par ailleurs constaté la profonde conviction du contraire chez les enseignants. L'article de Marcel Crahay tente d'expliquer la persistance de cette représentation. Il le fait en invoquant la prééminence d'une logique de la prédestination, du fatalisme, et de l'idéologie du don chez les enseignants, selon laquelle il existerait des élèves doués et d'autres qui ne le seraient pas. Convaincus que ce potentiel, favorable ou défavorable, est immuable, les enseignants *usent de l'évaluation comme moyen de sélection*, en cherchant à trier et à classer les élèves une fois pour toutes et le plus tôt possible sur l'échelle des compétences scolaires. Les données analysées montrent en effet que la logique de la prédestination se met en place dès l'accueil des élèves dans les classes, sur la base de pronostics établis en fonction de leur parcours scolaire. Ce positionnement scolaire dès le début de l'année reste ensuite stable. Dans les cas traités par Marcel Crahay, on observe en effet que les enseignants restent «*aveugles*» et «*insensibles*» aux progrès des élèves, selon les termes mêmes de l'auteur.

L'effet-école

La contribution de Marcel Crahay montre donc que l'effet-enseignement sur l'amélioration des apprentissages des élèves est plutôt mince. Aletta Grisay s'interroge, pour sa part, sur l'effet positif que pourraient produire certaines formes d'organisation des établissements scolaires sur les résultats des élèves. Ses recherches examinent les effets produits par le degré de structuration de l'enseignement, par les modalités de direction de l'établissement scolaire, par l'importance de l'effectif de la classe, par le degré d'autonomie de l'école, par les ressources mises à disposition ou par la qualification des enseignants. La plupart des recherches de ce type parviennent à des résultats contradictoires, et par ailleurs peu stables dans le temps pour le même établissement. Aletta Grisay, se référant aux travaux de Bosker et Scheerens, observe que lorsque des différences sont constatées, elles «*sont peu accusées et ne permettent*



que rarement de trancher de manière nette (...) et que les effets sur le rendement explicables par les caractéristiques de l'environnement scolaire et de l'enseignement dispensé ne constituent (...) qu'une proportion très modeste de la variance totale entre élèves (...)). Elle observe toutefois, par ses propres recherches, que dans un même établissement la variance des résultats entre classes s'accroît en quatre ans en raison des pratiques enseignantes d'affectation des élèves et d'orientation-sélection; ce qui vient confirmer les conclusions précédentes de Marcel Crahay. Cet accroissement de la variance interclasses apparaît par ailleurs plus important lorsque l'enseignement est organisé en filières plutôt que dans des classes hétérogènes, comme l'ont montré de nombreux travaux de Daniel Bain.

Devant la «*désespérante monotonie*» des résultats issus de ces recherches dans les pays développés, Aletta Grisay a tourné son regard vers un petit archipel du Pacifique, la République du Vanuatu, où elle parvient, dans un contexte typique de pays en voie de développement, à mettre en évidence un effet significatif d'une variable intitulée *qualité du service*. Elle observe toutefois qu'aux antipodes se manifestent les mêmes effets néfastes de la mise en concurrence des établissements scolaires (échec des élèves, sélection) que ceux qu'elle a obtenus en Belgique, où cohabitent plusieurs réseaux scolaires parallèles.

L'effet-système

Alors que, comme nous venons de le voir, les psychopédagogues doivent, au bout de leurs recherches, conclure à l'insensibilité des enseignants aux progrès des élèves et à l'incapacité des établissements à démontrer leur efficacité, le sociologue Philippe Perrenoud parle «*d'impuissance teintée de lassitude*» lorsqu'il étudie la capacité des systèmes éducatifs à lutter contre les inégalités et l'échec scolaire, tout en déclarant qu'«*il n'y a aucune raison de désespérer*» au vu de l'élévation du niveau de formation des élèves, de la scolarisation croissante des filles ou du taux élevé de scolarisation postobligatoire par exemple.

Philippe Perrenoud ne peut toutefois se satisfaire de ce mieux relatif et qualifie les mesures prises de dérisoires en regard de la force des mécanismes qui les neutralisent. Ces mécanismes, contrairement aux évidences des enseignants, ne se situent pas du côté des élèves et de leurs capacités – ou incapacités – mais relèvent de «*l'inconstance, de l'incohérence et de l'inefficacité programmée de nos politiques de l'éducation*». L'auteur dresse l'inventaire de ces mécanismes et les analyse: le refus de faire partie du problème, la volonté de ne pas savoir, le poids des lobbies, la recherche d'effets à court terme, les orientations de la formation des enseignants, la conception du changement et des organisations et les failles du pilotage des réformes.

Les préoccupations scientifiques de ces trois chercheurs ont également habité Daniel Bain tout au long de son parcours professionnel. Dès le début des années 70 jusqu'aux années 90, il a cherché à analyser et à comprendre le fonctionnement et l'or-



ganisation du Cycle d'orientation – école secondaire du 1^{er} degré – et particulièrement les procédures d'orientation-sélection, leurs effets discriminatoires et différenciateurs. Philippe Perrenoud a été son compagnon de recherche sur ces questions au tout début des années 70 déjà; et l'un et l'autre ont jeté un regard croisé sur la sélection, psychologique pour l'un, sociologique pour l'autre. Cette longue fréquentation scientifique incite Philippe Perrenoud à mentionner, dans le présent ouvrage, une réflexion de Daniel Bain et de sa collègue Dagmar Hexel pour qui *«l'école n'apprend pas, elle sait»*. Philippe Perrenoud trouve donc un écho à son analyse dans les propos de son collègue de la première heure. En retour, Daniel Bain pourrait bien donner son aval à cette assertion du sociologue selon laquelle *«la transformation de l'école ne peut résulter que de la résolution des crises qui pourraient la frapper»*.

Assurément ils tomberaient d'accord pour partager l'espoir qu'il y a, en dépit de ces analyses plutôt pessimistes, d'améliorer l'école, et pour tout entreprendre afin d'y parvenir envers et contre tout. Preuve en est l'énoncé par Philippe Perrenoud de *«Six thèses pour un changement de la culture scolaire»* et l'engagement de Daniel Bain dans des recherches en didactique.

La didactique

Cet optimisme actif, on le trouve en effet tout particulièrement du côté des didacticiens. Daniel Bain, à côté de l'analyste distant et rigoureux du système scolaire, est en effet aussi un didacticien engagé dans la recherche de démarches susceptibles d'améliorer l'apprentissage des élèves. À côté du psychologue et de l'analyste, il y a l'enseignant, le professeur de lettres, préoccupé de l'apprentissage du français par les élèves, mais aussi intéressé par celui des sciences. Daniel Bain a été simultanément observateur du monde de l'école et acteur de sa transformation, au risque de croiser la contradiction à chaque carrefour. Sur ce chemin, il s'est trouvé pèlerin au côté de Jean-Paul Bronckart et d'André Giordan.

Avec eux, Daniel Bain partage l'avis que le désir et le plaisir de la connaissance et de l'apprentissage ne sont pas d'emblée mobilisés par l'élève, et qu'ils ne suffisent pas à la construction de compétences. L'intervention didactique est fondée, nécessaire, et relève d'un enseignement construit. Si la nécessité de cette intervention apparaît claire et fondée, les modalités de mise en œuvre sont en revanche problématiques et encore largement inexplorées. Jean-Paul Bronckart et André Giordan ont donc travaillé à la conceptualisation et à l'élaboration de stratégies d'enseignement ainsi que de supports et de matériels pour les enseignants et les élèves. Le concept de «séquences didactiques» sur lequel Daniel Bain a investi beaucoup de son énergie et de ses compétences, de même que le modèle «allostérique» d'André Giordan, s'inscrivent dans ces démarches de structuration de l'enseignement/apprentissage par lesquelles les enseignants sont appelés à remplir une fonction de médiateur des apprentissages et de metteur en scène des situations d'acquisition.



Les didactiques du français et des sciences

S'appuyant sur les acquis relatifs à la didactique de la langue, Jean-Paul Bronckart, dans son article, tente d'articuler cette dernière à une didactique de la littérature dont l'objectif, défini par l'auteur, est de contribuer à la formation de la personne et à la construction de sa socialité par une fréquentation de «*la contradiction permanente*» et par la rencontre de l'écriture littéraire, «*témoin des mouvements dialectiques de la pensée humaine (...)*».

Alors que la démarche didactique idéale consisterait à faire débiter l'enseignement de la langue par des activités de lecture et de production de textes, et à articuler ensuite à cette démarche initiale des activités d'inférence puis de codification des régularités observables dans le corpus de textes ainsi mobilisé, Jean-Paul Bronckart considère cette procédure comme inapplicable. Face à ce qu'il considère comme une «*impossibilité sociale*», il envisage une position de compromis mettant en parallèle le traitement d'un corpus de phrases et celui d'un corpus de textes, avec toutefois passage progressif d'une didactique du «système» de la syntaxe phrastique à une didactique du fonctionnement textuel. Cette dernière est présentée alors comme lieu d'enclenchement d'une didactique de la littérature.

Jean-Paul Bronckart pour la didactique de la langue et André Giordan pour la didactique des sciences se réfèrent, tous deux, à des théories communes de l'apprentissage, à savoir les modèles piagétien et vygotskien, selon lesquels l'apprenant procède par construction de systèmes successifs de représentations en référence soit à l'objet soit à autrui, soit encore à tous les deux simultanément. Dans le présent ouvrage, André Giordan met tout spécialement en question le constructivisme piagétien dont ses travaux montrent les limites. Il observe en effet que des adultes, en pleine maîtrise du formalisme logico-mathématique, peuvent raisonner à l'égal d'enfants de 6-7 ans dans des situations ou sur des contenus inhabituels. Dans les cas d'apprentissage complexe, un travail de *déconstruction* des représentations initiales – sorte de «prisons» intellectuelles – est à entreprendre. Une médiation doit alors intervenir qui prend en compte les dimensions sociale et affectivo-émotionnelle de l'élève. Il appartient en effet à l'enseignant-médiateur de filtrer les informations et de favoriser les mises en relation... selon un modèle «*allostérique*» qui «*permet de prévoir les situations, les outils et les ressources qui facilitent l'apprendre*», en évitant toute scolarisation du savoir, menace qui, selon Daniel Bain, pèse sur toute transposition didactique.

À ce stade de l'exposé, il convient de situer une contribution originale de Daniel Bain à la conception de la démarche didactique que les auteurs de cet ouvrage n'ont pas abordée. Il a en effet ouvert un débat stimulant entre les tenants d'une évaluation formative interdisciplinaire et les tenants de l'évaluation intégrée à la démarche didactique disciplinaire. Son intervention au colloque de l'ADMÉE de 1986 où il déclarait que «*l'évaluation formative faisait fausse route*» a marqué le début d'un processus d'intégration progressive de cette forme d'évaluation dans la didactique elle-même. «*La validité d'un dispositif d'évaluation formative sera psychopédagogique ou elle*



ne sera pas», disait-il à ce colloque. «*J'entends par là que la pertinence des items ne pourra être jugée que par rapport à la perspective d'apprentissage adoptée dans une certaine stratégie didactique*». Daniel Bain est en effet non seulement un spécialiste de la didactique, il est également un fin connaisseur des pratiques d'évaluation scolaire, comme il l'est aussi des méthodes statistiques d'évaluation et de mesure.

Une théorie pour affiner la mesure

André Giordan et les didacticiens des sciences relèvent, comme on l'a vu plus haut, que les adultes peuvent avoir des représentations primitives des phénomènes scientifiques, construites sur la base d'évidences ou de généralisations résultant d'un échantillon réduit et aléatoire d'observations. Si la loi que l'on infère de ces observations répétées n'est pas contredite par l'expérience quotidienne, elle est alors appliquée à toutes situations semblables. Cette procédure «courante» d'inférence ne satisfait pas les logiciens, ni d'ailleurs les scientifiques, et par conséquent ni Jean Cardinet ni Daniel Bain. L'un et l'autre avaient besoin pour leurs recherches de méthodes statistiques plus sensibles et de mesures plus assurées.

Daniel Bain, et avec lui Dagmar Hexel et Gianreto Pini, ont donc étroitement collaboré avec Jean Cardinet à l'élaboration et la divulgation de la théorie de la généralisabilité. Par ces travaux, Daniel Bain se révèle tout particulièrement homme de rigueur et de précision, mais aussi pédagogue puisqu'il a collaboré activement à l'élaboration d'un mode d'emploi de cette théorie. Ce travail s'inscrit dans une recherche de construction d'un modèle statistique dont Jean Cardinet présente les derniers développements dans cet ouvrage, et d'un logiciel de traitement de données. Ces instruments conceptuels et de calcul ont pour but de permettre une généralisation «contrôlée» des résultats obtenus à partir d'échantillons de données ou d'individus, et d'assurer une meilleure mesure des apprentissages des élèves par une estimation plus précise des risques d'erreur lors d'enquêtes et de *surveys*. Si Jean Cardinet admet volontiers que ce modèle statistique répond mal à la logique du fonctionnement quotidien d'une classe, il considère toutefois qu'il devrait être utile à tous ceux qui cherchent à comprendre les distorsions du système éducatif et à jeter un regard lucide sur l'enseignement et ses effets.

Daniel Bain qui a fort bien perçu les limites méthodologiques des grandes enquêtes internationales – il en a fait d'ailleurs une critique sévère en 1998 – a donc entrepris, avec Jean Cardinet, le développement de cette théorie statistique afin de disposer de meilleurs instruments de mesure. De telles enquêtes peuvent en effet apporter aux enseignants des informations utiles sur et pour leurs élèves et leur classe. Elles permettent surtout aux pouvoirs organisateurs de rendre compte de la qualité et des résultats des systèmes de formation dont ils sont responsables, ainsi que de rechercher les moyens nécessaires à leur amélioration.



Armin Gretler a une claire conscience de l'importance de ces enquêtes pour le pilotage de l'école, sans pour autant en ignorer les limites. Dans son article, il décrit avec précision l'émergence des recherches internationales, leurs caractéristiques, et les avantages que l'on peut en tirer. Il montre notamment l'intérêt national, mais aussi cantonal et local, que peuvent avoir ces grandes enquêtes; il rejoint ainsi, dans une certaine mesure, la position de Daniel Bain. Il importe en effet que les données obtenues puissent être exploitées localement et avoir une incidence sur les politiques de formation du pays et/ou des cantons. Les retombées pour la Suisse et les différents systèmes de formation seront par conséquent d'autant plus efficaces que les chercheurs suisses auront pu intervenir dans la conception et la réalisation de ces enquêtes.

Une invitation au voyage dans un espace multidimensionnel

Nous avons vu l'extraordinaire richesse et diversité des intérêts et des travaux de Daniel Bain. Nous en avons examiné plusieurs facettes, dirait un «généralisabilien». Nous avons saisi la polyvalence de ses compétences au travers des travaux de chercheurs d'horizons très différents avec lesquels il a travaillé et qui ont apprécié sa collaboration.

Mon rôle de préfacier est d'inviter les lecteurs à le découvrir à leur tour au travers des textes de cet ouvrage. Ces écrits abordent des thèmes chers à Daniel Bain et rendent hommage à une vie professionnelle toute consacrée à la recherche, et par elle, à l'élève. Daniel Bain a accompli ce travail avec modestie, sans donner à ses résultats plus d'importance que les conditions de la recherche ne le permettaient. Cette attitude profondément scientifique ne surprendra pas tous ceux qui ont eu le bonheur et la chance de travailler avec lui.



Les enseignants, leurs croyances, leurs pratiques d'évaluation et l'échec scolaire

Analyse des réussites et des échecs dans quatre classes du secondaire

Marcel Crahay

Les enseignants croient-ils en l'efficacité du redoublement?

Incontestablement, les enseignants des régions francophones (Communauté française de Belgique, Genève, etc.) recourent régulièrement au redoublement. Les statistiques sont éclairantes à ce sujet; l'importance des retards scolaires singularisent ces régions par rapport à la plupart des autres systèmes d'enseignement de l'Europe occidentale (Crahay, 1996). Il est, par conséquent, légitime de traiter ce phénomène comme un fait sociologique et, partant, de parler de *culture de l'échec* (Crahay, 1996) ou d'*idéologie de l'excellence* (Perrenoud, 1984).

Les études menées par Pini (1991) à Genève, de Landsheere (1993) puis Stegen (1994) en Communauté française de Belgique (cités par Crahay, 1996) aident à cerner les représentations des enseignants qui sous-tendent le recours régulier au redoublement¹. Pour la majorité des enseignants, qu'ils soient genevois ou belges, qu'ils enseignent au niveau primaire ou secondaire, le redoublement joue un rôle essentiellement positif et ce, en vertu de deux mécanismes complémentaires:

- bénéficiant d'une année supplémentaire, certains élèves faibles ont l'occasion de mûrir et de mieux se préparer à affronter les difficultés de leur scolarité future;
- en répétant une année, les élèves faibles peuvent parcourir une deuxième fois la totalité du programme.

¹ L'intérêt de ces trois études vient de ce qu'elles portent sur les mêmes items (Crahay, 1996). Pini avait conçu une série de 19 propositions vis-à-vis desquelles il a invité 92 enseignants primaires du canton de Genève à exprimer leur accord ou désaccord. De Landsheere (1993) a repris 13 de ces questions et les a présentées à 450 enseignants de 1^e et 2^e années primaires de la Communauté française de Belgique. Stegen (1994) a repris 10 de ces questions et les a soumises à 263 enseignants secondaires de la même Communauté française.

Bref, tout semble indiquer que les enseignants de ces régions croient aux vertus pédagogiques du redoublement.

Peut-on conclure de ces observations que les enseignants voient dans le redoublement un outil de remédiation dont pourraient profiter les élèves faibles²? Ce n'est pas évident. Si tel était le cas, ceux-ci devraient se montrer nettement plus positifs dans leurs évaluations des élèves redoublants. Or, de nombreuses études révèlent que les élèves qui ont redoublé une fois subissent très souvent de nouveaux revers dans la suite de leur carrière scolaire. Nous savons que le redoublement procède essentiellement d'une décision des enseignants (Hutmacher, 1993; Grisay, 1984). On aurait donc pu imaginer que les enseignants, farouches adeptes du redoublement, soient tentés d'en montrer l'efficacité en favorisant ceux qui en ont été les bénéficiaires. Manifestement, il n'en est rien. Au contraire, il semble que les enseignants se montrent plus sévères que les tests à l'égard des élèves redoublants. C'est ce qui ressort de la méta-analyse de Holmes (1990). L'amplitude de l'effet du redoublement est de $-0,31$ lorsque les performances académiques de ces élèves sont mesurées à l'aide d'épreuves standardisées; elle est de $-0,78$ lorsqu'on prend pour référence les épreuves des enseignants.

A cet égard, la contribution de Bain (1988) est cruciale. Pour lui, *«une façon de vérifier les vertus et avantages pédagogiques du redoublement serait de montrer qu'à la fin de l'année répétée, les élèves ont fortement progressé, sont à nouveau à flot et peuvent repartir d'un bon pied (...). Le bénéfice attendu du redoublement est rarement défini précisément. On s'accordera cependant à admettre que le progrès doit être substantiel pour qu'on estime que le jeu en valait la chandelle»* (p. 64). A Genève, où la recherche a été menée, les notes varient entre 0 et 6; la barre du suffisant est à 2,5 à l'école primaire (au Cycle d'orientation, ceci n'est pas le cas). Conformément à l'interprétation officielle, Bain fixe à 4 la limite du satisfaisant; c'est aussi la note minimale pour passer en 7^e dans la filière supérieure (section latine-scientifique). Il considère qu'entre 3 et 3,9, les résultats restent médiocres.

Le constat est désolant: en grammaire, orthographe et mathématiques, un élève sur 6 ou sur 7 seulement obtient une note supérieure à 4; en composition française (qui n'est que rarement cause d'échec dans l'école genevoise), un tiers des élèves franchit le seuil de 4 (tableau 1, extrait de Bain, 1988, p. 65). Quant aux autres, ils ne dépassent pas la zone de performances relativement médiocres (c) et ceci alors qu'ils répètent l'année.

²C'est l'interprétation que, nous-mêmes, nous donnions aux résultats de ces enquêtes (Crahay, 1996, p. 116).



Tableau 1: Evolution des redoublants de 6^e de 1981 à 1982.
Elèves insuffisants dans la branche. Notes

Evolution	Composition	Grammaire	Orthographe	Mathématiques
a) Sont restés insuffisants ($n < 2,5$)	(0%)	21%	38%	0%
b) Sont restés proches de la limite ($2,4 < n < 3,1$)	(67%)	33%	33%	33%
c) Ont des résultats encore médiocres ($3,1 < n < 3,9$)	(0%)	29%	14%	53%
d) Ont des résultats satisfaisants ($n > 3,9$)	(33%)	17%	14%	13%
Effectifs	100% 3	100% 24	99% 42	99% 30

n: notes

Plus intéressante encore est la seconde analyse conduite par Bain. Puisque le redoublement constitue une mesure globale, certains élèves en échec en raison de lacunes dans une branche, doivent refaire, dans les autres branches, un programme qu'ils avaient acquis, si l'on en juge par les notes qu'ils avaient obtenues. Bain s'intéresse dès lors à l'évolution des redoublants dans ces branches en principe maîtrisées.

Comme le montre le tableau 2 (extrait de Bain, 1988, p. 66), moins d'un tiers des élèves ont progressé de façon appréciable (progrès égal ou supérieur à une unité, ligne d). La majorité n'ont pas ou guère progressé (c) et certains ont même régressé (b).

Tableau 2: Evolution des redoublants de 6^e de 1981 à 1982.
Elèves suffisants dans la branche. Notes

Evolution	Composition	Grammaire	Orthographe	Mathématiques
a) Sont devenus insuffisants ($n_{82} < 2,5$)	1%	2%	3%	2%
b) Ont régressé ($n_{82} < n_{81}$)	25%	20%	5%	11%
c) N'ont pas ou guère progressé ($n_{82} - n_{81} < 1,0$)	60%	46%	69%	65%
d) Ont progressé de façon appréciable ($n_{82} - n_{81} \geq 1,0$)	14%	32%	23%	22%
Effectifs	100% 77	100% 41	100% 39	100% 54

n_{81} : notes de 1981 / n_{82} : notes de 1982



Ces données montrent que les enseignants sont directement confrontés à l'inefficacité du redoublement et traduisent ce fait dans des notes par lesquelles ils signifient que l'année répétée par la majorité des élèves est bel et bien une année perdue.

Et si les enseignants ne croyaient pas dans la capacité d'apprendre de leurs élèves...

Mais alors pourquoi les enseignants font-ils redoubler des élèves?

Pour expliquer ce qui peut apparaître comme un paradoxe, il faut en revenir au concept d'idéologie de l'excellence. Pour Perrenoud (1984), l'école assume la fonction d'instituer des normes d'excellence et de classer les individus en fonction de celles-ci, créant ainsi des hiérarchies. Elle participe ainsi au fonctionnement général de nos sociétés occidentales, au sein desquelles les individus sont classés les uns par rapport aux autres en fonction de leur expertise dans un domaine, de leurs performances sportives, de leur beauté physique, etc. L'école n'a donc pas le monopole de la fabrication de hiérarchies d'excellence. Dans la vie quotidienne, les jugements d'excellence vont bon train. «Monsieur Tout le monde» est friand des appréciations quant aux qualités des uns et des autres. L'excellence scolaire présente, toutefois, une spécificité importante: elle s'applique à tous les enfants et, par voie de conséquence, à tous les adultes d'une société.

Chez beaucoup d'enseignants, cette idéologie de l'excellence se combine avec des conceptions psychologiques de type innéiste. La capacité de faire des études n'est pas donnée à tout le monde. Autrement dit, il existerait une (ou des) aptitude(s) intellectuelle(s) qui prédispose(nt) à réussir à l'école et, plus tard, à l'université. Si le redoublement peut s'avérer positif, c'est parce qu'il laisse la maturation faire son œuvre. Comme indiqué ci-dessus, environ 90% des enseignants primaires interrogés par Pini (1991) et de Landsheere (1993, cité par Crahay, 1996) concèdent que l'année supplémentaire dont certains élèves ont besoin leur permet de mûrir. Au secondaire, ils sont 78% à s'accorder avec cette idée. Dans les interviews approfondis menés par Burdevet (in Crahay, 1996) auprès de onze enseignants primaires genevois, nombreuses sont les réponses qui font référence à la maturation des enfants. On est, en définitive, tenté de supposer qu'aux yeux d'une majorité d'enseignants, le fait de parcourir une seconde fois la totalité du programme n'a de pertinence que dans la mesure où la maturation est susceptible de produire ses effets.

A la lumière de cette réinterprétation des enquêtes menées antérieurement, on comprend mieux le fatalisme des enseignants en matière de remédiation. Si le développement cognitif est essentiellement une affaire de maturation, la pédagogie ne peut pas grand-chose. En tout cas, l'efficacité de l'action éducative est subordonnée à l'œuvre de la nature: une explication, une leçon ou encore une remédiation ne peut engendrer l'apprentissage qu'à condition que l'esprit de l'élève soit prêt à la recevoir.



Rien d'étonnant alors si bon nombre d'enseignants ne considèrent pas le redoublement comme un échec de leur enseignement. Pas surprenant non plus si les enseignants refusent de s'attribuer la responsabilité des échecs et s'ils ne s'émeuvent pas à l'observation d'élèves qui ne progressent pas durant l'année redoublée. De même, l'hypothèse formulée ici est compatible avec le fait qu'une majorité d'enseignants considèrent que le redoublement est inévitable dans la mesure où l'échec est principalement dû à des facteurs extra-scolaires (57% à Genève et 53% en Communauté française de Belgique), et doutent qu'il soit possible d'éviter les redoublements en s'y prenant de façon adéquate (65% d'une part et 56% d'autre part).

Sans doute, le positionnement des enseignants face à l'échec des élèves procède-t-il de mécanismes psychosociaux complexes. Au terme d'une enquête menée en France, Gosling (1992) constate que les 320 enseignants du secondaire interrogés expliquent l'échec des élèves de façon très différente de leur réussite. Pour cet auteur, il faudrait compter avec une réaction de *déresponsabilisation* mobilisée par les enseignants lorsqu'ils sont invités à expliquer le pourquoi de l'échec de leurs élèves. En quelque sorte, face à un élève qui réussit, l'enseignant s'interroge sur l'action pédagogique qu'il peut avoir pour favoriser son épanouissement intellectuel, tandis que, face à l'élève qui échoue, il se confronte à la question «*qui est responsable?*» et tend à se disculper. Parallèlement, cet auteur relève de nombreuses réponses faisant référence à l'idéologie du don, renouant ainsi avec l'analyse avancée par Carugati et Mugny (1985).

S'il est vrai qu'un nombre important d'enseignants conservent une représentation naturaliste de l'intelligence, faisant de celle-ci une caractéristique individuelle stable et donc inchangeable, il est difficile d'imaginer une pédagogie de lutte contre l'échec qui puisse être efficace. Par ailleurs, si des enseignants véhiculent cette représentation de l'intelligence, certaines pratiques d'évaluation dénoncées par la plupart des pédagogues prennent sens. Si l'aptitude à apprendre est un attribut génétique de la personne comme l'est la taille, il est naturel de rencontrer des élèves intelligents et d'autres qui ne le sont pas, comme il est naturel de côtoyer des individus de grande taille et d'autres de petite taille. Cette observation de sens commun, les enseignants ont, en quelque sorte, pour mission institutionnelle de la consigner dans les bulletins ou livrets scolaires. A leurs yeux, l'évaluation ne serait rien d'autre qu'une activité perceptive au cours de laquelle ils constatent simplement et – croient-ils – en toute objectivité que certains sont aptes à poursuivre des études tandis que d'autres en sont incapables. Les premiers seraient promus et les seconds recalés, c'est-à-dire contraints au redoublement. Bref, l'évaluation procéderait d'une *logique de la prédestination* (Neave, 1977) et aurait pour fonction de «*séparer le bon grain de l'ivraie*» et il n'y aurait aucun sens à poser la problématique du redoublement en termes de vertu remédiate ou d'efficacité pédagogique.

Ce positionnement psychosocial des enseignants est, en définitive, sollicité par l'institution. Fin du primaire dans certains systèmes, début du secondaire dans d'autres, les enseignants ont pour fonction de participer à l'orientation des élèves dans une filière de type technique ou professionnelle ou, à l'inverse, dans une filière qui assure



la transition avec l'enseignement supérieur. Logiquement, c'est auprès des enseignants opérant à ce niveau du cursus scolaire que l'on devrait rencontrer le plus souvent le mode de fonctionnement décrit ci-dessus. On peut encore anticiper une différence entre les enseignants qui officient dans les classes qui se situent résolument dans la perspective des filières nobles et ceux qui pratiquent leur art dans des classes décalées par rapport à ces filières.

Ce sont ces différentes supputations que nous avons essayé de confronter aux faits au cours d'une étude exploratoire menée dans une école de la région liégeoise. Plus précisément, il s'agit de montrer que, par leur stabilité ou leur stéréotypie, les pratiques d'évaluation d'enseignants officiant au début du secondaire s'accordent avec la représentation de l'intelligence comme un trait stable. Fixant leur jugement relatif aux différents élèves de leurs classes dès le début d'année, les enseignants sont aveugles face aux progrès de certains.

La sélection est opérée dès la constitution des classes

En Communauté française de Belgique, les deux premières années de l'enseignement secondaire sont des années d'orientation: en cours de route ou à leur terme, les élèves sont maintenus dans l'orientation générale (filière noble) ou sont détournés vers les filières techniques ou professionnelles. Dès la première année, les élèves peuvent choisir l'option *latin* ou non; dans ce cas, ils sont dirigés vers la section *moderne*. L'option *latin* profile les étudiants dans la droite ligne de la filière noble de 3^e.

Dans la présente étude³, on s'attachera à étudier les pratiques d'évaluation de deux enseignants de français dans quatre classes de première année de l'enseignement secondaire. Pour ce faire, on recourra à l'occasion à la comparaison des notes scolaires avec les résultats obtenus par les élèves de ces quatre classes à deux épreuves externes: la première organisée dès le début d'année (fin septembre) et l'autre en fin d'année (début juin).

Les quatre classes sont issues du même établissement. Les deux premières regroupent des élèves qui ont choisi l'option *latin*; elles sont désignées par les sigles 1LA et 1LB. Les deux autres – IMA et IMB – relèvent de l'orientation *moderne*. La logique de la prédestination, que nous supposons à l'œuvre dans la quotidienneté du fonctionnement scolaire, devrait se manifester dès ce moment: les meilleurs élèves devraient se retrouver dans les classes avec l'option *latin* et les autres dans les classes avec l'option *moderne*. Le tableau 3 fournit une première confirmation de notre hypothèse.

³ Les données présentées ici ont été recueillies par Bounafaa (1992) dans le cadre de son mémoire de licence en Sciences de l'éducation (Université de Liège). Elles ont été complètement retraitées pour les besoins de cet article.



Les effectifs de classe varient entre 18 et 23 élèves. La répartition entre les sexes est similaire dans les quatre classes; cette impression est confirmée par le calcul du chi-carré (34,369)⁴. En revanche, il apparaît que les quatre classes se différencient quant au nombre d'élèves étrangers et de sujets qui ont déjà redoublé. Ces impressions sont à nouveau confirmées par le calcul du chi-carré: en ce qui concerne le premier de ces deux paramètres, sa valeur est de 12,338; elle est de 9,206 en ce qui concerne le second.

Tableau 3: Composition des quatre classes (effectif, sexe, Belges *versus* étrangers UE/hors UE, ont doublé *versus* n'ont jamais doublé)

Désignation	1 LA	1 LB	1 MA	1 MB
Effectifs	22	23	18	23
Garçons	8	12	12	14
Filles	13	11	6	9
Belges	18	20	12	17
Etrangers UE	0	3	1	0
Etrangers hors UE	4	0	5	6
N'ont jamais doublé	20	19	10	14
Ont déjà doublé	2	4	8	9

Il est manifeste que les classes 1MA et 1MB (soit les deux classes dites *modernes*) comportent un plus grand nombre d'élèves étrangers hors UE (plus d'un quart de l'effectif) que les classes avec l'option *latin* (4 sur 22 dans la classe 1LA et aucun dans la classe 1LB). Les proportions d'élèves étrangers hors UE sont éloquentes: 1LA = 0,18; 1LB = 0; 1MA = 0,27; 1 MB = 0,26. Par ailleurs, les deux classes avec l'option *latin* se caractérisent en sus par un faible nombre d'élèves qui ont redoublé au cours de leur enseignement primaire (2 dans la classe 1LA et 4 dans la 1LB, pour 8 dans la 1MA et 9 dans la 1MB). Les classes dites *modernes* comportent plus d'un tiers de redoublants et les *latines* moins d'un cinquième. Tout ceci laisse penser que l'option *latin* est l'occasion de regrouper les élèves de nationalité belge (et étrangers UE) qui n'ont pas redoublé au cours de l'enseignement primaire.

Par ailleurs, les élèves des quatre classes ont été soumis, en début d'année, à un test de connaissances. Ce test (dont le score maximal est de 136) porte sur les compétences en langue maternelle qui doivent être maîtrisées en fin de sixième année primaire. On peut dès lors comparer les quatre classes du point de vue des compétences maîtrisées par les élèves au moment d'entamer l'enseignement secondaire. On ne s'étonnera pas d'apprendre que les moyennes obtenues dans les classes *latines* (1LA

⁴Dans ce texte, les statistiques sont utilisées à des fins descriptives (et non inférentielles); on se contentera donc d'indiquer les valeurs des indices sans faire référence au seuil de probabilité correspondant.



= 108,903; 1LB = 103,209) sont supérieures aux classes *modernes* (1MA = 90,526; 1MB = 88,287). Une analyse de variance appliquée aux scores obtenus par les élèves des quatre classes à ce test met en évidence une différence nette entre les classes *latines* et *modernes* ($F = 7,32$).

Ce dernier constat n'a rien d'étonnant. La littérature de recherche a montré à satiété (Crahay, 1996) que les élèves redoublants se caractérisent par un niveau de compétence moyennement inférieur à celui des autres élèves. C'est également le cas dans les quatre classes de l'établissement sous étude: 23 élèves ont redoublé une année (ou deux) du primaire, la moyenne de leurs scores au test de début d'année est de 86,82 alors que la moyenne des scores des 60 élèves qui n'ont jamais redoublé vaut 102,17; la différence est appréciable ($F = 12,86$).

Les enseignants de l'établissement concerné par notre étude ont une connaissance intuitive des constats évoqués ci-dessus. Selon les dires du professeur de français responsable des classes 1LB et 1MB, «un élève qui a déjà redoublé au primaire, c'est plutôt mauvais signe». Ou encore, «on ne va pas envoyer en latine un élève qui a déjà eu des difficultés en primaire». Le professeur responsable des classes 1LA et 1MA est plus nuancé. Il concède: «il est rare qu'un élève qui a déjà redoublé au primaire n'ait pas de problème au secondaire», mais ajoute «il ne faut pas être fataliste». Cet enseignant n'est évidemment pas seul à décider de l'affectation des élèves dans les classes. Son rôle, comme celui de ses collègues, consiste à recevoir les élèves avec leurs parents au moment de l'inscription, à prendre connaissance de leur parcours scolaire antérieur et à suggérer une option. Par ailleurs, des recommandations ont été formulées par le chef d'établissement: il faut «tenir compte du potentiel des élèves pour composer les classes»; en particulier, il faut «éviter d'imposer le latin à des élèves qui ont déjà eu des difficultés en primaire». En définitive, sous le couvert de justifications psychopédagogiques (tenir compte du potentiel des élèves), la logique de la prédestination se met en place, dans cet établissement secondaire, dès l'accueil des élèves: au moment de l'inscription des élèves, les enseignants examinent le parcours des uns et des autres et établissent des pronostics, négatifs pour ceux qui ont redoublé et neutres ou positifs pour les autres⁵. Ainsi sont créées des classes de niveau pour le meilleur, mais surtout pour le pire...

Logique de la prédestination et effet de stéréotypie

Si les enseignants sont convaincus de la stabilité des capacités cognitives des élèves, ils sont susceptibles d'ignorer les coups d'éclat des plus faibles et les accidents des

⁵Nous avons vécu ce processus d'orientation de l'intérieur en tant que parent. Audrey ne voyait pas l'utilité du latin. Au moment de son inscription dans un établissement de l'enseignement secondaire, l'enseignant qui l'accueillait lui a expliqué qu'avec les bons résultats qu'elle avait obtenus au primaire, elle devait faire du latin au moins pendant la première année... le temps que la sélection soit opérée.



plus forts. Autrement dit, nous supposons que les enseignants sont enclins à l'effet de stéréotypie afin d'assurer l'homéostasie de leur système de représentations.

L'examen des notes attribuées par les enseignants de français dans les quatre classes considérées ici révèle une grande stabilité en cours d'année. Nous avons, en effet, pu recueillir les notes inscrites dans les livrets scolaires (ou bulletins) des 83 élèves à six évaluations périodiques: octobre, décembre, février, mars, mai et juin. Les résultats aux examens de français de Noël et de fin d'année ont également été pris en considération. Lorsqu'on calcule des corrélations entre toutes ces séries de notes, on obtient la matrice présentée au tableau 4.

Tableau 4: Corrélations entre les notes de français aux différentes périodes de l'année scolaire pour l'ensemble des classes

	Décembre	Examen de Noël	Février	Mars	Mai	Juin	Examen de fin d'année
Octobre	0,62	0,74	0,65	0,64	0,59	0,64	0,80
Décembre		0,76	0,67	0,71	0,69	0,68	0,66
Noël			0,76	0,77	0,77	0,80	0,84
Février				0,80	0,82	0,91	0,81
Mars					0,88	0,82	0,81
Mai						0,93	0,80
Juin							0,87

Les corrélations des notes de français sont toutes supérieures à 0,60 et la plupart à 0,80. Les notes attribuées par les enseignants en octobre prédisent de façon forte celles des examens de fin d'année scolaire (0,80, soit 64% de la variance). A partir de février, les choses semblent figées: les corrélations recensées dans la moitié inférieure du tableau sont toutes supérieures à 0,80.

Ce sont les évaluations de décembre qui sont les moins sûres en matière de prédiction: les corrélations de notes de cette période avec les autres se situent entre 0,60 et 0,75; leur corrélation avec les notes de fin d'année correspond seulement à 0,65 (soit 42% de la variance). Les enseignants expliquent aisément ce constat: en décembre, ils sous-évaluent délibérément les étudiants afin de mettre les élèves sous pression; en particulier, les bons élèves qui seraient «tentés de se reposer sur leurs lauriers».

La stéréotypie des notes attribuées aux élèves est peut-être tributaire de la réalité des choses. Il se peut, en effet, que les performances scolaires des 83 élèves de notre étude soient restées stables tout au long de l'année. Ce serait donc faire un mauvais procès aux enseignants que de parler de logique de la prédestination ou d'effet de stéréotypie. Pour mettre notre hypothèse à l'épreuve des faits, il suffit de comparer



l'évolution des notes attribuées par les enseignants à l'évolution des scores obtenus par les 83 élèves à une épreuve externe de français. C'est ce que nous sommes en mesure de faire puisqu'un test de français a été présenté dans trois des quatre classes en début et en fin d'année⁶.

En début d'année, les compétences des élèves en français mesurées par le test varient en fonction de l'appartenance à un type de section. Elles varient également selon que les élèves ont ou non redoublé en primaire et selon qu'ils sont belges ou étrangers. Qu'en est-il en fin d'année? Le tableau 5 permet d'ébaucher la réponse à cette question puisqu'il présente les moyennes par classe aux tests de début et de fin d'année (ainsi que les pourcentages de réussite et les résultats d'une analyse de variance visant à cerner l'effet-classe).

Tableau 5: Moyenne des scores par classe aux tests de français (début et fin d'année) et pourcentage de réussite (fixée à 60% des points)

Classes	Début d'année		Fin d'année	
	(max.= 136)	% de réussite	(max.= 144)	% de réussite
ILA	108,9	80,07	102,7	70,62
1LB	103,2	75,88	101,7	71,29
1MA	90,5	66,56	94,9	65,91
F	5,65		1,11	

Manifestement, les différences de résultats entre classes tendent à s'estomper en cours d'année; les progrès relatifs de la classe moderne (1MA) sont plus importants que ceux des deux classes latines, ce qui se traduit au niveau du F qui est nettement plus bas lorsqu'on compare les moyennes en fin d'année (1,11) et en début d'année (5,65).

Par ailleurs, lorsqu'on compare les scores des élèves qui ont redoublé à ceux des autres, on aboutit à un constat analogue. En début d'année, l'écart était important; il s'est atténué en fin d'année. Les données présentées dans le tableau 6 en attestent⁷.

⁶Les élèves des deux classes latines et d'une seule classe moderne ont été soumis à un test de connaissances en langue maternelle au mois de juin. Ce test porte sur les compétences qui doivent être maîtrisées en fin de première année secondaire. C'est pour des raisons de disponibilité d'horaire que les élèves de la classe 1MA n'ont pu être testés.

⁷Notons que, dans ce cas-ci et a fortiori dans le tableau suivant, la comparaison et, plus encore, l'analyse de variance est délicate vu le petit nombre de sujets et surtout la répartition inégale dans les cellules.



Tableau 6: Moyenne des scores des élèves redoublants *versus* les autres aux tests de français (début et fin d'année) et pourcentage de réussite

	Début d'année		Fin d'année	
	N	Moyenne	N	Moyenne
Redoublants	14	88,66	13	95,1
Les autres	46	105,55	46	101,99
F	10,66		1,88	

En début d'année, les scores au test se distinguaient nettement selon la nationalité des élèves. Il n'en est plus vraiment de même en fin d'année, comme le montrent les données du tableau 7.

Tableau 7: Moyenne des scores des élèves belges *versus* étrangers UE *versus* étrangers hors UE aux tests de français (début et fin d'année) et pourcentage de réussite

	Début d'année		Fin d'année	
	N	Moyenne	N	Moyenne
Belges	47	106,76	48	102,4
Etrangers UE	4	91,75	4	100,5
Etrangers hors UE	9	79,2	9	87,3
F	13,05		2,87	

Même si les analyses statistiques présentées dans les tableaux qui précèdent sont précaires, un constat se dégage avec netteté: dans cette école, un processus d'égalisation des acquis en français s'est opéré au cours de la première année d'enseignement secondaire. En début d'année, les élèves étrangers, ayant redoublé et dotés de moindres connaissances en français se sont principalement retrouvés en moderne. Si l'on en juge par les analyses effectuées sur une classe de section moderne et deux de section latine, les inégalités de compétences se sont estompées en fin d'année. Comment? Est-ce le fait d'une démarche délibérée des enseignants? Probablement pas, puisque les notes des enseignants de français attribuées à ces mêmes élèves sont restées stables tout au long de l'année. Bref, il semble bien que les enseignants sont restés insensibles aux évolutions de ceux qui démarraient dans les conditions les plus difficiles.

L'évaluation des enseignants serait-elle aveugle aux progrès des élèves?

Le verdict murmuré ci-dessus est grave. Il importe de l'étayer. Pour ce faire, il a été fait appel à l'analyse de régression. Cette analyse statistique a été appliquée à trois reprises afin de mettre en relation:



- les scores aux deux tests (régression 1),
- les scores au test de début d'année et les notes de fin d'année (régression 2),
- les scores au test de fin d'année et les notes de fin d'année (régression 3).

Ce type d'analyse permet d'estimer dans quelle mesure les résultats au test de début d'année prédisent les résultats au test de fin d'année d'une part, les notes de fin d'année d'autre part. Il sera également possible d'estimer dans quelle mesure les résultats au test de fin d'année prédisent les notes de fin d'année.

La première équation de régression est essentielle à notre démonstration. Si les positions relatives des performances des élèves n'ont pas bougé entre le premier et le second test, le coefficient de régression sera élevé. Ceci signifierait que le classement des élèves entre eux est resté stationnaire en cours d'année et ceci malgré le fait que les écarts entre leurs scores se sont estompés (cf. les analyses précédentes). La stéréotypie des évaluations des enseignants s'expliquerait alors aisément: les progrès des élèves étant symétriques, il serait malaisé de les apercevoir. En revanche, si nos craintes sont fondées c'est-à-dire s'il est vrai que les enseignants sont aveugles aux progrès de leurs élèves faibles, le coefficient de régression devrait être relativement faible.

Si nos hypothèses sont correctes, la seconde équation devrait aboutir à un coefficient de régression élevé. En effet, si les évaluations des enseignants sont profondément marquées par l'état initial des élèves, les performances de ceux-ci au test de début d'année devraient permettre de prédire les notes de fin d'année de façon relativement sûre. Enfin, toujours si nos hypothèses sont exactes, la troisième analyse de régression devrait déboucher sur un coefficient plutôt faible, signe que la relation entre les performances des élèves au test de fin d'année et les notes des enseignants est peu solide.

Les résultats obtenus vont globalement dans le sens de nos hypothèses. La première équation aboutit à un coefficient de régression de 0,49 ($F = 53,43$), la seconde à un coefficient de 0,69 ($F = 176,11$) et la troisième à un coefficient de 0,49 ($F = 55,77$). Autrement dit, les scores au test de début d'année prédisent 24% de la variance des scores au test de fin d'année, alors qu'ils prédisent 48% de la variance des notes de fin d'année. Quant aux scores au test de fin d'année, ils prédisent seulement 24% de la variance des notes de fin d'année. On constate donc que la liaison la plus forte est celle qui unit le test de début d'année et les examens de fin d'année. Ceci est compatible avec l'hypothèse selon laquelle les évaluations des enseignants sont assez fortement déterminées par l'état initial des élèves. Quant à la faiblesse de la relation entre le test d'entrée et celui de sortie, elle est cohérente avec les constats établis ci-dessus. En définitive, il apparaît que les résultats de l'évaluation interne continuent à refléter étroitement les différences de départ alors que les inégalités de compétences (mesurées par les tests) se sont estompées.



Pour estimer les progrès que chaque élève a pu réaliser entre les deux évaluations externes, on peut recourir d'une seconde manière à l'équation de régression. En effet, l'équation obtenue permet de calculer pour chaque élève un score (ou une note) prédit(e) à partir d'une première mesure et de comparer ce score (ou cette note) prédit(e) au score (ou à la note) observé(e). La différence entre les deux scores est appelée résidu ou score résiduel. Si le résidu est positif, cela signifie une progression plus importante que la prédiction. A l'opposé, si le résidu est négatif, on peut en déduire une progression inférieure à la prédiction établie à partir du profil moyen d'évolution de l'ensemble des élèves.

En calculant la moyenne des résidus par classe, on peut obtenir une estimation de l'effet-classe. Cette analyse, appliquée aux scores résiduels obtenus en mettant en relation les résultats aux deux tests, aboutit à montrer le sous-rendement d'une classe latine. En effet, la moyenne des scores résiduels de la 1LB est de $-37,76$ alors qu'elle est de $+1,913$ pour la 1LA et $+1,19$ pour la 1MA. Rappelons que ces deux classes bénéficient du même professeur de français.

Les résultats de cette analyse sont présentés de façon plus précise dans le tableau 8. Les valeurs résiduelles sont réparties dans des classes d'intervalle et le nombre de sujets qui se situent dans chaque intervalle est indiqué pour chacune des trois classes.

Tableau 8: Distribution des scores résiduels (calculés à partir du test d'entrée et du test de sortie) dans les trois classes

Intervalles	Classes		
	1LA	1LB	1MA
En deçà de -20	1	3	0
-19 à -10	3	1	2
-9 à $+9$	16	11	9
$+10$ à $+19$	1	6	3
Au delà de $+20$	0	1	1
Effectifs ⁸	21	22	15

Ce tableau est éclairant quant à la répartition des sujets dans les marges d'intervalle. Les classes se singularisent quant à leur dispersion respective. Les scores résiduels de la classe 1LB se situent principalement dans la zone de -9 à $+9$. Quatre sujets sur 21 obtiennent des scores inférieurs à ce qu'on pouvait attendre d'eux, mais un seul se retrouve en deçà de -20 . La classe 1LB se distingue de la précédente par une disper-

⁸ Certains étudiants étaient absents lors de la passation du test de fin d'année. Les effectifs de ce tableau ne sont donc pas équivalents à ceux du tableau 3.



sion importante des élèves: la moitié d'entre eux se retrouvent dans l'intervalle médian (-9 à +9); sept se caractérisent par un score résiduel supérieur à +10 et quatre obtiennent des scores inférieurs à la prédiction (dont trois en deçà de -20). Dans cette classe, la différence initiale entre les élèves a été fortement amplifiée. Fait notable: dans la classe 1MA, aucun élève ne se trouve dans l'intervalle en deçà de -20; deux seulement obtiennent un score résiduel situé entre -19 et -10.

Eu égard à la stéréotypie des évaluations des enseignants, il n'est pas étonnant de constater que la majorité des élèves mis en échec en fin d'année l'étaient déjà en octobre. Le tableau 9 est éloquent à cet égard. Dans les classes 1LA, 1LB et 1MB, les élèves qui éprouvaient des difficultés en début d'année sont en échec en français en fin d'année. Dans la classe 1MA, quatre élèves qui étaient en difficulté en début d'année se sont *sauvés* en cours d'année.

Tableau 9: Elèves en échec en début et en fin d'année dans chacune des classes

	1LA	1LB	1MA	1MB	Total
Echecs en octobre	2	2	8	5	17
Echecs en fin d'année	2	2	4	5	13

Etudions de plus près le profil de huit élèves mis en échec en fin d'année, dans les classes 1LA, 1LB et 1MA. Dans le tableau 10, on trouve, pour chacun d'eux, leurs notes (en pourcentages et en notes réduites) au bilan des enseignants ainsi que leurs scores aux tests de début et fin d'année. On peut également y lire le score résiduel.



Tableau 10: Résultats d'élèves menacés de redoublement au bilan des enseignants et à deux tests de français (début et fin d'année). Les notes réduites sont entre parenthèses.

Elèves	Classes	Evaluations de l'enseignant	Test de début d'année	Test de fin d'année	Scores résiduels
E5	1LA	22% (-3,43)	43% (-2,33)	39% (-2,77)	-14,24
E6	1LA	44% (-1,84)	64% (-0,79)	59% (-0,99)	-3,79
E33	1LB	58% (-0,83)	58% (-1,21)	52% (-1,54)	-8,34
E35	1LB	56% (-0,96)	76% (0,14)	69 (-0,07)	-2,05
E50	1MA	53% (-0,74)	70% (-0,32)		
E53	1MA	47% (-1,63)	45% (-2,2)	58% (-1,01)	+12,55
E54	1MA	47% (-1,63)	47% (-2,1)	55% (-1,33)	+6,03
E60	1MA	42% (-1,99)	56% (-1,4)	51% (-1,65)	-7,95

Tous ces élèves se caractérisent par des notes faibles à l'évaluation de l'enseignant (moins de 60%), mais aussi par des scores bas aux tests; pour ces trois paramètres, les notes réduites (notes Z) sont négatives. Il convient, toutefois, de remarquer le cas du sujet E6 et surtout celui de E35 dont les résultats au test de fin d'année ne sont pas catastrophiques. En fait, l'examen minutieux des données montre que les élèves menacés de redoublement⁹ sont ceux qui, dans leurs classes respectives, ont les notes les plus basses au bilan de l'enseignant. L'évolution des élèves en cours d'année n'est pas prise en considération par les enseignants. A cet égard, le cas de deux sujets (E53 et E54) est parlant: leurs progrès en cours d'année sont supérieurs à la tendance générale (respectivement +12,55 et +6,03). Leur faiblesse par rapport à leurs condisciples subsiste, toutefois; c'est ce dont témoignent notamment les notes réduites à l'évaluation de l'enseignant (-1,63 pour les deux) comme au test final (respectivement -1,01 et -1,33). Bref, bien que ces deux élèves se soient fortement améliorés en cours d'année, ils restent parmi les élèves les plus faibles de la classe 1MA.

⁹Le redoublement est décidé collégalement par l'ensemble des professeurs de 1^e année. Par conséquent, les élèves mis en échec en français ne devront pas nécessairement répéter leur année; un examen de repêchage peut leur être imposé à la fin du mois d'août.



Indéniablement, les deux enseignants considérés ici ne prennent pas en considération l'évolution des élèves en cours d'année pour décider de l'échec ou de la réussite. Il est aisé d'en faire la démonstration. Dans le tableau 11, nous avons regroupé huit élèves qui ont été promus, mais dont les progrès en cours d'année (progrès mesurés par les scores résiduels) sont largement inférieurs à ceux des huit élèves menacés de redoublement (tableau 11).

Tableau 11: Résultats de huit élèves promus au bilan des enseignants et à deux tests de français (début et fin d'année)

Elèves	Classes	Evaluations de l'enseignant	Test de début d'année	Test de fin d'année	Scores résiduels
E8	1LA	64% (-0,40)	81% (-0,44)	60% (-0,91)	-17,92
E12	1LA	65% (-0,33)	70% (-0,32)	57% (-1,16)	-13,63
E27	1LB	63% (-0,47)	66% (-0,65)	54% (-1,44)	-13,99
E32	1LB	66% (-0,26)	86% (-0,82)	63% (-0,62)	-19,36
E34	1LB	80% (0,75)	87% (0,93)	60% (-0,90)	-25,39
E43	1LB	74% (0,31)	70% (-0,32)	45% (-2,19)	-29,54
E51	1MA	60% (-0,69)	64% (-0,81)	53% (-1,46)	-12,20
E52	1MA	77% (0,53)	78% (0,25)	59% (-0,99)	-18,19

Si le jugement de réussite ou d'échec s'était appuyé sur le test de fin d'année, les élèves E12, E27, E51 et E52 auraient dû être traités de la même manière que les élèves E6, E33, E53, E54 et E60. En définitive, le cas des premiers est bien plus inquiétant que celui des seconds: en cours de première secondaire, ils ont nettement moins amélioré leurs compétences en français que leurs camarades, ce qui se traduit par des scores résiduels fortement négatifs. Manifestement, les pratiques de ces deux enseignants de français sont fort éloignées de l'esprit de l'évaluation formative.



Les enseignants croient-ils en leur pouvoir d'action?

Il serait présomptueux de tirer des conclusions générales de cette étude de cas menée dans quatre classes d'un établissement d'enseignement secondaire en Communauté française de Belgique. Les analyses présentées ici portent sur deux enseignants seulement et, de surcroît, certaines données sont manquantes. La prudence s'impose de façon évidente. Toutefois, les résultats obtenus offrent une belle cohérence interne tout en se révélant compatibles avec les grandes tendances observées dans la littérature de recherche. Il importe, par conséquent, de s'interroger sur la portée des observations rapportées ici et de lancer un appel à d'autres études qui pourraient venir confirmer, nuancer ou réfuter nos hypothèses. Car celles-ci débouchent sur un questionnement qui nous paraît essentiel: dans quelle mesure les enseignants croient-ils pouvoir influencer positivement les capacités cognitives des élèves et, partant, dans quelle mesure se sentent-ils responsables de leurs apprentissages?

Les deux enseignants concernés par la présente étude de cas semblent, pour leur part, fonctionner selon ce que nous avons appelé une logique de la prédestination. Considérant que les difficultés éprouvées au primaire obligent à formuler un pronostic réservé quant à la réussite de certaines élèves, ils ferment leurs regards aux progrès manifestes réalisés par certains élèves, mais aussi au *statu quo* d'autres. Ce faisant, ils ne remarquent pas que la dispersion interindividuelle des compétences s'est atténuée. En quelque sorte, ils sont aveugles face aux effets engendrés par leur action pédagogique.

Pour les élèves, tout se joue au cours des premières semaines. En fonction des compétences acquises ou non pendant l'enseignement primaire, ils susciteront une image positive ou réservée chez les enseignants. Dès la première évaluation périodique (fin octobre), le sort de la majorité est scellé. Les efforts consentis par certains pour s'améliorer risquent fort de rester vains. Eu égard à ce que les psychologues ont découvert dans le champ de la motivation scolaire (Viau, 1994), il faut craindre que ne se développe, chez ces élèves, un sentiment de résignation: à quoi sert-il d'essayer de s'amender si les enseignants restent insensibles à leurs progrès?

En définitive, au terme de cette modeste étude de cas, on est en droit de mettre en question l'idée selon laquelle les enseignants pratiquent le redoublement parce qu'ils croient dans la vertu remédiate d'une répétition des enseignements. Au secondaire notamment, des enseignants se donnent – semble-t-il – pour mission de *séparer le bon grain de l'ivraie* et, convaincus de ce que le potentiel des élèves est immuable, ils usent de l'évaluation comme moyen de sélection.



Références bibliographiques

- Bain, D. (1988). Le redoublement, une mesure pédagogique, pour qui? In D. Bain & al. *La recherche au service de l'enseignement?* (pp. 57-72). Genève: Centre de recherches psychopédagogiques.
- Bounafaa, N. (1992). *Pratiques d'évaluation et échec scolaire dans un établissement d'enseignement secondaire*. Mémoire de licence. Liège: Faculté de Psychologie et des Sciences de l'Education de l'Université.
- Carugati, F. & Mugny, G. (1985). *L'intelligence au pluriel. Les représentations sociales de l'intelligence et de son développement*. Cousset: Delval.
- Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire?* Bruxelles: De Boeck.
- Gosling, P. (1992). *Qui est responsable de l'échec scolaire?* Paris: Presses Universitaires de France.
- Grisay, A. (1984). Les mirages de l'évaluation scolaire (1). Rendement en français, notes et échecs à l'école primaire? *Revue de la Direction Générale de l'Organisation des Etudes*, XIX (5) 29-42.
- Holmes, C. T. (1990). Grade Level Retention Effects: A Meta-Analysis of Research Studies. In L. A. SHEPARD & M. L. SMITH (Ed.), *Flunking Grades. Research and Policies on Retention* (16-33). Bristol: Falmer Press.
- Hutmacher, W. (1993). *Quand la réalité résiste à la lutte contre l'échec scolaire. Analyse du redoublement dans l'enseignement primaire genevois*. Genève: Service de la recherche sociologique, Cahier 36.
- Neave, G. (1977). *Equality, Ideology and educational policy: an essay in the history of ideas*. Amsterdam: Institut d'éducation, Fondation Européenne de la Culture, Cahier 4.
- Perrenoud, Ph. (1984). *La fabrication de l'excellence scolaire*. Genève: Droz.
- Pini, G. (1991). Effets et méfaits du discours pédagogique. *Education et Recherche*, 13, 3, 255-272.
- Viau, R. (1994). *La motivation scolaire*. Bruxelles: De Boeck.





Réflexions sur l'effet-école (et sur ses manifestations dans un archipel du bout du monde)

Aletta Grisay

Il y a longtemps que les chercheurs en sciences de l'éducation s'intéressent aux multiples différences observables entre systèmes scolaires pour tenter d'identifier les facteurs susceptibles d'expliquer les meilleures ou les moins bonnes performances de ces systèmes en termes d'acquis des élèves. C'est là une des principales raisons d'être des vastes enquêtes de l'Association internationale pour l'évaluation du rendement scolaire (IEA) et, plus récemment, du programme PISA de l'OCDE (Programme international pour le suivi des acquis des élèves). En effet, un grand nombre de caractéristiques des systèmes éducatifs (par exemple, les dispositions relatives à l'obligation scolaire, le curriculum officiel, les horaires, les normes d'effectifs par classe, les qualifications requises pour l'accès au métier d'enseignant) ne varient généralement pas, ou de manière peu significative au sein d'un même pays, ce qui ne permet d'en apprécier empiriquement les effets qu'en effectuant des comparaisons entre pays différents.

Il est vrai que dans un même système scolaire, l'homogénéité des conditions est souvent plus structurelle que réelle: des différences de rendement parfois importantes s'observent d'une école à l'autre – même lorsque l'on tient sous contrôle statistique les caractéristiques d'«entrée» (origine socio-économique, compétences initiales) des populations d'élèves qui fréquentent ces établissements. Le courant de recherches sur l'«école efficace» (*School effectiveness studies*) a trouvé là son terrain de chasse privilégié. Moins amples que les enquêtes précédentes par le nombre d'établissements et d'élèves étudiés, mais prenant plus soigneusement en compte les processus mis en œuvre dans l'environnement scolaire, ces travaux ont donné la priorité à une comparaison systématique entre établissements performants et peu performants (c'est-à-dire des établissements où les résultats des élèves s'avèrent significativement supérieurs – ou inférieurs – à ceux que l'on observe, en moyenne, chez des élèves de même origine sociale et de mêmes aptitudes initiales inscrits dans d'autres écoles). Ils ont permis, depuis une vingtaine d'années, d'effectuer une avancée indiscutable dans l'identification des caractéristiques de l'environnement scolaire les plus favorables aux progrès de l'apprentissage.



Un profil cohérent de facteurs d'efficacité

Par-delà le foisonnement d'informations et de résultats apportés par ces paradigmes de recherches, on ne peut qu'être frappé par une double série de convergences:

1. En ce qui concerne les aspects quotidiens du travail en classe, les travaux sur l'école efficace tendent à confirmer, de plus en plus clairement, la robustesse du «vieux» modèle de l'apprentissage scolaire de Carroll (1963 et 1989). L'apprentissage que réalisera un élève particulier est fonction:

- a) de ses *aptitudes* (intelligence générale, capacité d'apprendre);
- b) de ses *prérequis* (dispose-t-il des bases nécessaires pour comprendre la nouvelle matière qu'on lui présente?);
- c) de sa *motivation* (a-t-il envie d'apprendre cette matière? Est-il prêt à s'engager dans l'effort que cela requiert, et d'y persévérer?);
- d) du fait que *l'occasion d'apprendre* cette matière lui soit effectivement présentée (et cela pendant un temps suffisant);
- e) de la *qualité de l'enseignement* dispensé.

A l'exception du premier de ces facteurs (aptitudes générales de l'élève), tous sont au moins partiellement gérables par l'école. Les établissements efficaces, nous dit en substance la littérature (Purkey & Smith, 1983; Good & Brophy, 1986; Scheerens & Bosker, 1997), sont ceux où les enseignants donnent la priorité aux apprentissages fondamentaux, où des objectifs clairs sont fixés aux élèves, où il leur est fait comprendre qu'on les considère comme capables d'atteindre ces objectifs, où des évaluations fréquentes permettent de réguler l'enseignement et de s'assurer que chacun progresse à un rythme satisfaisant, où l'on parvient à éviter les pertes de temps dues à une organisation défectueuse ou à des épisodes trop fréquents d'indiscipline, et où les pratiques d'enseignement sont de type plutôt structuré (les leçons sont planifiées avec soin, en courtes séquences alternant exposé de la matière, questions/réponses et travail individuel ou en groupe; l'enseignant fait souvent le lien entre ce qui a été vu et ce qui va suivre; il rappelle les points essentiels; il multiplie les exemples, insiste sur les transitions, résume ou fait résumer ce qui a été vu, etc.).

2. En ce qui concerne les caractéristiques plus globales de l'établissement, de son fonctionnement et de ses modalités de gestion, il est intéressant de constater que les principaux facteurs «facilitants» identifiés par la littérature mettent l'accent sur la cohérence. Le milieu scolaire est une bureaucratie qui se caractérise le plus souvent (en termes de théorie des organisations) par un mode de fonctionnement «faiblement couplé» (Weick, 1976), où les mécanismes de communication et de concertation sont peu développés et où chaque enseignant dispose d'une relativement large autonomie.



Les établissements performants, nous dit encore la littérature, sont ceux où une direction forte parvient à faire partager à tous une même vision des objectifs à poursuivre, où un *ethos* positif commun a pu se développer, où règne un climat de calme et de sécurité propice au travail, où le personnel se perçoit comme collectivement responsable des progrès des élèves – ce qui se traduit souvent par l'existence de dispositifs de régulation: les résultats des élèves font l'objet d'analyses au niveau de l'établissement et sont utilisés pour décider des améliorations à apporter à l'enseignement dispensé. Une politique concertée de formation continuée du personnel est menée au sein de l'école; les contacts avec les familles des élèves et leur implication dans le fonctionnement de l'école sont encouragés.

Tout cela fait sens, indiscutablement, et les praticiens reconnaissent sans peine, dans les listes de variables ainsi isolées par les travaux des chercheurs, des facteurs d'efficacité qui leur sont familiers. Si familiers qu'on tend parfois à oublier que ces résultats n'ont rien de trivial: bien d'autres facteurs sont souvent considérés comme indispensables au bon fonctionnement de l'enseignement, sans que leur pertinence ait pu être mise en évidence de manière empirique. C'est le cas, pour ne citer que quelques exemples, de l'utilisation de classes à effectifs réduits, de la plupart des pratiques d'individualisation des apprentissages, et, du moins dans les pays industrialisés, de la richesse de l'établissement en matériel pédagogique, ou du niveau de qualification du personnel enseignant.

Des points d'interrogation tenaces

L'avancée est réelle, mais d'innombrables problèmes subsistent, en particulier à propos de la stabilité de ces résultats et de leur interprétation exacte.

1. Les méta-analyses disponibles sur ces études (Witziers & Bosker, 1997) montrent que le profil des facteurs mis en évidence varie bien davantage d'une recherche à l'autre et d'un pays à l'autre qu'on ne l'avait cru tout d'abord. En particulier, l'importance du rôle joué par la direction de l'école ne paraît confirmée que dans un certain nombre d'études, pour la plupart anglo-saxonnes. Aux Pays-Bas (un pays où pourtant le courant de recherches sur l'école efficace est très actif, et où plusieurs enquêtes inspirées de ce paradigme ont été menées), aucun effet significatif n'a pu être trouvé pour cette variable. Il en va de même pour les quelques recherches menées en France (Bressoux, 1993; Grisay, 1997). Plusieurs autres des dimensions souvent citées dans la littérature présentent (quoique à un moindre degré) ce même comportement («à éclipses»): confirmées par certaines des études, elles sont loin de l'être par toutes.

2. Les travaux relatifs à la stabilité de l'effet-école dans le temps et pour les divers aspects du curriculum (un établissement «performant» le reste-t-il pour l'ensemble des matières enseignées et d'une année scolaire à l'autre?) font apparaître des résultats qui demeurent contradictoires. Une relative cohérence est observée dans certains



des échantillons d'écoles étudiés. Par exemple, Rutter & al. (1977) trouvent que les établissements scolaires londoniens qu'ils ont étudiés sont le plus souvent «performants» ou «peu performants» pour l'ensemble des critères cognitifs et socio-affectifs qu'ils ont examinés, et le restent généralement au fil du temps. Mais de nombreuses autres études (Rowan & Denk, 1982; Bosker & Scheerens, 1989; Luyten, 1994) mettent plutôt l'accent sur la faible stabilité des performances d'un même établissement d'une matière à l'autre et d'une année à l'autre. Slater & Teddlie (1992) attribuent ces fortes variations au fait que des modifications importantes peuvent se produire au cours du temps, tant en ce qui concerne la population qui fréquente l'établissement que sa direction ou la composition de son personnel enseignant.

Luyten (1994), dans une étude sur un large échantillon d'écoles secondaires néerlandaises, se demande pour sa part si l'unité pertinente n'est pas plutôt le «département» (ensemble des enseignants titulaires d'une même matière) que l'école proprement dite: il trouve que les performances des établissements dans une même *matière* sont un peu plus stables dans le temps que ce n'est le cas pour le rendement global de l'école (toutes matières confondues).

Il est légitime de s'interroger sur la nature exacte des systèmes de causalité que recouvrent plusieurs des dimensions mises en évidence par le courant de l'école efficace. La meilleure performance de certains établissements est associée, on l'a vu, à un curriculum plus fouillé et plus exigeant, à des attentes plus positives des enseignants vis-à-vis des élèves, à un climat plus calme et plus studieux. Mais comment interpréter ces associations? Le bon rendement des élèves est-il favorisé par les attentes positives des enseignants, le curriculum proposé, la discipline qui règne dans les classes? N'est-ce pas plutôt le fait que la population recrutée par ces établissements est «meilleure» qu'ailleurs, ce qui explique la bonne discipline, le curriculum exigeant et les attentes élevées des maîtres? Il n'est pas interdit de penser que les deux hypothèses sont simultanément correctes...

Un statut causal encore à préciser

Un effort intéressant pour mieux identifier les réseaux de causalité à travers lesquels agissent les divers facteurs de performance mis en évidence par le courant de l'école efficace a été tenté par Bosker & Scheerens (1994). Ces auteurs distinguent six mécanismes possibles:

1. Des effets de miroir. On peut penser que dans certains cas, le bon rendement collectif est tout simplement la résultante de conditions positives présentes au niveau individuel (une classe est «bonne» parce qu'elle est fréquentée par des élèves initialement bons, une école est bonne parce qu'elle compte une majorité de «bonnes» classes). Dans ce cas, certains des facteurs d'efficacité ne feraient que refléter la relative homogénéité des bonnes conditions initiales (élèves motivés et disciplinés, entraînant de fortes attentes de la part des enseignants).



2. *Des effets contextuels.* Il est cependant possible aussi que l'addition de caractéristiques individuelles positives ou négatives ait un effet allant au-delà d'un simple cumul. Des élèves faibles ou des enseignants médiocrement motivés seront «tirés vers le haut» par leur groupe de pairs dans une école dont l'*ethos* dominant est positif. L'inverse serait vrai dans des établissements fréquentés par une majorité d'élèves faibles, ou recrutant un personnel en majorité médiocre.

3. *Des effets directs.* La bonne performance des établissements efficaces serait due aux dispositifs qui y sont mis en place pour obtenir cette performance. C'est évidemment l'hypothèse la plus plausible pour des variables telles que le dynamisme de la direction, les mesures prises pour gérer efficacement le temps de travail (lutte contre l'absentéisme, remplacement des enseignants en congé de maladie, maintien de la discipline, formation du personnel à des pratiques qui améliorent la participation active et l'implication en classe des élèves), ou encore l'utilisation de systèmes bien conçus d'évaluation des progrès réalisés.

4. *Des effets liés au renforcement (incentives).* Il est rare que des dispositifs de gratification matérielle soient mis en œuvre dans les systèmes scolaires (salaires «au mérite», ou compléments de subventions pour les établissements atteignant certaines normes de performance). Mais divers types de renforcements symboliques jouent probablement un rôle dans le maintien ou la détérioration de la qualité de l'enseignement d'une école: la réputation de l'établissement, l'existence éventuelle d'établissements concurrents, la pression des collègues ou de la direction, celle exercée par des associations de parents plus ou moins sourcilleuses sur la discipline ou sur le niveau d'exigence du curriculum...

5. *Des effets facilitants liés aux conditions matérielles.* Les auteurs citent l'exemple des établissements organisant des relevés informatiques des résultats des élèves, ce qui peut faciliter ou encourager un pilotage systématique de leurs progrès. On peut sans doute ajouter à cela les quelques résultats indiquant que les bonnes écoles bénéficient souvent d'une administration efficiente, qui évite au personnel d'inutiles pertes de temps et lui permet de donner la priorité aux tâches d'enseignement.

6. *Des effets de tampon.* Il s'agirait ici des dispositifs qui permettent de protéger le bon fonctionnement de l'école contre les influences négatives pouvant le menacer de l'extérieur. Plusieurs travaux semblent indiquer qu'une certaine autonomie de l'établissement constitue un facteur positif autorisant une meilleure adaptation aux besoins spécifiques des élèves qui le fréquentent, et qu'un des rôles importants joué par les directeurs efficaces est celui de médiateur entre l'établissement et diverses instances extérieures (la hiérarchie du système éducatif, les autorités locales, l'association des parents, etc.). Par ailleurs, de multiples études de cas indiquent que, lorsqu'une école est située dans une aire géographique «sensible» (violence, drogue, etc.), les conditions minimales de travail ne peuvent y être maintenues que si elle arrive à élever une barrière de sécurité la protégeant, et protégeant ses élèves, contre les perturbations venant de son environnement immédiat.



Utilisant un sous-échantillon de données recueillies lors d'une évaluation nationale menée dans les écoles secondaires des Pays-Bas (3^e année de l'enseignement secondaire), les auteurs font une ingénieuse tentative pour vérifier, d'une part si certains de ces modèles explicatifs peuvent être techniquement spécifiés (par exemple, en mettant à l'épreuve des modèles de régression multi-niveaux où l'on suppose, alternativement, que certaines variables ou groupes de variables n'ont qu'un effet principal, ou qu'elles agissent à la fois par leur effet principal et par des effets d'interaction avec d'autres variables); d'autre part si, dans ce cas, certains modèles sont plus puissants que d'autres en termes de variance de rendement expliquée.

Ils obtiennent des résultats mitigés. Bien que certains modèles paraissent plus pertinents que d'autres, les différences sont peu accusées et ne permettent que rarement de trancher de manière nette entre l'une ou l'autre hypothèse (sans doute parce qu'une même corrélation entre variables scolaires et rendement résulte souvent de plusieurs déterminismes à la fois, les uns «directs», les autres «contextuels», etc.).

D'autre part, ils observent, une fois de plus, que les effets sur le rendement explicables par des caractéristiques de l'environnement scolaire et de l'enseignement dispensé ne constituent, en tout état de cause, qu'une proportion très modeste de la variance totale entre élèves, par comparaison avec les composantes traditionnelles majeures (les caractéristiques socio-économiques de la population, les filières fréquentées, les aptitudes des élèves mesurées à l'entrée à l'école).

La notion de valeur ajoutée serait-elle réductrice?

Ce dernier constat, qui se répète avec une désespérante monotonie depuis le début des travaux sur l'effet-école il y a plus de quarante ans, a conduit quelques chercheurs (dont je fais partie) à s'interroger sur la pertinence réelle qu'il y a, dans ce courant de recherches, à définir l'effet-école en termes de «valeur ajoutée» (c'est-à-dire la variance entre écoles qui peut être expliquée par les seules différences de qualité de l'environnement scolaire ou de l'enseignement dispensé, lorsque l'on tient statistiquement sous contrôle l'origine socio-économique et les aptitudes initiales de la population recrutée).

Dans un article récent (Grisay, 1999), j'ai tenté de montrer que la pratique consistant à mesurer l'effet des variables scolaires «toutes choses égales par ailleurs en ce qui concerne les caractéristiques d'*input* des élèves» conduit vraisemblablement à sous-estimer l'ampleur véritable de l'effet-école – en attribuant par exemple à l'origine socio-économique des élèves une partie de la variance entre établissements qui est probablement due, en réalité, à la meilleure qualité de l'enseignement dispensé dans les écoles fréquentées par une population d'origine favorisée.

Les disparités de rendement en mathématiques observées dans un échantillon de collèges français (Grisay, 1997) semblent ainsi pouvoir être attribuées pour une part



importante (25% de la variance totale, sur 40% de variance relevée entre classes et entre écoles) au recouvrement entre les caractéristiques initiales de la population recrutée et la qualité de l'enseignement reçu. La variance «jointe» ainsi mise en évidence est attribuable au fait que l'enseignement est meilleur, à différents égards, dans les écoles comptant une majorité d'élèves d'origine favorisée et/ou dont le niveau d'aptitudes à l'entrée au secondaire était élevé.

A noter que ce constat concerne un système éducatif et un niveau scolaire où les élèves se trouvent encore dans un tronc commun – où donc, en principe, des disparités d'une telle ampleur ne peuvent guère s'expliquer par l'existence de filières inégalement «fortes» en raison de la sélection qui s'y opère et du curriculum inégalement exigeant qu'elles proposent en mathématiques.

Pour la même raison, on s'étonne de constater que, dans le même pays, la variance entre classes au sein des établissements *s'accroît* de plus du tiers en quatre ans (entre le début et la fin du premier cycle du secondaire); ce qui traduit probablement l'existence de différenciations tacites dues aux pratiques d'affectation des élèves aux classes et au curriculum réel qui y est mis en œuvre.

En dépit des problèmes persistants d'analyse et d'interprétation que pose l'étude des différences de rendement entre établissements, de tels phénomènes semblent confirmer à quel point il serait essentiel, pour les autorités responsables des systèmes éducatifs, de pouvoir disposer à tout le moins *d'informations systématiques sur l'ampleur de ces disparités et sur leur évolution d'un cycle à l'autre du cursus scolaire des élèves.*

Il paraît important, par exemple, de tenter de comprendre pourquoi, dans le groupe de pays où le premier cycle du secondaire est organisé en tronc commun, les disparités de rendement entre établissements tendent à *diminuer* légèrement entre le niveau primaire et secondaire dans certains pays (pays scandinaves, Italie, Japon) alors qu'elles *s'accroissent* significativement dans d'autres (France, Belgique, Nouvelle-Zélande).

Dans les pays où des filières sont organisées dès la sortie du primaire (Pays-Bas, Allemagne, Autriche, Suisse), la variance entre établissements est, bien évidemment, beaucoup plus élevée au secondaire qu'au primaire, et elle s'accompagne souvent d'un net accroissement des effets d'agrégation liés à l'origine socio-économique des élèves (les élèves de milieu défavorisé tendant à s'orienter, plus souvent que les autres, vers les écoles proposant un enseignement technique ou professionnel). Des différences s'observent cependant d'un pays à l'autre en ce qui concerne l'ampleur de ce «regroupement social» au sortir de l'enseignement de base, et il serait intéressant d'en connaître les raisons.

Faute d'études de cas nationales permettant de mieux comprendre ces tendances, les relevés de variance entre écoles que l'OCDE a intégrés depuis quelques années dans le jeu d'indicateurs internationaux de l'éducation qu'elle publie régulièrement (OCDE/CERI: *Regards sur l'éducation*) demeurent, hélas, bien difficiles à interpréter.



Un cas d'école pour l'étude de l'effet-école

J'aimerais conclure ces quelques pages en proposant à la réflexion du lecteur une description succincte du «cas d'école» le plus extraordinaire qu'il m'ait été donné d'étudier au cours de mes travaux sur les disparités entre systèmes éducatifs et entre établissements d'un même système. Elle concerne le système scolaire d'un petit archipel du Pacifique, la République du Vanuatu (anciennement Nouvelles Hébrides).

Le pays est minuscule (environ 280 000 habitants, répartis sur plus de 80 îles situées à mi-chemin entre l'Australie et Hawaï). Il doit à sa géographie une forte fragmentation linguistique (une centaine de langues mélanésiennes, dont la majorité ne sont parlées, en moyenne, que par un millier d'individus). Son passé historique d'ancien protectorat franco-britannique a été marqué par les turbulences dues à une constante rivalité entre les deux puissances coloniales, et lui a légué deux réseaux scolaires concurrents: un tiers des écoles du pays sont francophones, les deux autres tiers anglophones. Par un étrange paradoxe, la langue de communication couramment utilisée dans l'ensemble du pays et dans les médias écrits et audiovisuels (le Bislama, un pidgin dérivé de l'anglais) est devenue la langue nationale officielle lors de l'accès du pays à l'indépendance en 1980, mais n'est enseignée dans aucun des deux réseaux scolaires.

Vingt ans d'indépendance, caractérisés par une alternance mouvementée de gouvernements à majorité tantôt anglophone, tantôt francophone, n'ont guère atténué la rivalité entre les deux systèmes scolaires; cependant, un effort important a été entrepris depuis quelques années pour unifier le curriculum et le contenu de l'examen de sortie de l'enseignement secondaire. Curriculum et examen de sortie restent en revanche distincts dans l'enseignement primaire.

La mise en place de l'examen national unifié à la fin de l'année 10 a mis en évidence d'importantes disparités de rendement entre les écoles du réseau francophone et celles du réseau anglophone – les résultats du groupe francophone s'avérant très inférieurs à ceux de l'autre groupe. C'est ce constat qui a été à l'origine de l'étude résumée ci-dessous: une assistance technique «neutre» a été sollicitée par les autorités du pays pour étudier l'origine de ces disparités (sous-rendement effectif des élèves francophones? Biais liés à l'inégale difficulté des versions anglaise et française des épreuves d'examen? Inégale sévérité des jurys de correcteurs anglophones et francophones?). L'étude, co-financée par l'Aide au développement britannique (DFID) et par la Coopération française, a été confiée à une équipe formée d'un expert anglophone et d'un expert francophone (Carr Hill et Grisay, 1999).

Le caractère particulier de la situation scolaire au Vanuatu vient du fait que les deux réseaux scolaires servent en réalité une même population. La plupart des villages ont à la fois une école primaire anglophone et une école primaire francophone – les familles inscrivent souvent certains de leurs enfants dans l'une et les autres dans l'autre, dans un souci pragmatique de ménager l'avenir en cas de retournement des



majorités politiques. Il en va de même au secondaire où, de plus, en raison des distances, la plupart des élèves des deux réseaux fréquentent l'internat de leur école, souvent dans une autre île que celle où se situe leur village. Par ailleurs, la quasi totalité des élèves sont confrontés à un même problème de trilinguisme: ils parlent leur langue maternelle avec leurs parents, la langue véhiculaire (Bislama) avec les camarades venant d'un autre village, et l'anglais ou le français en classe. Enfin, à l'exception d'une minorité de familles vivant en milieu urbain, tous ou quasi tous les élèves sont issus d'un environnement rural de type traditionnel, dont l'économie est fondée sur la pêche et/ou une agriculture familiale.

Compte tenu de ces conditions, si la présence de fortes disparités de rendement entre les deux groupes se confirme, il paraît improbable qu'elle puisse être due pour une large part à des facteurs externes à l'école: très vraisemblablement elle est plutôt attribuable à des différences de qualité de l'enseignement dispensé par les deux réseaux.

A ce titre, le cas du Vanuatu pourrait être considéré comme une illustration particulièrement nette d'une tendance déjà observée à plusieurs reprises dans les études sur les facteurs associés aux différences de rendement menées dans les pays en voie de développement (voir notamment Heyneman & Loxley, 1983). Les clivages dus à l'origine socio-économique des élèves sont souvent moins accusés dans ces pays, où une partie considérable de la population relève d'une économie de subsistance de type traditionnel et où la hiérarchisation sociale est moins développée que dans les sociétés industrialisées. Les facteurs proprement scolaires y jouent donc souvent un rôle beaucoup plus important que dans les pays développés, d'autant que les ressources éducatives, le niveau de qualification du personnel enseignant et la qualité du service scolaire tendent à y être plus inégalement répartis que dans les pays où la scolarisation de masse est beaucoup plus ancienne¹⁰.

L'étude, qui a porté sur l'ensemble des élèves du pays fréquentant la dernière année primaire (6000 sujets) et sur l'ensemble des élèves de l'année 10 (1200 sujets) a consisté à administrer dans les écoles anglophones et francophones des tests rigoureusement parallèles (lecture en 6^e année primaire; lecture, mathématiques et sciences en 10^e année). Les items sont extraits (avec de légères adaptations) des versions anglaise et française du matériel utilisé lors des enquêtes internationales IEA/RLS de 1991 et IEA/TIMSS de 1995. Les données ainsi recueillies ont été utilisées pour vérifier, par calibrage IRT, l'existence de biais éventuels liés à une inégale difficulté des examens nationaux passés par ces mêmes élèves au terme de l'année scolaire 1997.

¹⁰ A noter que ce constat recoupe indirectement une tendance relevée dans les pays industrialisés eux-mêmes: les variables décrivant la plus ou moins grande qualité de l'enseignement dispensé semblent constituer des déterminants du rendement dont le poids est sensiblement plus accusé en ce qui concerne la frange la plus défavorisée de la population scolaire qu'en ce qui concerne les élèves d'origine favorisée.



Des disparités considérables, qui s'aggravent du primaire au secondaire

L'analyse permet de confirmer que les différences relevées ne résultent pas d'un biais ou d'un artefact. Elles sont bien réelles, et d'une ampleur tout à fait considérable: à la fin du degré 10, les scores des élèves francophones sont, en moyenne, inférieurs à ceux des anglophones de plus d'un écart type en lecture et d'un écart type en sciences. L'écart est un peu moins accusé en mathématiques (près d'un demi écart type). Les résultats demeurent stables, que la différence soit mesurée par les tests IEA ou par les notes aux examens nationaux après calibrage IRT. Des disparités d'un tel ordre de grandeur sont, d'un point de vue pédagogique, très significatives: il y a tout lieu de penser qu'elles équivalent, du moins en lecture et en sciences, à plus d'une année scolaire de retard des élèves francophones sur leurs camarades anglophones.

Les disparités relevées en fin de primaire (où la lecture seule a été testée) sont significatives également, mais d'ampleur moindre (un demi écart type), indiquant que le handicap des francophones, déjà présent au primaire, s'élargit considérablement durant la scolarité secondaire.

Par ailleurs, la part de variance de rendement qui se situe entre écoles est énorme. Elle représente en 10^e année 58% de la variance totale des scores de lecture, 51% en sciences et 40% en mathématiques. Il est vrai que ce constat recoupe en partie le précédent: une part importante des différences entre écoles relève, en fait, des différences entre réseaux scolaires. Cependant les disparités sont très significatives même entre écoles d'un même réseau (de l'ordre de 20% de la variance totale en lecture et en sciences, et de plus de 30% en mathématiques).

A noter, enfin, que la variance entre écoles, déjà importante au primaire (32% de la variance totale des scores en lecture) s'accroît sensiblement au secondaire. Cette tendance semble liée à la nette aggravation du handicap des écoles francophones qui prend place au secondaire (et non à une augmentation des disparités entre écoles au sein de l'un et l'autre réseau).

Le brusque accroissement des écarts entre les deux groupes observé au secondaire pourrait cependant aussi avoir pour origine un artefact structurel. Pour des raisons historiques, le pays compte un nombre d'écoles secondaires francophones qui est proportionnellement supérieur au pourcentage d'élèves francophones dans la population. De ce fait, l'accès des francophones à l'enseignement secondaire fait l'objet d'une sélection moins sévère que ce n'est le cas pour les anglophones (leur seuil d'admission, en termes de scores à l'examen national de fin de primaire, est fixé plus bas). Les moins bonnes performances des élèves francophones à la fin de la 10^e année pourraient donc refléter aussi, pour une part, la moindre sévérité de la sélection opérée dans ce groupe.



Essai d'explication

Pour tenter de démêler l'origine de ces disparités, les renseignements recueillis par questionnaire sur les caractéristiques des élèves, de leur milieu familial et de leur environnement scolaire ont été introduits dans une série de modèles de régression multi-niveaux où les unités d'analyse sont d'une part les écoles fréquentées (niveau 1), d'autre part les élèves au sein des écoles (niveau 2). Ces modèles font intervenir, isolément ou en combinaison, les trois blocs de prédicteurs suivants:

- la langue d'enseignement (anglais *versus* français);
- l'ensemble des caractéristiques décrivant l'environnement éducatif des élèves (milieu familial et milieu scolaire);
- (pour les analyses portant sur l'année 10): les notes reçues à l'examen national de fin du primaire.

Les résultats des sept modèles mis en œuvre sont présentés dans un tableau à la fin de ce texte.

- Le *premier modèle* (modèle «vide» A) se borne à partitionner la variance totale de chacune des variables critère en deux composantes: variance entre écoles et variance entre élèves dans les écoles. C'est ce modèle qui met en évidence la très importante variance entre établissements évoquée ci-dessus.
- Le *second modèle* (modèle B) et le *troisième modèle* (modèle C) permettent d'estimer quelle est la part de la variance totale entre écoles qui peut être «expliquée», respectivement par la langue d'enseignement et par le score obtenu en 1993 à l'examen de 6^e année par les élèves fréquentant l'année 10 au moment de l'enquête.
- Le *quatrième modèle* (modèles B + C) permet d'étudier le recouvrement éventuel entre les deux facteurs précédents (et donc d'estimer dans quelle mesure les disparités entre écoles secondaires s'«expliquent» par l'inégale sévérité des examens de 6^e passés par les élèves anglophones et francophones). Il semble montrer que le facteur «sélectivité» ne fait qu'ajouter ses effets à ceux de la langue d'enseignement, sans vraiment s'y confondre; il ne peut en tout cas pas expliquer à lui seul l'aggravation du handicap des écoles francophones relevée au secondaire.
- Le *cinquième modèle* (modèle D) donne une idée de la variance totale des divers scores qui peut être expliquée lorsqu'on prend en considération les caractéristiques de l'environnement familial et scolaire des élèves. Certaines de ces variables ne peuvent expliquer que d'éventuelles différences entre écoles (p. ex. les variables décrivant les pratiques pédagogiques). D'autres (p. ex. les variables



décrivant le milieu socio-économique et culturel des élèves) sont susceptibles d'expliquer, pour partie, les différences observées entre élèves fréquentant une même école, et, pour partie, les différences entre écoles liées à un recrutement «moyen» plus ou moins favorisé.

Le détail de l'analyse (qu'il serait trop long de présenter ici), fait apparaître plusieurs phénomènes intéressants:

- Les différences liées à l'origine socio-économique des élèves, tout en étant (comme on pouvait s'y attendre) de faible ampleur, contribuent cependant à expliquer en partie la variance de rendement. Les scores sont plus élevés, dans les diverses matières et aux deux niveaux de scolarité, chez les élèves dont le père reçoit un salaire (par opposition à ceux dont le père est pêcheur ou agriculteur) et/ou chez ceux issus de familles qui disposent d'électricité, qui ont un raccordement à la distribution d'eau courante, qui se servent du gaz pour préparer les repas (par opposition aux familles qui cuisinent sur un feu de bois, puisent l'eau à la rivière, s'éclairent avec des lampes à pétrole). Les effets d'agrégation s'avèrent non négligeables (ces variables expliquent davantage de variance entre écoles qu'elles n'expliquent de variance interindividuelle!), indiquant que, même dans une société dont le niveau de vie est en apparence uniformément très modeste, des clivages entre écoles fréquentées par une population un peu moins ou un peu plus favorisée sont toujours présents.
- Les clivages d'ordre socio-culturel (le fait que les parents aient ou non fréquenté l'école primaire, l'existence de livres à la maison, l'usage plus ou moins fréquent de la langue d'enseignement en dehors de l'école) sont plus accusés que les clivages d'ordre socio-économique. Ils s'avèrent, comme toujours, significativement associés au rendement, particulièrement en lecture et en sciences.
- Parmi les variables décrivant l'environnement scolaire, les prédicteurs les mieux corrélés avec le rendement sont ceux relatifs au matériel pédagogique disponible dans la classe – un résultat fréquemment rencontré dans les enquêtes relatives à des pays en voie de développement (voir par exemple Heyneman, 1981; Levin & Lockheed, 1993). Les scores sont significativement inférieurs dans les écoles où seul le maître dispose d'un manuel (par opposition à celles – rares – où chaque élève en a un, ou celles où il y a au moins un manuel pour deux élèves).
- Le *sixième modèle* (modèles B + D) aide à comprendre dans quelle mesure ces variables environnementales permettent d'«expliquer» les disparités relevées entre langues d'enseignement. La comparaison entre les résultats des modèles B, D et B + D permet en effet de constater que la variance entre écoles expliquée par la langue d'enseignement est liée, pour une part significative, aux handicaps dont



souffre le groupe d'élèves francophones en termes d'origine socio-économique et culturelle et, surtout, en termes de qualité du service scolaire offert par les écoles que fréquentent ces élèves.

- Enfin, le *septième modèle* (modèles B + C + D) permet d'estimer la variance de rendement qui peut être expliquée par l'*ensemble* des prédicteurs évoqués ci-dessus et, par la même occasion, de voir que la variance dont l'ensemble des informations recueillies *n'a pas pu rendre compte*¹¹ est, pour l'essentiel, de la variance entre élèves dans les écoles. Les disparités entre écoles sont, elles, relativement bien expliquées.

Ce sont ces deux derniers résultats qui méritent l'attention dans le cadre du présent article. Le profil qui émerge des analyses statistiques (et que permettent de conforter, dans une large mesure, les nombreuses observations qualitatives, plus informelles, réalisées dans les classes) montre en effet que les facteurs responsables du grave handicap de la population francophone du pays appartiennent, en gros, à trois grandes catégories:

1) L'exposition à l'apprentissage est plus réduite dans les établissements francophones, particulièrement en lecture et en sciences. Les professeurs de sciences francophones, par exemple, déclarent avoir abordé en classe un nombre sensiblement moins élevé de points de matière prévus par le programme que ce n'est le cas pour leurs collègues anglophones. En lecture, les élèves sont moins souvent invités à utiliser la bibliothèque de l'école, lisent moins souvent en classe, ont moins de travaux à domicile impliquant des activités de lecture. Une enquête collatérale portant sur les compétences en lecture des futurs enseignants en cours de formation à l'école normale du Vanuatu montre par ailleurs que le niveau en lecture des futurs *instituteurs* anglophones est supérieur à celui de leurs collègues francophones (y compris ceux des sections préparant à l'enseignement *secondaire*). Les élèves-enseignants francophones fréquentent eux-mêmes plus rarement la bibliothèque de leur institut, ont moins de livres dans leur chambre d'internat, se voient moins souvent demander des travaux impliquant des activités de lecture. Au cours des visites de classes francophones, pratiquement aucune leçon de lecture n'a pu être observée – les cours de français auxquels les deux consultants ont assisté étaient presque tous des leçons de grammaire. Enfin, les familles francophones possèdent moins souvent des livres (alors que le taux de scolarisation des parents est très semblable à celui relevé dans les familles anglophones).

¹¹ Cette variance «non expliquée» vient de ce que (a) bien d'autres facteurs que ceux étudiés ici interviennent dans les résultats des élèves, (b) ces facteurs eux-mêmes ne sont sans doute pas mesurés avec précision par les variables recueillies, (c) les scores qui servent de critères sont, tout comme les prédicteurs, affectés d'une certaine part d'erreur de mesure.



2) La pénurie de manuels est nettement plus accusée, au secondaire, dans les classes francophones. Entre 10 et 15% des élèves anglophones signalent qu'il n'existe qu'un seul manuel de sciences, de mathématiques, etc. *pour toute la classe*. La proportion est de 30 à 50% (selon les matières) chez les élèves francophones. Cette disparité résulte en partie de la résistance des enseignants francophones du secondaire, qui semblent avoir massivement boudé les manuels développés dans le cadre de l'unification du curriculum (auxquels ils reprochent, non sans quelque raison, d'être de médiocres traductions d'un matériel originellement produit en anglais). Aucune des classes francophones visitées ne possédait, par exemple, la collection de supports écrits officiels développée pour l'enseignement des sciences.

3) Cela a, entre autres conséquences, un impact négatif sur les pratiques d'enseignement mises en œuvre dans les classes francophones. L'absence de supports écrits conduit la plupart des professeurs francophones à utiliser l'essentiel du temps de classe à écrire au tableau et à faire copier laborieusement par les élèves le contenu de la leçon du jour, dans d'épais classeurs à anneaux dont ils n'ont que rarement le temps de vérifier la gestion (la variable «utilise un classeur à anneaux» *versus* «utilise des cahiers» apparaît dans les analyses de régression avec un coefficient négatif extrêmement élevé). La comparaison avec les visites des classes effectuées dans le réseau anglophone confirme ce constat et met en outre en évidence un style d'enseignement plus interactif, plus diversifié, plus orienté vers l'acquisition de compétences fonctionnelles.

Les murs des classes anglophones sont, par exemple, couverts de documents écrits produits par les élèves, alors que les rares affiches présentes dans les classes francophones sont le plus souvent des rappels de règles d'orthographe ou de grammaire calligraphiés par l'enseignant, des cartes géographiques ou des posters distribués lors de campagnes d'hygiène. Les travaux de groupe, le dialogue entre enseignants et élèves, la correction collective de travaux sont fréquents dans les classes anglophones, alors que ce type d'activités n'est que rarement observé dans les classes francophones.

Dans le même ordre d'idées, la version anglaise de l'examen national de 10^e année donne la priorité à l'évaluation de compétences fonctionnelles de base (trois épreuves d'expression écrite, de nombreuses questions de compréhension de la lecture), tandis que la version française privilégie les questions de connaissance de type formel (vocabulaire et grammaire).



Conclusions

Sans doute, le double système scolaire du Vanuatu constitue-t-il un exemple extrême, où des circonstances peu communes rendent l'effet-école à la fois particulièrement visible et plus «explicable» qu'ailleurs. Ce cas offre néanmoins une illustration particulièrement nette de quelques-uns des aspects qui me paraissent importants dans la problématique de l'effet-école:

- Au Vanuatu comme ailleurs, le taux de variance entre établissements fait partie du groupe restreint d'indicateurs virtuellement les plus «riches de sens» sur le fonctionnement d'un système scolaire. Même si l'interprétation en est moins transparente dans nos pays, il y aurait lieu de poursuivre nos efforts pour apprendre à lire ce qu'un tel voyant peut nous signaler sur les disparités historiques, sociologiques, pédagogiques qui affectent nos systèmes scolaires.
- Ceci est particulièrement vrai aux périodes où des innovations à large échelle sont introduites dans les structures éducatives d'un pays. Les 50% de variance entre écoles du rendement en sciences au Vanuatu constituent un témoignage éloquent sur l'inégale implantation dans les établissements du pays du curriculum secondaire «unifié». Mais un taux très similaire de variance entre écoles avait été observé pour le rendement en mathématiques lors de la mise en œuvre du programme de mathématiques modernes dans l'enseignement primaire belge au cours des années 70...
- L'hypothèse selon laquelle un marché scolaire caractérisé par une forte concurrence entre établissements tendrait à favoriser une accentuation des disparités paraît se vérifier, non seulement dans les pays où de puissantes rivalités opposent établissements publics et établissements subventionnés conventionnels (Belgique, Pays-Bas, Irlande, etc.), mais aussi dans un pays comme le Vanuatu, où les réseaux scolaires concurrents relèvent tous deux de l'autorité publique.
- En ce qui concerne, enfin, les facteurs scolaires associés à ces disparités, on notera avec intérêt que, sous les traits spécifiques aux conditions d'enseignement propres à un pays en voie de développement, on reconnaît sans peine la déclinaison de quelques-unes des composantes les plus universellement «robustes» mises en évidence par la littérature sur les écoles efficaces: meilleure exposition à l'apprentissage, priorité donnée aux *skills* fondamentaux, utilisation plus efficiente du temps, pratiques d'enseignement plus structurées, plus interactives et plus diversifiées.



Répartition de la variance entre écoles et entre élèves dans les écoles au Vanuatu: divers modèles

Modèles	Année 6		Année 10	
	Lecture		Lecture	Mathématiques
(A) Modèle vide				
Niveau 1 (Ecole)	0.655 (31.9%)	5670 (57.9%)	4954 (39.1%)	3864 (50.6%)
Niveau 2 (Elève)	1.396 (68.1%)	4120 (42.1%)	7711 (60.9%)	3769 (49.4%)
(B) Langue d'enseignement				
Niveau 1 (Ecole)	0.603 (29.4%)	2234 (22.8%)	4094 (32.3%)	1433 (18.8%)
Niveau 2 (Elève)	1.396 (68.1%)	4122 (42.1%)	7715 (60.9%)	3782 (49.5%)
Effet sur la variance Ecole	-2.4%	-35.1%	-6.8%	-31.8%
(C) Score de 6^e				
Niveau 1 (Ecole)	Nd	4787 (48.9%)	3268 (25.8%)	3123 (40.9%)
Niveau 2 (Elève)		3856 (39.4%)	6458 (51.1%)	3593 (47.1%)
Effet sur la variance Ecole		-9.0%	-13.3%	-9.7
Effet sur la variance Elève		-2.7%	-9.8%	-2.3%
(B+C) = (B) Langue d'enseignement + (C) Score de 6^e				
Niveau 1 (Ecole)	Nd	1282 (13.9%)	2336 (18.5%)	833 (10.9%)
Niveau 2 (Elève)		3862 (39.4%)	6468 (51.1%)	3611 (47.3%)
Effet sur la variance Ecole		-44.0%	-20.6%	-39.7%
Effet sur la variance Elève		-2.7%	-9.8%	-2.1%
(D) Caractéristiques de l'élève et de son environnement culturel et scolaire				
Niveau 1 (Ecole)	0.464 (22.6%)	3279 (33.5%)	3654 (28.9%)	2384 (31.2%)
Niveau 2 (Elève)	1.203 (58.6%)	3737 (38.2%)	6976 (55.1%)	3397 (44.5%)
Effet sur la variance Ecole	-9.3%	-24.4%	-10.2%	-19.4%
Effet sur la variance Elève	-9.5%	-3.9%	-5.8%	-4.9%
(B+D) = (B) Langue d'enseignement + (D) Caractéristiques de l'élève et de son environnement culturel et scolaire				
Niveau 1 (Ecole)	0.443 (21.6%)	1424 (14.5%)	3539 (28.0%)	1166 (15.3%)
Niveau 2 (Elève)	1.203 (58.6%)	3739 (38.2%)	6974 (55.1%)	3409 (44.7%)
Effet sur la variance Ecole	-10.3%	-43.4%	-11.1%	-35.3%
Effet sur la variance Elève	-9.5%	-3.9%	-5.8%	-4.7%
(B+C+D) = (B) Langue d'enseignement + (C) Score de 6^e + (D) Caract. de l'élève et de son environn' cult. et scol.				
Niveau 1 (Ecole)	Nd	742 (7.6%)	2055 (16.2%)	714 (9.4%)
Niveau 2 (Elève)		3536 (36.1%)	5956 (47.0%)	3296 (43.2%)
Effet sur la variance Ecole		-50.3%	-22.9%	-41.2%
Effet sur la variance Elève		-6.0%	-13.9%	-6.2%

Tous les pourcentages sont calculés par référence à la variance totale du modèle vide (par exemple, dans la colonne *Sciences* du modèle C, 3726 représente 50% de 7453 (3684+3769) et non de 5215 (=1433+3726)).

L'«effet sur la variance école» et l'«effet sur la variance élève» correspondent, techniquement, à la réduction de la variance totale entre écoles ou entre élèves dans les écoles que l'on obtient lorsqu'on introduit dans l'équation un indicateur (p. ex. la langue d'enseignement) ou un groupe d'indicateurs (p. ex. les caractéristiques de l'environnement éducatif de l'élève).

Nd: Les modèles (C), (B+C) et (B+C+D) où entre en jeu la variable *Score obtenu à l'examen de 6^e* ne concernent, pour des raisons évidentes, que la 10^e année.



Références bibliographiques

- Bressoux, P. (1993). *Les performances des écoles et des classes: le cas des acquisitions en lecture*. Paris: Direction de l'évaluation et de la prospective, Dossier Education et Formation N° 30.
- Bosker, R. & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness studies. *International Journal of Educational Research*, 13, 741-751.
- Bosker, R. & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research*, 21, 159-180.
- Carr Hill, R. & Grisay, A. (1999). *Comparaison des performances des élèves fréquentant l'année 6 et l'année 10 dans les écoles anglophones et francophones du Vanuatu*. Londres: DFID, rapport de recherche non publié.
- Carroll, J. B. (1989). The Carroll model. A 25 year retrospective and prospective view. *Educational Researcher*, 18, 20-31.
- Good, T. L. & Brophy, J. E. (1986). School effects. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 570-601). New York: Macmillan.
- Grisay, A. (1997). *Evolution des acquis cognitifs et socio-affectifs des élèves au cours des années de collège*. Paris: Direction de l'évaluation et de la prospective, Dossier Education et Formation N° 88.
- Grisay, A. (1999). Comment mesurer l'effet des systèmes éducatifs sur les inégalités entre élèves? In D. Meuret (Ed.), *La justice du système éducatif*, (pp. 113-138). Bruxelles: De Boeck.
- Heyneman, S. P., Farrel, J. P. & Sepulveda-Stuardo, M. A. (1981). Textbooks and achievement in developing countries: What we know. *Journal of Curriculum studies*, 13 (3), 227-246.
- Heyneman, S. P. & Loxley, W. A. (1983). The distribution of primary school quality within high and low income countries. *Comparative Education Review*, 27(2), 108-118.
- Levin, H. M. & Lockheed, M. E. (1993). *Effective schools in developing countries*. London-Washington: Falmer Press.
- Luyten, H. (1994). Stability of school effects in Dutch secondary education: the impact of variance across subjects and years. *International Journal of Educational Research*, 21(2), 197-216.
- Purkey, S. C. & Smith, M. S. (1983). Effective schools: A review. *Elementary School Journal*, 83 (4), 427-452.
- Rowan, B. & Denk, C. E. (1982). *Modelling the academic performance of schools using longitudinal data: An analysis of school effectiveness measures and school and principal effects on school-level achievement*. San Francisco: Far West Laboratory.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J. & Smith, A. (1979). *Fifteen Thousand Hours: Secondary schools and their effects on children*. Cambridge, M. A.: Harvard University Press.
- Scheerens, J. & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Slater, R. O. & Teddlie, C. (1992). Towards a theory of school effectiveness and leadership. *School Effectiveness and School Improvement*, 3, 242-257.
- Witziers, B. & Bosker, R. (1997). *A meta-analysis on the effects of presumed school effectiveness enhancing factors*. Communication présentée au symposium ISCEI, Memphis, Etats-Unis.
- Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1-19.



Les systèmes éducatifs face aux inégalités et à l'échec scolaire: une impuissance teintée de lassitude

Philippe Perrenoud

Un enseignant de quarante ou cinquante ans pourrait croire, de bonne foi, que le système éducatif qui l'emploie a «tout fait» pour éliminer l'échec scolaire et les inégalités sociales devant l'éducation: réformes de structures, cycles d'orientation, développement de la scolarité préobligatoire, diversification des voies postobligatoires, modernisation des programmes, ouverture sur la vie, formation plus poussée des enseignants, audiovisuel et nouvelles technologies, dialogue plus ouvert avec les familles, respect des droits des élèves, sensibilité aux différences culturelles, soutien pédagogique, travail des enseignants en équipe, projets d'établissements, zones d'éducation prioritaire. A force d'en entendre parler, on pourrait même ajouter: évaluation formative, pédagogie différenciée, cycles d'apprentissage et approche par compétences, en s'imaginant que ces idées sont déjà mises en œuvre...

De ce foisonnement de mesures, il est tentant de conclure que, si les échecs scolaires, les abandons en cours d'études et les inégalités subsistent, si «*la réalité résiste*» (Hutmacher, 1993), c'est qu'il faut se rendre à l'évidence: on arrive au bout de nos moyens. Ce sentiment d'*impuissance* pourrait nous saisir d'autant plus facilement que les finances publiques sont dans les chiffres rouges et la croissance économique durablement ralentie, si bien que la fuite en avant vers de nouvelles réformes se heurte à des intérêts plus terre à terre: maintenir l'emploi et les conditions de travail, faire fonctionner l'école dans un monde en crise, où l'intention d'instruire n'est plus acquise. La violence et l'incivilité semblent, en cette fin du XX^e siècle, préoccuper les systèmes éducatifs davantage que l'échec scolaire. Comme s'il s'agissait d'un tout autre problème.

Alors que, dans les années 1950, on assistait à «l'émergence de l'échec scolaire comme problème social» (Isambert-Jamati, 1985), nous nous trouvons près d'un demi-siècle plus tard dans une tout autre configuration idéologique: à la fois moins de résistance au principe d'une éducation pour tous et plus de scepticisme sur les chances d'y arriver.



On peut comprendre ce sentiment d'impuissance teintée de lassitude. On ne compte plus les lois, les réformes, les programmes de formation, les projets qui se sont réclamés de la lutte contre l'échec scolaire. Les militants sont un peu comme ces pacifistes qui ont cru au «plus jamais ça» et qui voient avec désespoir éclater de nouvelles guerres.

Pourtant, à y regarder de près, il n'y a aucune raison de désespérer. Les tentatives de lutte contre les inégalités ne sont pas restées sans effet. Le niveau monte, même si cette démocratisation au sens large n'affaiblit pas les écarts entre les classes sociales: l'échelle a pris l'ascenseur, mais reste une échelle, selon la formule d'Hutmacher. Le destin scolaire des filles et des garçons est en voie de se rapprocher, voire la hiérarchie de s'inverser. Quant aux taux de scolarisation, ils indiquent que peu de jeunes sortent aujourd'hui du système éducatif en fin de scolarité obligatoire. Si tous n'ont pas un diplôme, presque tous ont eu l'occasion d'élargir leur formation de base. On peut donc être tenté d'opposer l'image d'un verre à moitié plein à celle d'un verre à moitié vide. Ne nous cachons pas cependant le fait que les compétences et les savoirs d'une partie des jeunes restent largement en deçà des objectifs visés par les systèmes éducatifs, alors même que ces objectifs sont censés correspondre à une dotation de base, à un «SMIG culturel» en deçà duquel on ne peut prétendre maîtriser son existence et participer en connaissance de cause à la vie de la cité.

Si nous restons «loin du compte», il ne faut pas s'en étonner: nos tentatives demeurent dérisoires en regard de la force des mécanismes à neutraliser, nos didactiques et nos politiques de démocratisation restent encore plus proches des gesticulations des médecins de Molière que d'une approche rigoureuse des problèmes.

Peut-être est-ce bien d'impuissance qu'il s'agit. Mais elle ne doit rien à la «nature des élèves». Elle tient à l'inconstance, à l'incohérence et à l'inefficacité programmée de nos politiques de l'éducation. Ce constat sévère ne vise pas en priorité les ministères. Leurs décisions ne sont que l'expression d'une société civile, d'un système politique, d'un appareil administratif et d'un corps enseignant qui ont l'art de vivre avec des problèmes en les déplaçant plutôt qu'en s'y attaquant.

Je tenterai d'analyser quelques-uns des mécanismes qui sous-tendent notre impuissance:

- Le refus de faire partie du problème
- La volonté de ne pas savoir
- Le poids des *lobbies*
- La recherche d'effets à court terme
- Les orientations de la formation des enseignants
- La conception du changement et des organisations
- Les failles du pilotage des réformes



Il est évidemment confortable de prendre une posture critique. C'est le rôle des chercheurs, mais il leur appartient aussi de contribuer, entre «engagement et distanciation» (Duru-Bellat, 1999) à poser les problèmes et à proposer des solutions. On ne saurait leur reprocher de n'avoir pas encore dégagé toutes les connaissances nécessaires: le volontarisme ne fait pas de miracles en ce domaine, comme on s'en aperçoit aussi en médecine ou en physique, en dépit de moyens sans commune mesure avec ceux qu'on accorde à la recherche en éducation. Pour le reste, les chercheurs font partie du système et n'ont pas davantage de vertu...

Le refus de faire partie du problème

L'échec des politiques, c'est, comme l'enfer, toujours les autres. Or, que voit-on?

- Les enseignants persistent à rejeter la responsabilité de l'échec sur les élèves et leurs familles. Comme si le désir de s'instruire était censé être déjà là, comme si le métier d'enseignant consistait encore et toujours à transmettre des savoirs à des élèves parfaitement disposés à les assimiler, si l'on excepte quelques moments de chahut rituel ou de fatigue. Dire «*Donnez-moi des élèves qui veulent apprendre, aiment apprendre, maîtrisent la langue scolaire et tous les codes, bénéficient de tous les appuis familiaux, et je me fais fort d'éliminer l'échec scolaire*», c'est dire en fait: «*Apportez-moi un problème résolu et je me ferai fort de le résoudre en un instant*». Les élèves «sont ce qu'ils sont», la seule compétence professionnelle qui vaille est de faire avec. On gaspille dans le système éducatif une énergie disproportionnée à chercher des *coupables*, alors qu'il n'y a que des *problèmes*, qu'on ne peut résoudre qu'en les considérant comme normaux et en les affrontant avec méthode. Nul ne saurait reprocher à un problème de ne pas être résolu! C'est celui qui s'y attaque qui fait la différence en lui trouvant ou non une solution.
- Les politiques ont pour logique (à quelques louables exceptions près) de rejeter tous les maux sur la détestable médiocrité de leurs prédécesseurs (fût-ce au sein de la même coalition) et d'annoncer qu'on va enfin s'attaquer aux vrais problèmes. J'y reviendrai à propos de la recherche d'effets à court terme.
- Les parents se plaignent de l'école, mais la plupart se comportent aujourd'hui comme des *consommateurs d'école* (Ballion, 1982) assez égocentriques, qui ne se soucient pas des effets pervers de leurs stratégies d'orientation. Au nom des intérêts de leur enfant – ce qu'on peut bien entendu comprendre – ils remanient le sens de la carte scolaire, des filières, des options, des diplômes, réintroduisant des hiérarchies cachées là où le système éducatif allait vers le collège unique et la coexistence de toutes les classes sociales. Les parents qui parlent le plus haut sont en général ceux dont les enfants reçoivent presque tout du système éducatif, alors que ceux dont les enfants sont, dès trois ou six ans, promis aux filières dévalorisées, ne trouvent pas d'aussi efficaces porte-parole.



- Les chefs d'établissements et les cadres scolaires sont plus soucieux de conserver ou d'améliorer leur position et d'administrer leur organisation que de transformer les pratiques pédagogiques. Lorsque le changement menace leur pouvoir gestionnaire ou simplement la tranquillité de l'organisation, ils choisissent la sécurité.
- Les formateurs, les didacticiens, les chercheurs en éducation, les inspecteurs, bref, les membres de la noosphère, sont souvent plus prompts à jeter la pierre aux praticiens qu'à s'impliquer dans les transformations sur le terrain. Cela s'explique sans doute par le fait que, dans le système éducatif, ceux qui pensent les pratiques sont en majorité des enseignants qui ont «réussi» à quitter la classe. Ils ne sont pas pressés de replonger dans les contradictions de l'action pédagogique.
- Les employeurs jugent très souvent l'école par le petit bout de la lorgnette et refusent d'admettre qu'ils ne peuvent exiger des salariés *à la fois* capables d'évoluer au gré des progrès technologiques et des restructurations de l'appareil de production et immédiatement adaptés à leur première tâche professionnelle.

La volonté de ne pas savoir

Elle se présente sous une double forme:

- il y a des choses que l'on ne sait pas, faute de s'être donné les moyens de les établir;
- il y a des choses que l'on sait, mais dont on ne fait rien, ce qui revient au même.

Notre ignorance effective porte sur divers aspects majeurs de l'échec:

- dans certains pays, les statistiques scolaires de base font défaut; on ne sait toujours pas, par exemple, dans quelle proportion les élèves redoublent, à quel âge, dans quelles disciplines cela se joue; on ne sait pas davantage l'importance des abandons (*drop out*, décrochage scolaire) après 15 ans;
- les indicateurs de compétences à l'issue de la scolarité de base et à l'âge adulte sont encore insuffisants; certes, les efforts américains (*National Assessment*), les grandes enquêtes de l'IEA, celle de TIMMS, puis les travaux de l'OCDE sur les indicateurs et les recherches dans divers pays (par exemple en France sous l'impulsion de Claude Thélot) permettent peu à peu de cerner les effets des systèmes éducatifs, mais nous sommes encore loin d'avoir les moyens d'une forte lucidité sur les *faits*;
- les inégalités devant l'école et entre écoles restent un tabou dans plusieurs systèmes éducatifs; on peut y affirmer sans risque d'être démenti par des données incontestables que tous les élèves ont les mêmes chances, que la classe sociale n'a pas d'influence sur la réussite scolaire ou que l'échec est surtout un problème de langue et d'immigration;



- les recherches expliquant les mécanismes de l'échec restent trop peu nombreuses, elles fonctionnent comme des «coups de sonde» dispersés et difficiles à comparer plutôt que comme des éléments de programmes cohérents; lorsqu'elles existent, elles éclairent davantage les caractéristiques des élèves que les fonctionnements de l'école; les travaux sur l'effet-maître et l'effet-établissement sont assez récents;
- les innovations et recherches-actions qui permettraient de vérifier et d'affiner les connaissances résultant des enquêtes sont trop rares.

Toutefois, nous en savons nettement plus qu'il y a trente ans, à la fois dans le registre descriptif et dans le registre de l'élucidation des causes et des mécanismes. Se sert-on de ce qu'on sait pour agir? Dans une mesure insuffisante:

- la culture statistique des acteurs du système éducatif reste limitée, un tableau à double entrée les effraie, une analyse de variance leur parle chinois; or, aucun des constats n'est simple, toutes les analyses sont multivariées, il faut contrôler nombre de variables pour isoler, par exemple, l'effet de la nationalité ou de la composition de la classe sur les résultats scolaires ou l'efficacité du redoublement;
- dans de nombreux systèmes, la formation des enseignants et des dirigeants les initie très marginalement, voire pas du tout à la recherche et aux sciences de l'éducation;
- les liens entre la recherche et les décideurs sont assez frustes, poussant une partie des chercheurs dans une position purement critique et inclinant une partie des décideurs à conclure qu'on ne peut rien attendre de ce monde de «coupeurs de cheveux en quatre», de «scientifiques» qui se prennent au sérieux, répondent aux questions par des questions, se manifestent lorsque les décisions sont prises ou préfèrent ne pas se compromettre dans des opérations politiques à hauts risques;
- quant à la connaissance de ce qu'on fait de la connaissance, en dépit d'intéressantes études sur la dissémination des savoirs et sur l'usage de la recherche par les enseignants ou les décideurs (Huberman et Gather Thurler, 1991), on ne peut pas dire qu'elle soit un moyen de régulation satisfaisant des modes actuels de transfert des connaissances et de gestion des innovations.

Les savoirs sont certes fragiles, controversés et changeants. Il y a cependant des secteurs où nous en savons assez pour agir. Nous savons par exemple que le redoublement est à la fois coûteux (pour les personnes comme pour les systèmes, financièrement et symboliquement) et qu'il n'est pas efficace (Crahay, 1996). Or, que font les acteurs de ce savoir? Ils lui préfèrent leur «conviction intime» que le redoublement, c'est «tout de même utile». Le «*Tenez, je connais un enfant qui, grâce au redoublement, a pu faire des études longues*» fait contrepois à toutes les statistiques. Second



exemple: les expériences de scolarité sans notes montrent que leur suppression n'a aucun effet négatif à condition de continuer à évaluer sérieusement les progressions et les acquis et à en informer les intéressés. Or, à cette évidence empirique, les acteurs, de l'enseignant au ministre, opposent leur expérience personnelle et leur attachement à la notation comme source de motivation. Troisième exemple: les classes censées être capables de suivre avec succès le même programme annuel sont de fait fort hétérogènes, si bien qu'on garantit, au gré d'un enseignement frontal, soit un nivellement par le bas, soit un échec d'une partie de la classe. Que fait-on de cette évidence? A l'heure où une pédagogie différenciée paraîtrait une règle de base dans tout système éducatif, certains en sont à envisager quelques heures de soutien, les plus avancés réservent la pédagogie différenciée aux zones d'éducation prioritaire et aux filières les plus dévalorisées, façon subtile de prouver qu'elle ne peut pas faire de miracles...

Ces constats n'expriment pas l'amertume de quelque inventeur déçu de voir le monde ignorer sa potion magique. Le problème de l'écart entre ce qu'on sait et ce qu'on en fait ne se réduit pas à un malentendu entre les producteurs et les utilisateurs des savoirs. Toutefois, ce malentendu existe et le monde de la recherche est loin d'être irréprochable. Il se montre très ambivalent à l'égard des orientations politiques fortes des programmes de recherche. Nombre de chercheurs – au nom bien entendu de l'indépendance de la science – préfèrent cultiver leur petit jardin que participer à une entreprise d'envergure, en aliénant une partie de leur liberté. Les programmes de recherche orientés vers l'action attirent sans doute des projets, mais n'est-ce pas tout simplement parce qu'ils promettent des crédits que la recherche fondamentale ne reçoit pas, du moins en sciences humaines?

Ici, comme sous d'autres angles, je ne cherche pas un bouc émissaire; ce sont des effets de système qui ne pourront être maîtrisés que par l'ensemble des acteurs, au prix d'une analyse de leurs modes de faire. Mais avant d'être une question de mémoire collective et de méthode, la volonté de savoir est une volonté *politique*. Nous sommes certainement sortis de l'obscurantisme et de la dénégation des faits qui ont accueilli, par exemple, le rapport Coleman (1966) ou les travaux de Bourdieu et Passeron (1964, 1970) sur les héritiers et la reproduction. De là à une mobilisation collective pour dresser «l'état des savoirs» et en tirer quelques conclusions majeures, il y a un pas que nulle société développée n'a véritablement franchi. On laisse le souci de définir de magnifiques politiques de l'éducation aux organisations internationales (OCDE, UNESCO, BIE, Communauté européenne, Conseil de l'Europe), qui se meuvent dans un *no man's land* qui permet d'accumuler des recommandations sans avoir les moyens de les mettre en œuvre... On peut en conclure que même si les stratégies de perpétuation des inégalités et de reproduction de l'ordre social ne dominent plus la scène politique (Berthelot, 1983), elles demeurent fort actives et conservent assez de force pour neutraliser les constats et surtout les conséquences politiques qu'on pourrait en tirer. La récente consultation sur les lycées, en France, suggère que l'état des rapports de force permet aujourd'hui d'analyser les fonctionnements et les effets du système éducatif, de proposer des réformes audacieuses,



mais pas de transformer l'essai. Les ministères ne comptent plus les livres blancs et autres rapports qu'ils ont commandés et qui sont restés sans suites. Ce qui nous amène au problème suivant.

Le poids des *lobbies*

Le système éducatif est un édifice plusieurs fois centenaire qui a mis en place des structures, des filières, des programmes, des examens. Derrière chaque élément de l'édifice, à chaque étage, il y a des emplois à préserver, des territoires à défendre, des pouvoirs, des privilèges et des publics à conserver. Ce n'est pas original, toutes les organisations souffrent des mêmes blocages. Ils sont, dans le système éducatif, aggravés du fait:

- de l'appartenance de la majorité des écoles au secteur public;
- de la dispersion des établissements et de leur relative autonomie de gestion;
- des alliances avec des autorités ou des forces associatives ou économiques locales;
- du poids des parents;
- de l'existence d'un corps professionnel (les enseignants) qui occupe l'immense majorité des emplois;
- de la nature d'un métier qui ne se prête ni à une rationalisation poussée, ni à un contrôle serré;
- des enjeux idéologiques qui font rapidement peser sur toute action un peu énergique le soupçon de s'en prendre au savoir, au maintien du fameux «niveau», à la citoyenneté, aux droits de l'homme, aux libertés, à l'égalité des chances, à l'équité devant la loi et l'évaluation, aux minorités, à la décentralisation, à la modernité, etc.;
- de l'existence de disciplines et de statuts bien séparés et contrastés, qui sont autant de champs clos qui développent leur logique propre et se considèrent volontiers comme des «systèmes dans le système».

Si bien qu'à de nombreux égards, le système éducatif s'apparente à la société bloquée dont parle Crozier (1971). Aucune force ne parvient à entraîner l'ensemble dans le changement, mais toutes ont assez de pouvoir pour neutraliser les initiatives des autres.

Les *lobbies* disciplinaires et statutaires sont particulièrement puissants dans tous les systèmes éducatifs. Il faut un combat de plusieurs années pour modifier la répartition des heures, introduire ou supprimer une discipline, rapprocher des statuts ou mettre fin à certains privilèges. C'est ainsi qu'en France la puissante société des agrégés défend les privilèges archaïques d'une caste, ce qui vaut à ses membres moins d'heures d'enseignement et les dispense, par exemple, d'avoir à présenter un mémoire dans leur formation en IUFM, la «leçon d'agrégation» donnée devant un jury tenant lieu de validation de la compétence pédagogique. Mais d'autres *lobbies*



ont un pouvoir impressionnant: les mathématiciens, depuis toujours, qui contrôlent les programmes scolaires à partir de l'Université, ou plus récemment les tenants des technologies nouvelles. Certains syndicats tout-puissants ou certaines associations de parents sont capables de mettre en crise une réforme. L'Eglise peut mobiliser l'épiscopat et les fidèles dès qu'un gouvernement s'attaque si peu que ce soit aux écoles confessionnelles.

Les ministres successifs, lorsqu'ils ne sont pas prisonniers de tel ou tel *lobby* en vertu de leur origine, ont fort à faire pour dessiner et surtout mettre en œuvre une politique cohérente. Sitôt connue, ses aspects les plus novateurs sont déchiquetés ou ridiculisés s'ils touchent à des intérêts particuliers. Il serait vain d'attendre du politique qu'il débloque un système bloqué, puisqu'il en fait largement partie. Il arrive qu'à la faveur d'une crise ou d'un rapport de force particulier, un magistrat fasse fortement progresser le système, mais de telles conditions sont loin d'être remplies constamment.

La recherche d'effets à court terme

Le système éducatif change lentement. Entre l'énoncé d'une idée, d'une méthode, d'un programme, d'une réforme et sa mise en œuvre à large échelle, il peut se passer plusieurs années, durant lesquelles le changement n'existe à l'état accompli que dans les textes et une partie des classes concernées.

A supposer que les structures, les fonctionnements et les pratiques aient vraiment changé à large échelle – ce qui est très rare – il reste à produire des effets sur les élèves. Or, il faut leur laisser le temps de traverser les nouveaux programmes, au moins un cycle d'étude, parfois l'ensemble du cursus. C'est donc souvent dix ans plus tard qu'on a des chances de voir des effets significatifs d'un changement, par exemple de l'introduction d'une approche communicative de la langue, des «mathématiques modernes», de cycles d'apprentissage ou d'une pédagogie différenciée.

Qui, dans notre société, peut et veut attendre dix ans pour évaluer les politiques publiques? Cette impatience ne caractérise pas l'école seulement, mais elle est dans ce domaine encore plus grave: on condamne des idées nouvelles – jugées décevantes! – sans avoir eu l'occasion d'observer leurs effets, ni même, assez souvent, leur réalisation dans le terrain. Les décideurs se comportent comme des médecins qui essaieraient toutes sortes de médicaments sans attendre l'effet des précédents.

La raison la plus évidente de cette «fuite en avant», c'est bien sûr un système politique qui fonctionne à coup de promesses électorales, d'effets d'annonce et de bilans spectaculaires. Comment attendre dix ans s'il faut faire la preuve tous les quatre ans qu'on a introduit des changements décisifs, «réformé le collège», «pris le virage des nouvelles technologies», «réconcilié l'école avec les langues», etc.?



On aurait tort, cependant, de mettre la précipitation sur le compte de quelques ministres désireux de réussir «leur réforme» avant de poursuivre leur carrière. L'ensemble du système cultive l'effervescence. En dehors des rituels articles et émissions sur «la rentrée des classes», les médias ne s'intéressent qu'aux polémiques sur l'école, qui se nouent à propos de crises et des réformes qui prétendent y apporter des réponses. Les parents n'ont pas dix ans devant eux, leurs enfants grandissent et si leurs associations travaillent aussi pour les générations suivantes, elles auraient du mal à ne pas exiger des changements rapides. Quant aux syndicats d'enseignants, moins pressés, ils ne s'engagent pas pour autant sur des politiques à long terme. Ils oscillent plutôt entre deux tactiques: exiger la satisfaction immédiate de certaines revendications relatives aux conditions de travail et temporiser pour tout ce qui touche aux changements de fond: programmes, évaluation, pratiques pédagogiques.

La «culture de l'évaluation» qui nous envahit ne contribue pas à calmer le jeu. Mandater des experts externes n'a de sens que sur une courte période: ils demandent que le système explicite ses objectifs, ils construisent des indicateurs de réussite, recueillent des données, rédigent un rapport et vont sous d'autres cieux répéter la même opération (Perrenoud, 1998). Quant aux experts et services d'évaluation internes, ils plaident souvent pour des calendriers plus longs, mais qui les entendent?

Le problème est en partie de l'ordre des représentations sociales: la plupart des acteurs ne conçoivent pas les médiations lentes et incertaines par lesquelles une décision est prise, comprise, mise en œuvre et porte des effets sur des générations d'élèves traversant le système. Il n'est pas rare, deux ans après une réforme de l'école *primaire*, d'entendre dire qu'elle «ne donne rien» puisque la réussite au baccalauréat n'a pas augmenté! La faible culture sociologique, démographique et statistique des acteurs ne contribue pas à les détacher d'une forme de pensée magique qui conduit à refuser de donner du temps au temps.

On peut comprendre ce mécanisme sans pouvoir le maîtriser. On se trouve dans une impasse institutionnelle: la faible durée de vie des gouvernements et souvent de la haute administration fait que, dix ans plus tard, aucune instance n'a la mémoire des intentions initiales et ne se sent véritablement comptable de leur réalisation. Qui, en 1999, se souvient des décisions prises en 1989 et tente de vérifier leur mise en œuvre? En France, par exemple, depuis 1989, plusieurs majorités de gauche et de droite se sont succédé, plusieurs ministres ont œuvré. Par une heureuse coïncidence, le ministre de l'éducation de 1989 est devenu, dix ans plus tard, premier ministre, alors que son conseiller d'alors est devenu ministre de l'éducation. Lionel Jospin et Claude Allègre ont donc tous les atouts pour assurer une certaine continuité. Même dans ce cas, ils sont emportés par le souci d'ouvrir de nouveaux chantiers – lycée, puis collège – et de suivre de nouvelles priorités, par exemple autour de la violence.

Il n'est évidemment pas question de rêver de la continuité que manifestent certaines sociétés totalitaires. Le problème est plutôt de savoir si l'on peut, dans les démocraties, mettre en place des instances moins obsédées par le court terme, des instances



indépendantes et légitimes qui, sans se substituer aux gouvernements et aux parlements, leur rappellent les engagements pris et les changements décidés il y a des années. Une instance comme le Conseil supérieur de l'Éducation paraît jouer ce rôle au Québec. Inutile de dire que si l'on donne à de telles instances les moyens de construire une mémoire et d'observer les changements, elles peuvent à certains moments contrecarrer des politiques ministérielles ou des compromis au sein d'une coalition.

Les orientations de la formation des enseignants

On n'en a pas fini, en éducation, avec les changements de structures et de programmes, mais on sait désormais qu'il ne faut pas en attendre de miracles dans la lutte contre l'échec scolaire. La lutte contre l'échec scolaire passe par des «changements du 3^e type» (Perrenoud, 1990), qui touchent aux représentations et aux pratiques des enseignants *au-delà* de ce qui peut être infléchi par les structures et les programmes, dans la sphère qui relève de l'autonomie professionnelle et ne peut faire l'objet de prescriptions. C'est largement le cas dans le domaine de la pédagogie différenciée, du rapport au savoir, de l'évaluation formative, des méthodes actives, de la relation pédagogique, des visions de l'enseignement et de l'apprentissage. Bref, presque tout ce qui pourrait jouer un rôle déterminant dans la lutte contre l'échec scolaire fait partie des changements de 3^e type.

Puisque ces changements de 3^e type ne se décrètent pas, mais ne se produisent qu'en vertu d'une libre conviction, la question stratégique est de savoir comment *convaincre* les enseignants de changer volontairement. La formation, continue et initiale, est à l'évidence l'un des leviers majeurs, même s'il en faut d'autres. La formation est un levier à un double titre:

- parce qu'elle agit sur les valeurs et les représentations, par exemple sur l'adhésion au principe d'éducabilité;
- parce qu'elle donne des savoirs et des compétences qui rendent les changements possibles.

Or, aujourd'hui, qu'observe-t-on? La plupart des réformes se heurtent à un déficit de formation, à la fois spécifique et général:

- le déficit est *spécifique* lorsque les thèmes de la réforme n'évoquent rien, au départ, dans l'esprit de la plupart des enseignants. C'est ainsi que la plupart des systèmes éducatifs sont en passe d'introduire des cycles d'apprentissage pluriannuels; mais, au moment où survient cette réforme, la représentation de ce qu'est un cycle, de sa définition, de ses raisons d'être, de son fonctionnement, ne fait pas partie de la culture professionnelle de base des enseignants; au moment où il faudrait pouvoir s'appuyer sur ces représentations pour engager le débat de fond puis construire des compétences opératoires, tout reste à faire;



- le déficit est *général* lorsque la réforme se heurte à un niveau de professionnalisation et de compétence peu compatible avec ce qui est proposé. Ainsi, introduire le projet d'établissement apparaît très aléatoire lorsque le corps enseignant n'a pas développé de fortes compétences de coopération et de gestion de projets; de même, introduire de nouvelles technologies d'enseignement ou d'évaluation est un peu vain lorsque l'alphabétisation informatique de base fait défaut.

J'ai soutenu ailleurs l'idée d'une «*introuvable synchronisation entre réformes scolaires et rennovations de la formation des enseignants*» (Perrenoud, 1999d). La temporalité des réformes scolaires ne paraît pas conciliable avec une forte anticipation de leur contenu. Lorsqu'une réforme surgit, on peut au mieux informer les enseignants en formation initiale. Si l'on adapte le curriculum, il est fort possible que cette réforme-là soit passée de mode lorsque les nouveaux enseignants arriveront au terme de leurs études...

Il serait en revanche possible:

- de concevoir les programmes de formation initiale en fonction des problèmes assez permanents auxquels les réformes successives tentent d'apporter des réponses. Il est sans doute difficile de former les enseignants spécifiquement à des formes d'aide méthodologique, à des parcours diversifiés ou à des grilles d'évaluation qui auront peut-être une durée de vie limitée, mais on peut parfaitement les former à la prise en compte des problèmes sous-jacents: hétérogénéité des niveaux, des rythmes, des intérêts, des manières d'apprendre, des rapports au savoir;
- de mettre largement la formation continue, plus mobile, au service des axes forts des politiques de l'éducation. Certes, dans une perspective de professionnalisation, il faut que la formation continue conserve une certaine indépendance et soit partiellement gérée par les enseignants eux-mêmes; mais entre une formation continue qui conduirait à réciter comme un catéchisme les slogans du ministre en fonction et une formation continue complètement déconnectée des politiques de l'éducation, il existe un moyen terme.

Dans ces deux domaines, on peut agir tout de suite, car il ne s'agit pas de mettre la formation à la traîne de la dernière réforme en date, mais d'anticiper. L'école est aux prises avec des contradictions et des problèmes récurrents, qui changent en partie de visage, mais présentent une forte permanence. Il «suffit» de les identifier pour en tenir compte dans les curricula de formation initiale et les actions de formation continue. Dans le domaine de l'échec scolaire, la liste des incontournables est assez facile à dresser, même si on peut en débattre à l'infini. Si la volonté politique existe, on peut incorporer aux référentiels de compétences tout ce qui permettra aux professeurs de participer, au niveau de l'établissement et de la salle de classe comme à celui du système, à la lutte contre l'échec scolaire (Perrenoud, 1999a).



La conception du changement et des organisations

Il se peut que les systèmes éducatifs vivent encore dans la nostalgie des changements du 1^{er} et du 2^e type. Une âme bien trempée de technocrate et de bureaucrate peut en effet regretter l'heureux temps où quelques textes bien conçus modifiaient les structures et les programmes sans trop se soucier de convaincre les acteurs. Au vrai, ces derniers ne demandent pas à être convaincus aussi longtemps qu'ils ont l'impression que les changements n'exigeront guère d'infléchissement de leur propre pratique, qu'ils pourront en quelque sorte «passer entre les gouttes» selon le principe: «*Le système change, mais moi, je ne change pas*». Les changements de structures sont à cet égard privilégiés: les enseignants intéressés investissent les nouvelles structures et les autres «se replient» sur les filières et structures existantes. La modernisation des programmes est un peu plus délicate, mais il est rare qu'elle exige davantage qu'une forme d'*aggiornamento* des savoirs à enseigner. Aussi longtemps qu'un nouveau programme incite à enseigner autrement, mais n'y contraint pas, le conflit reste modéré, car ceux qui n'y adhèrent pas adoptent les paroles sans la musique. Les rénovations de l'enseignement du français, par exemple, ont fait accepter de nouvelles conceptions de la grammaire, de la morphosyntaxe et du lexique relativement facilement, alors que le développement de compétences langagières à partir de situations de communication et d'écrits sociaux semble n'avoir véritablement convaincu que les enseignants déjà convertis aux pédagogies actives et aux démarches de projet.

Les autorités scolaires ont expérimenté les limites du modèle *top down*, mais le modèle *bottom up* n'est pas une alternative crédible, du moins pas pour faire évoluer à un certain rythme et en maintenant une certaine homogénéité l'ensemble du système éducatif. On parie aujourd'hui sur des modèles *middle out* qui tentent de concilier des impulsions venues d'en haut (un plan-cadre, avec des axes non négociables et l'exigence de rendre compte) et des initiatives locales. Dans ce contexte, l'autonomie des établissements prend tout son sens. Encore faut-il dépasser ses ambiguïtés (Perrenoud, 1999b) et penser l'établissement comme un lieu privilégié de rencontre entre les politiques de l'éducation et les pratiques des acteurs. Ce qui exige une vision neuve et audacieuse des établissements, de leur leadership, de leur fonctionnement (autorité négociée, démarches de projet, formation commune, partenariats locaux) et de leurs façons de rendre compte (Gather Thurler, 1998a et b; 1999b).

Aujourd'hui, dans les pays latins, les administrations scolaires vivent encore sur des conceptions bureaucratiques du changement, craignant constamment de perdre le contrôle des opérations et verrouillant en conséquence le système. A l'inverse, dans les cultures anglo-saxonnes, l'autonomie locale et le service à la communauté priment sur la mise en œuvre de politiques nationales. Bref, les systèmes éducatifs sont à la recherche d'une conception de l'organisation faisant la part de la différence, de la négociation, du conflit et du désordre (Alter, 1990) sans pour autant sombrer dans l'anarchie et les rationalités inégalitaires (Grisay, 1988).



Aussi longtemps que les visions de l'organisation scolaire oscilleront entre l'autoritarisme centralisateur des pouvoirs organisateurs et la volonté des enseignants de faire ce qu'ils veulent sans rendre de comptes, il est probable que toute réforme, aussi prometteuse soit-elle sur le fond, sera vouée à l'échec, parce que les jeux des acteurs la réduiront d'un côté à quelques prescriptions de plus, de l'autre à quelques ruses nouvelles pour échapper au changement.

Les failles du pilotage des réformes

L'un des aspects des stratégies de changement passe par une réponse claire aux questions suivantes: qui veut le changement, qui le pilote, qui l'évalue?

Dans une démocratie, la réponse «juridiquement correcte» à cette question est bien entendu: *les autorités légitimes*, qu'on entende par là des pouvoirs organisateurs locaux, comme en Belgique, des Eglises, des collectivités locales ou l'Etat régional ou national. Or, la légitimité des décideurs entre en conflit avec la nécessité d'associer à la politique d'innovation des acteurs qui ne sont pas formellement représentés dans les instances de décision, notamment les usagers (parents, grands élèves) et les enseignants, mais aussi les cadres scolaires et parfois les autorités locales et divers partenaires du système éducatif. Il importe donc de construire des systèmes complexes d'orientation du changement.

Dans une entreprise, ces questions devraient se poser encore moins que dans une collectivité publique: ne revient-il pas à la direction le rôle de vouloir, de piloter et d'évaluer le changement? Or, les entreprises innovantes, celles qui se veulent des «organisations apprenantes», mettent en place des structures de participation, des dispositifs de consultation, des «démarches qualité» ou des audits pour tenter de concilier la légitimité du pouvoir, l'expertise et la coopération des forces vives de l'organisation.

Les administrations scolaires sont en retard sur cette évolution et l'on peut regretter qu'elles retiennent du secteur privé les contrats de prestation, fer de lance du *New Public Management*, plutôt que les stratégies d'innovation participative. Alter (1990) montre notamment que dans les organisations innovantes, loin de vouloir tout contrôler, la direction adopte la posture d'un médiateur et d'un arbitre entre les fractions les plus conservatrices et les plus progressistes du système.

J'ai plaidé ailleurs (Perrenoud, 1999c; voir aussi Gather Thurler, 1999a) pour que le pilotage des réformes et plus globalement des politiques de l'éducation soit confié à des instances regroupant les acteurs sociaux concernés ou leurs représentants, en laissant aux directions administratives à la fois un pouvoir d'initiative et un droit de veto, mais sans leur donner le total leadership des opérations. Actuellement, la tendance est à négocier davantage que par le passé dans la phase de genèse des réformes, mais d'en confier la mise en œuvre aux administrations scolaires, en leur laissant le



soin de mobiliser des forces d'accompagnement et de formation. Or, la phase de conception n'est pas, dans une réforme, la plus problématique si elle survient dans une phase où le consensus s'établit au moins sur l'idée qu'un changement est indispensable. C'est le coût de ce changement qui divise les acteurs; c'est lorsque la réforme bute sur des obstacles et qu'il faut prendre des options difficiles que la négociation devient vitale. Certes, si la réforme a été conçue en vase clos, il serait vain d'attendre des acteurs sociaux qu'ils se solidarisent de décisions qu'on leur a imposées. Mais si on les dépasse de la concertation une fois les textes établis, en confiant «l'application des textes» à la seule administration, leur coopération fera défaut au moment de la mise en œuvre et videra la réforme d'une partie de son sens.

Les systèmes éducatifs n'ont pas encore trouvé des systèmes de pilotage négocié alliant cohérence et débat, légitimité formelle et assise sociologique.

Réformer la pensée?

«*Repenser la réforme, réformer la pensée*». C'est à cette vaste entreprise que nous convie Edgar Morin. La formule indique bien qu'il s'agit d'un cercle vicieux: notre éducation nous empêche de penser les problèmes de manière systémique et donc de transformer le système éducatif.

Hexel et Bain, dans un congrès intitulé «*L'institution scolaire est-elle capable d'apprendre?*» affirment: «*L'école n'apprend pas, elle sait*». C'est bien une partie du problème. L'école n'est pas une organisation apprenante, elle change sous l'empire de la contrainte et de la crise.

On peut proposer un changement de culture, par exemple à travers les six thèses suivantes défendues dans le même congrès (Gather Thurler et Perrenoud, 1991).



1. La valeur de la diversité: l'école apprend lorsqu'elle reconnaît que la force d'un système vivant procède de sa diversité plus que de son uniformité, lorsqu'elle permet et encourage la mise en commun et la valorisation des expériences locales.

2. Le droit à l'erreur: l'école apprend lorsqu'elle adopte des procédures de résolution de problèmes, qu'elle accepte le caractère provisoire et inachevé des programmes, des didactiques, des structures, qu'elle abandonne l'esprit de système et le mythe de la réforme définitive, qu'elle substitue le tâtonnement concerté aux directives et recettes venues d'en haut.

3. Une épistémologie réaliste et critique: l'école apprend lorsqu'elle accepte les limites de la connaissance de l'enfant et de l'apprentissage, reconnaît les impasses et les impuissances de toute action pédagogique, refuse la pensée magique, se dégage des mécanismes défensifs et des effets de façade.

4. Le souci de la méthode: l'école apprend lorsqu'elle s'en donne le droit et les moyens, lorsqu'elle s'organise pour formuler les problèmes, inventorier les hypothèses, ne pas tourner en rond, identifier les variables changeables.

5. Une certaine objectivation: l'école apprend lorsqu'elle accepte de se prendre et d'être prise comme un objet d'analyse et de théorisation, lorsque les structures et les pratiques, les représentations et les attitudes peuvent être décrites, expliquées plutôt que jugées.

6. Une ouverture vers l'extérieur: l'école apprend lorsqu'elle accepte de regarder au-delà de ses murs, de chercher des hypothèses, des paradigmes, des stratégies dans d'autres organisations et d'autres champs sociaux, de s'exposer telle qu'elle est au regard extérieur.

Il reste à savoir comment faire évoluer la culture de l'organisation et la culture politique dans ce sens...



Références bibliographiques

- Alter, N. (1990). *La gestion du désordre en entreprise*. Paris: L'Harmattan.
- Bain, D. (1991). *L'école est-elle capable d'apprendre?* In Société Suisse de Recherche en Education (SSRE), *L'institution scolaire est-elle capable d'apprendre?* (pp. 29-36). Lucerne: Zentralschweizerischer Beratungsdienst für Schulfragen.
- Ballion, R. (1982). *Les consommateurs d'école*. Paris: Stock.
- Berthelot, J.-M. (1983). *Le piège scolaire*. Paris: PUF.
- Bourdieu, P. (dir.) (1993). *La misère du monde*. Paris: Seuil.
- Bourdieu, P. & Passeron, J.-C. (1964). *Les héritiers. Les étudiants et la culture*. Paris: Ed. de Minuit.
- Bourdieu, P. & Passeron, J.-C. (1970). *La reproduction. Eléments pour une théorie du système d'enseignement*. Paris: Ed. de Minuit.
- Coleman, J. & al. (1966). *Equality of Educational Opportunity*. Washington: Office of Education, U.S. Department of Health, Education and Welfare.
- Collège de France (1985). *Propositions pour l'enseignement de l'avenir*. Paris: Collège de France.
- Crozier, M. (1971). *La société bloquée*. Paris: Seuil.
- Crozier, M. (1996). *La crise de l'intelligence*. Paris: InterEditions.
- Crozier, M. & Friedberg, E. (1977). *L'acteur et le système. Les contraintes de l'action collective*. Paris: Seuil.
- Duru-Bellat, M. (1999). La sociologie des inégalités sociales à l'école entre «engagement et distanciation». In D. Meuret (dir.), *La justice du système éducatif* (pp. 21-36). Bruxelles: De Boeck.
- Gather Thurler, M. (1998a). Savoirs d'action, savoirs d'innovation des chefs d'établissement. In G. Pelletier (dir.), *Former les dirigeants de l'éducation. L'apprentissage par l'action* (pp. 91-129). Bruxelles: De Boeck.
- Gather Thurler, M. (1998b). Rénovation de l'enseignement primaire à Genève: vers un autre modèle de changement. Premières expériences et perspectives. In F. Cros (dir.), *Dynamiques du changement en éducation et en formation. Considérations plurielles sur l'innovation* (pp. 229-257). Paris: INRP.
- Gather Thurler, M. (1999a). Au-delà de l'innovation et de l'évaluation: instaurer un processus de pilotage négocié. In L. Demailly (dir.), *Suffit-il d'évaluer?* Bruxelles: De Boeck (à paraître).
- Gather Thurler, M. (1999b). *L'établissement scolaire, un lieu où construire le sens du changement*. Paris: ESF (à paraître).
- Gather Thurler, M. & Perrenoud, Ph. (1991). L'école apprend si elle s'en donne le droit, s'en croit capable et s'organise dans ce sens! In Société Suisse de Recherche en Education (SSRE), *L'institution scolaire est-elle capable d'apprendre?* (pp. 75-92). Luzern: Zentralschweizerischer Beratungsdienst für Schulfragen.
- Grisay, A. (1988). La pédagogie de maîtrise face aux rationalités inégalitaires des systèmes d'enseignement. In M. Huberman (dir.), *Maîtriser les processus d'apprentissage. Les propositions de la pédagogie de maîtrise* (pp. 235-265). Paris: Delachaux et Niestlé.



Huberman, M. & Gather Thurler, M. (1991).
De la recherche à la pratique. Bern: Lang.

Isambert-Jamati, V. (1985). Quelques rappels de l'émergence de l'échec scolaire comme «problème social» dans les milieux pédagogiques français. In E. Plaisance (dir.), *«L'échec scolaire»: Nouveaux débats, nouvelles approches sociologiques* (pp. 155-163). Paris: Ed. du CNRS [repris In B. Pierrehumbert (dir.), (1992) *L'échec à l'école: échec de l'école*. Paris: Delachaux et Niestlé].

Morin, E. (1999). *La tête bien faite. Repenser la réforme, réformer la pensée*. Paris: Seuil.

Perrenoud, Ph. (1990). La formation équilibrée des élèves, chimère ou changement du troisième type? *C.O. Informations* (Genève), 8, 16-41.

Perrenoud, Ph. (1998). Evaluer les réformes scolaires, est-ce bien raisonnable? In G. Pelletier (dir.), *L'évaluation institutionnelle de l'éducation* (pp. 11-47). Montréal: Editions de l'AFIDES.

Perrenoud, Ph. (1999a). *Dix nouvelles compétences pour enseigner. Invitation au voyage*. Paris: ESF.

Perrenoud, Ph. (1999b). *L'établissement scolaire entre mandat et projet: vers une autonomie relative*. Université de Genève: Faculté de psychologie et des sciences de l'éducation.

Perrenoud, Ph. (1999c). Le pilotage négocié du changement dans les systèmes éducatifs. In J. Lurin & C. Nidegger (dir.), *Expertise et décisions dans les politiques de l'enseignement* (pp. 88-103). Genève: Service de la recherche en éducation, Cahier N° 3.

Perrenoud, Ph. (1999d). Réformes scolaires et rénovations de la formation des enseignants: une introuvable synchronisation. In M. Carbonneau & M. Tardif (dir.), *Les réformes en éducation, leur impact sur l'école et sur la formation des maîtres*. Bruxelles: De Boeck (à paraître).



De la didactique de la langue à la didactique de la littérature

Jean-Paul Bronckart

Le propos de cette contribution est de tenter de démontrer que le texte, comme objet d'enseignement, peut constituer le lieu d'une véritable **articulation** des démarches de didactique de la langue et de didactique de la littérature.

Une telle démonstration requiert deux conditions préalables.

La première est de montrer en quoi le texte constitue un objet pertinent pour une didactique de la langue. En soi, cette démonstration n'est plus à faire, tant sont nombreux les travaux théoriques et empiriques qui montrent qu'une didactique de la langue articulée aux finalités communicatives modernes, ou encore aux finalités pratiques de lecture-écriture, conduit nécessairement à une prise en compte de la dimension textuelle. Pour éviter la banalité et les redites, je proposerai dans la deuxième partie de cette contribution une conception des textes, de leur statut externe de production sociale d'une part, des différents niveaux de leur architecture interne d'autre part, susceptible de fonder et d'organiser la longue démarche de didactique de la langue qui se déploie dans le cursus scolaire. Longue démarche qui, conformément à la Tradition que je ne contesterai pas sur ce point, devrait servir d'instrument à (et se prolonger dans) l'enseignement de la littérature.

Mais il existe une seconde condition, comme chacun le sait beaucoup plus problématique, qui est d'établir en quoi consiste, en quoi pourrait consister ou encore en quoi devrait consister cet enseignement de la littérature. Plus spécifiquement en quoi cet enseignement relève-t-il d'une didactique, dans l'acception contemporaine de ce terme: *didactique des disciplines scolaires*. Cette acception implique d'une part que la littérature est une discipline, organisable en termes de savoirs, et notamment de savoirs scientifiques, et elle implique d'autre part une démarche structurée d'enseignement-apprentissage au terme de laquelle les élèves seraient susceptibles de s'appropriier ou de *re-construire* une partie au moins de ces savoirs. C'est, bien sûr et nécessairement, à l'examen de cette problématique que sera consacrée la première partie de mon exposé.



Problèmes et enjeux d'une didactique de la littérature

Puisque la question est difficile, commençons par deux constats d'évidence:

- il existe bien une littérature, au sens très provisoire de sous-ensembles de textes ou d'œuvres auxquels des jugements sociaux d'ordres divers attribuent un statut, positif, de littérarité;
- dans les pratiques éducatives de nos sociétés occidentales, cette littérature, en raison même de la valeur éminemment positive qui lui est attribuée, constitue un objet digne d'enseignement.

A l'énoncé de ces évidences, trois questions émergent immédiatement, que les didacticiens ne peuvent pas ne pas se poser.

- 1) Comment définir, de manière plus fine ou moins provisoire, notre objet: la littérature et les évaluations sociales qui la constituent?
- 2) Au-delà du constat de la valorisation sociale de cet objet, quels sont les objectifs précis qui justifient son enseignement, et qui devraient en conséquence l'organiser?
- 3) Quelle est enfin la démarche, ou la méthodologie (ou plutôt les démarches et méthodologies) qui est (sont) pertinente(s) et appropriée(s) pour atteindre ces objectifs?

Dans le cadre de cette première partie, je me centrerai sur les deux premières questions.

Qu'est-ce que la littérature?

Nous avons tous été formés dans le cadre d'une tradition dans laquelle cette question ne se posait pas. Tradition qui posait l'existence d'un corpus littéraire quasiment statique, et qui, au-delà de conditions et d'arguments proprement idéologiques qu'il n'est guère utile, je crois de ré-évoquer et de re-récuser, se soutenait du caractère *montrable* de ce corpus. Celui-ci était présent et présenté dans des manuels, recueils ou encyclopédies confectionnés par un sous-ensemble de ces successeurs modernes des *législateurs de la langue* qu'avait, il y a fort longtemps, mis en scène Platon dans son célèbre *Cratyle*. Par ces autorités, qui, à l'instar des grammairiens pour le système de la langue, et des puristes normatifs pour l'usage de cette même langue, savaient distinguer le bon du mauvais, les œuvres de qualité et les autres, plus concrètement les textes dignes d'entrer dans leurs recueils et ceux qui devraient éternellement demeurer aux portes de ce panthéon.



De cette évocation, une fois de plus banale, une indication forte émerge néanmoins: dans son fondement, dans son origine première et quelque extraordinaire qu'ait été par ailleurs sa destinée sociale, la notion même de littérature est liée à l'enseignement; elle en est un produit. Cette thèse, l'histoire de l'enseignement de la langue la confirme. A m'en tenir, et on voudra bien m'en excuser, à l'histoire de l'enseignement en France, on peut relever que si, du XVI^e au XVIII^e siècle, l'ex-langue vulgaire qu'est le français était progressivement devenue un objet d'enseignement et d'étude, ainsi que le medium d'interaction, dans l'enseignement élémentaire ou primaire d'un côté, dans l'enseignement universitaire d'un autre côté, sa pénétration dans l'enseignement secondaire est restée bien plus lente: au XVIII^e, s'il était officieusement ou *de facto* devenu l'instrument des échanges scolaires, l'objet d'étude à la fois central et terminal des collèges demeurait le latin, avec sa double valeur d'instrument cognitif et de facteur de distinction sociale. Ce n'est qu'à l'issue de la révolution bourgeoise que le tableau se modifiera notablement, et ce en un double mouvement. D'une part, la langue française est érigée en objet digne d'enseignement, à côté du latin, et d'autre part la référence aux «auteurs du temps» à laquelle s'étaient adossés les jugements normatifs depuis Vaugelas se transforme, pour les besoins de cet enseignement, en référence aux «bons auteurs». «Auteurs du temps», «bons auteurs», le changement de syntagme est significatif de ce que la référence ne se situera plus désormais chez les contemporains; au contraire, dans le cadre d'une idéologie explicite selon laquelle la langue française a atteint sa perfection (cf. Rivarol, 1784), la référence est nécessairement celle du passé, passé récent au début du XIX^e, puis passé de plus en plus lointain. Et comme chacun le sait, les conditions historiques de cette première délimitation du corpus de la littérature continuent de produire leurs effets jusqu'à ce jour, comme en témoignent les difficultés et ambiguïtés que suscitent les tentatives d'intégration à ce corpus des auteurs modernes et/ou contemporains.

Cela dit, la littérature était née, et son maintien allait exiger, en permanence, une légitimation du statut d'excellence des auteurs et des œuvres qui l'habitaient: qu'est-ce qui justifie qu'une œuvre soit déclarée littéraire?

La réponse qui a dominé au XIX^e siècle et au cours de la première moitié de ce siècle est bien connue. Une œuvre est littéraire si d'une part elle témoigne d'une mise en œuvre originale et habile des ressources de la langue, et si d'autre part elle véhicule et organise des thématiques et des valeurs susceptibles d'«enrichir» les lecteurs. Combinaison donc d'un critère esthétique et d'un critère moral qui, en soi, n'est guère contestable. Si ce n'est que cette légitimation a longtemps été associée à une conception **statique** de ces valeurs esthétiques et morales, sans prise en compte de leurs dimensions historiques et sociales: enrichissement moral pour qui, au bénéfice de quelle formation sociale? Beauté formelle aux yeux de qui, de quelle classe et/ou de quelle génération?



Dès lors que ces dimensions sont prises en compte, la justification du statut de littérarité se révèle bien plus problématique. En introduisant la notion de *champ littéraire*, la sociologie de Bourdieu (1977, 1980) a cependant fourni un cadre de réponse. Comme les autres champs culturels, le champ littéraire est un système de positions et d'agents (éditeurs, presse spécialisée, jurys de prix, auteurs, etc.), système organisé par des relations de solidarité et des rapports de force, et régi par des intérêts et des enjeux spécifiques. Ce champ organise la production de *biens littéraires* en *sphères de production* qui peuvent être très larges (œuvres à grand tirage pour grand public), restreintes (œuvres de conformité classique), voire confidentielles (poésie contemporaine tirée à quelques centaines d'exemplaires). Ces biens sont caractérisés par une valeur symbolique, qui est généralement d'autant plus forte que la sphère de production est limitée et qui, en tout état de cause, semble en contradiction avec les valeurs économiques dominantes de la société dans son ensemble. Cette valeur symbolique est désignée et construite par des discours attribuant aux œuvres (et à leurs auteurs) des propriétés absolues ou universelles, propriétés qui transcendent ou transgressent les contraintes et normes thématiques, techniques et idéologiques ayant cours chez les contemporains. Et l'imposition de ces valeurs désignées (ou l'adhésion à ces valeurs) contribue à un découpage de l'espace social, distribue des positions fondées sur les normes et les valeurs, positions qui sont elles-mêmes, en retour ou secondairement, garantes d'un réel pouvoir économique.

De cette approche fort éclairante (en dépit du caractère plus que sommaire du résumé que nous venons d'en proposer), je retiendrai un thème central: la littérarité est une construction historico-sociale, qui s'effectue dans les commentaires sur les œuvres, ou encore dans l'évaluation socio-discursive des œuvres. Et de là découle le caractère **relatif** de la littérarité que R. Balibar soulignait il y a plus de 20 ans:

«... les interprétations et les commentaires manifestent au grand jour l'effet esthétique (littéraire). Est littéraire le texte reconnu comme tel, et il est reconnu comme tel précisément dans le temps, et dans la mesure où il provoque pratiquement des interprétations, des critiques, des "lectures". Ce pourquoi un texte peut très bien *cesser* réellement d'être littéraire, ou le *devenir* dans des conditions qui d'abord n'existaient pas.» (1974, p. 44).

Ces éléments permettent d'établir un premier critère, *externe* et de ce fait encore très général, de délimitation de la littérarité. Critère qu'Y. Reuter a présenté de la manière suivante:

«Le bien littéraire (la littérarité) n'est ni le texte, ni le commentaire (ce qui ne ferait que déplacer le problème) mais le rapport texte-commentaire, lui-même intégré dans un champ spécifique.» (1990, pp. 9-10).

En d'autres termes, la littérarité est le produit d'interactions complexes entre le texte lui-même et les effets qu'il suscite chez diverses catégories de lecteurs (les agents du champ littéraire, l'auteur lui-même quand il décrit le projet sous-tendant son texte, et



l'ensemble des lecteurs «ordinaires»). En conséquence se trouve donc être littéraire, ou plutôt potentiellement littéraire, le texte qui suscite ce type de commentaires, qui est l'objet d'un débat légitime dans le champ.

Cela dit, les évaluations socio-discursives, si elles sont bien dans leur principe même génératrices de relativité, produisent néanmoins des classements qui comportent une part importante de stabilité. Et ceci semble tenir au fait qu'elles s'appuient sur d'autres critères, *internes* cette fois (au sens de «fondés sur les effets de lecture») et donc plus strictement définitoires, que je présenterai en m'inspirant largement des nombreux écrits d'Y. Reuter (cf. notamment, 1990, 1992, 1996).

- Critère de *porosité*

Ce terme désigne, selon son inventeur, «la capacité spécifique des biens littéraires d'intégrer une grande diversité de textes, de thèmes, de méthodologies, de théories, de notions, etc.», mais de les intégrer en partie seulement, ou encore en ne retenant que certaines de leurs propriétés. Il me semble que cette caractéristique est de même ordre que celle que Bakhtine (1979) avait décrite et théorisée, dans le cadre notamment de ses commentaires de l'œuvre de Rabelais, pour autant qu'on l'étende, comme Bakhtine l'a fait ensuite, à d'autres genres que le *roman*. Le texte littéraire constitue un *genre second*, en ce sens qu'il a la capacité d'absorber diverses formes de langage ordinaire, à l'œuvre dans un contexte culturel donné (conversations, langages spécialisés, comme ceux des médias, des professions ou des catégories d'âge), et de transformer partiellement ces formes usuelles, dans le mouvement même qui les insère dans des schémas plus généraux et plus conventionnels. Le texte littéraire, en déplaçant légèrement la formule de Bakhtine, absorbe des formes vivantes ou contextualisées pour les remettre en scène dans un autre espace d'interaction, qui est précisément celui du *champ littéraire* analysé par Bourdieu.

«Au cours du processus de leur formation, ces genres seconds absorbent et transmutent les genres premiers (...) de toutes sortes, qui se sont constitués dans les circonstances d'un échange verbal spontané. Les genres premiers, en devenant composantes des genres seconds, s'y transforment et se dotent d'une capacité particulière: ils perdent leur rapport immédiat au réel existant et au réel des énoncés d'autrui (... ils ne s'intègrent) au réel existant qu'à travers (le genre second) pris comme un tout, c'est-à-dire (...) conçu comme phénomène de la vie littéraire-artistique et non de la vie quotidienne.» (1979, p. 267).

- Critère de *désancrage*

Cette caractéristique, qui prolonge et complète celle de *porosité*, était déjà annoncée dans les commentaires qui précèdent: le texte littéraire introduit une coupure ou une distance par rapport aux contextes d'interaction du monde ordinaire (elle a été commentée par de multiples auteurs et conceptualisée notamment en termes de «relativisation de la fonction référentielle», par Marghescou [1974]). Alors que ces concep-



tualisations accentuent le caractère négatif, ou privatif de ce processus, nous en soulignerons pour notre part le versant plus positif: le désancrage (eu égard aux contextes d'interaction du monde ordinaire) n'est qu'une condition pour un *ré-ancrage*, pour une réinsertion des thèmes évoqués dans des mondes «autres», qui sont les mondes construits par l'activité discursive, ou encore les *mondes discursifs*. Et ces mondes discursifs *médiatisent* le rapport du lecteur aux contextes d'interaction du monde ordinaire; les formes représentatives-sémiotiques que ces mondes discursifs organisent imposent une orientation à la relecture des propriétés du monde ordinaire. Et cette réorientation discursive débouche alors sur le troisième critère.

- Critère de *finalisation humaine*

Appelée parfois «manifestation archétypale», cette caractéristique a été particulièrement mise en évidence dans les travaux de Ricœur (1983, 1984, 1985, 1986); mais, comme pour Bakhtine, nous la généraliserons à d'autres genres que les seuls genres narratifs analysés par cet auteur: le texte littéraire propose des *re-configurations* de l'activité humaine, destinées à fournir au lecteur ou à l'interprétant des instruments de *compréhension* de sa situation d'humain. Les savoirs accumulés (qu'ils soient d'ordre scientifique, social ou historique), et plus particulièrement les décisions d'action prises par les humains dans diverses circonstances de leur vie, sont des phénomènes difficiles à interpréter et à comprendre sans le secours d'une mise en forme verbale. Et la fonction du texte littéraire est de mettre en scène ces phénomènes d'une manière telle qu'ils puissent être évalués. Que cette mise en scène soit très directive ou au contraire très ouverte (ce facteur constituant un élément d'appréciation de l'évolution historique de la littérature), elle fournit un cadre à partir duquel se construisent, s'imposent, se rejettent, se négocient les valeurs des organisations sociales ou humaines. Et en outre, dans la mesure où, selon la formule de Hjelmslev (1968), le langage est l'*interprétant* incontournable des autres systèmes sémiotiques, le texte littéraire constitue aussi, comme le souligne Reuter (1990, p. 13), le «réfèrent du système esthétique»: il est le médium au travers duquel se réinterprètent les enjeux et les caractéristiques de la fonction de re-configuration à l'œuvre dans les œuvres architecturales, musicales, picturales, etc.

Une fois ces critères de définition posés, il me semble que la réponse à notre deuxième question devient beaucoup plus aisée: – pourquoi enseigner la littérature?

Les objectifs de l'enseignement de la littérature

Nous rejetterons d'abord, sans y consacrer trop de temps, les thèses romantiques, émanant du champ éducatif ou du champ littéraire, selon lesquelles la littérature ne s'enseigne pas, et *a fortiori*, qu'elle n'est pas didactisable. Thèse sous-tendant par exemple l'ouvrage de Pennac, *Comme un roman* (1992), qui met l'accent sur le caractère ludique, gratuit, du contact avec le littéraire. Sur un plan pédagogique, s'agissant du rôle que jouent les désirs, la motivation ou le plaisir dans les processus



d'apprentissage, ces attitudes me paraissent procéder d'une analyse inversant les causes et les effets. Comme les enseignants ne le savent que trop, la quête du plaisir est un mobile qui n'est que rarement disponible, au départ, chez les élèves; pour la plupart d'entre eux, il doit se construire: le plaisir est donc un effet de la connaissance, non une de ses conditions. Sur un plan plus général et plus profond, ces attitudes me paraissent participer, paradoxalement, d'une sous-estimation du rôle de la littérature, plus particulièrement de sa fonction générale d'instrument et de lieu d'évaluation de l'activité humaine que nous venons d'évoquer.

C'est cette fonction générale qui justifie que la littérature soit enseignée, et qui confère à cet enseignement cette validité éducative elle-même générale, qu'on lui a toujours attribuée. Mais cette validité générale n'est toutefois pas celle de la Tradition: imposer la lecture d'un corpus de textes restreint et imposer simultanément un type de commentaire de ces textes visant à la reproduction des valeurs des forces sociales dominantes, telles qu'elles sont organisées dans les interactions subtiles qui régissent les rapports entre le champ littéraire et les champs économique-politiques. A mes yeux, cette validité éducative générale tient à ce que la littérature est le lieu du *débat permanent* sur les valeurs de tous ordres qui orientent les activités humaines, et que les citoyens en devenir que constituent les élèves doivent être aptes à entrer dans ce débat, à participer à ce travail interprétatif permanent, qui est constitutif de la socialité et donc des personnes humaines autonomes.

Cet objectif éducatif général peut être décomposé en sous-rubriques, que je détaillerai en me centrant sur la problématique de l'interprétation des actions humaines, telle qu'elle est formulée par la sociologie compréhensive issue de Weber (1971) et de Habermas (1987). Ce courant de pensée pose d'abord une unité d'analyse d'ordre anthropologique et sociologique: l'**activité collective**, ou les activités collectives, comme formes d'organisation des conduites historiquement élaborées par les formations sociales. Activité dont l'organisation même, dans ses aspects les plus «pratiques», est indissolublement liée à la présence d'un mode d'interaction particulier, le **langage**, comme processus de négociation et d'entente sur ce que sont les contextes d'activité, et par là même, comme instrument d'évaluation de cette activité. Selon Habermas, cette évaluation langagière des activités collectives s'effectue sur la base de trois ordres de critères: 1. les critères de vérité et d'efficacité, à partir desquels se construisent des représentations s'organisant en un «monde objectif»; 2. les critères de conformité aux normes et aux valeurs, à partir desquels se construisent les représentations s'organisant en un «monde social»; 3. les critères de véracité ou d'authenticité, à partir desquels se construisent les représentations s'organisant en un «monde subjectif». Et ce même courant de pensée pose encore que les **actions humaines** sont, généalogiquement d'abord des produits de cette évaluation de l'activité collective; les actions sont des *portions de l'activité collective* qui ne sont attestables que parce qu'elles ont été découpées et attribuées à un agent singulier par le jeu des évaluations. Et elles ne deviennent ensuite seulement des unités psychologiques que dans la mesure où l'agent singulier *s'est approprié et a intériorisé ces évaluations*. L'agent actif devient alors une **personne**, disposant de repré-



sentations concernant les diverses facettes de sa propre responsabilité dans le déroulement de l'activité collective.

Sur fond de cette approche théorique, on peut réinvestir les thèses de Ricœur. On peut considérer d'abord que cette auto-attribution de responsabilité qui est constitutive de la personne est toujours problématique, qu'il est toujours difficile, dans l'interprétation des actions, celles des autres ou les siennes, de délimiter le poids respectif des représentations propres à l'agent d'un côté (auto-attribution de capacités, d'intentions, de motifs), et des déterminismes ou contraintes issues des mondes externes d'un autre côté. Qu'il est toujours difficile de savoir en outre si les déterminismes auxquels nous pensons être soumis relèvent plutôt du monde objectif, du monde social ou encore du monde subjectif. Pensons à nos interventions éducatives; nos décisions d'actions sont-elles régies par le souci d'efficacité, par le souci de conformité aux valeurs en usage ou par le souci de préserver notre image de bon pédagogue? On peut considérer ensuite que les textes littéraires constituent un instrument visant à résoudre ce genre de problème; que la littérature propose des situations dans lesquelles les déterminismes internes ou externes des actions sont clarifiés ou, à tous le moins, discutés, problématisés. Et que c'est en entrant dans cette production littéraire, qui est bien, comme le soulignait Reuter, fondamentalement rapport entre le texte et son commentaire, que l'humain développe, solidairement, sa socialité et sa personnalité.

Les objectifs premiers de l'enseignement de la littérature sont donc la contribution à la **formation de la personne**, indissolublement liée à la **construction de la socialité**, et ce, par confrontation aux textes explicitant la manière dont les générations antérieures et les générations contemporaines ont abordé ou abordent ce type de question.

Mais à ces deux objectifs premiers s'ajoutent quasi nécessairement deux objectifs supplémentaires. Les questions relatives aux enjeux des décisions d'action et à la responsabilité des agents ont fait l'objet de traitements évidemment différents selon les positions de classe, les appartenances culturelles ou linguistiques, etc., des binômes auteur-lecteurs. Et l'enseignement de la littérature est une occasion d'entrer dans le **débat sur la diversité sociale et culturelle**, pour autant que l'on accepte de proposer et de confronter des corpus de textes constitués de littérature universelle, de littérature nationale et de littérature régionale. Par ailleurs et indépendamment des différenciations socio-culturelles, les textes littéraires témoignent, dans leur structure énonciative même, de l'évolution des positionnements philosophiques à l'égard de la question du sujet: du point de vue d'un énonciateur (ou narrateur) monolithique, reflet du sujet idéal cartésien, on est passé, comme l'a démontré Bakhtine, à un énonciateur démultiplié, témoin de la prise en compte, par la pensée du XIX^e siècle, du statut social et donc différentiel de ce même sujet. Et une littérature actuelle, inspirée du travail philosophique de déconstruction plus radicale encore du sujet cartésien, élabore des textes qui semblent s'interdire toute prise de position énonciative. L'enseignement de la littérature est donc aussi une occasion d'entrer dans le **débat philosophique**, pour autant que l'on accepte de proposer et de confronter des corpus de textes qui attestent de l'évolution qui vient d'être évoquée.



A ces quatre objectifs, j'ajouterai encore un dernier, qui n'est pas le moindre. Tout comme, selon la formule célèbre de Saussure, la langue est comme un manteau rapiécé, construit et reconstruit à partir des mêmes bouts d'étoffe, les textes sont des formes toujours nouvelles, construites et reconstruites à partir des mêmes moyens linguistiques: les unités et structures, par essence finies ou limitées, qui définissent le système de la langue. Et l'enseignement de la littérature est dès lors aussi l'occasion d'une **formation linguistique**, d'une mise en évidence des possibilités multiples, voire infinies, de structuration et de restructuration des ressources de la langue, au service de l'activité communicative ou discursive.

Si la littérature est bien un patrimoine, ce patrimoine est avant tout un patrimoine de débats, de travail interprétatif à propos de la personne humaine, de sa socialité, de la diversité socio-culturelle et des possibilités d'usage de la langue. Et la littérature est aussi, en cela même, le lieu de la contradiction permanente:

- contradiction entre, d'un côté, l'ancrage social et différentiel du champ littéraire et, d'un autre côté, la finalisation humaine tendanciellement universelle des textes produits dans ce champ;
- contradiction entre, d'un côté, la visée de construction de l'identité de la personne et, d'un autre côté la vision de l'insertion dans la socialité;
- contradiction encore, entre, d'un côté, l'unicité du système de la langue, et d'un autre côté, l'infinie diversité de ses réalisations textuelles.

Et c'est à ce titre qu'elle me paraît devoir être enseignée, comme témoin des **mouvements dialectiques** de la pensée humaine, et comme initiation à l'insertion dans ce processus permanent.

Si le pourquoi de l'enseignement de la littérature paraît ainsi assez clair, le «comment l'enseigner» reste, comme chacun le sait, beaucoup plus problématique, et les ébauches de réponse que je tenterai de formuler exigent un détour par la problématique de la didactique de la langue.

La didactique du texte comme aboutissement de la didactique de la langue

La didactique de la langue est, à l'évidence, tributaire des conceptions générales du statut même de la langue, dans ses rapports à la représentation du monde d'un côté, dans ses rapports aux pratiques verbales ou communicatives d'un autre côté.

La conception qui reste aujourd'hui dominante est celle du *représentationalisme* issu de la tradition aristotélicienne, c'est-à-dire la conception selon laquelle la fonction de représentation de la langue prime sur (et précède) sa fonction communicative. Dans cette optique, le système de la langue serait premier, et les textes ne constitueraient qu'une mise en fonctionnement «seconde» de ce même système. Le langage servirait



donc d'abord à traduire la logique du monde et/ou celle des opérations de pensée, et il ne servirait qu'ensuite seulement à transmettre ces représentations verbalisées aux autres. C'est cette conception qui continue de sous-tendre explicitement les travaux de Chomsky (cf. *Le langage et la pensée*, 1970) et des cognitivistes contemporains; elle implique la préexistence d'un substrat langagier, universel et inné, sur la base duquel se seraient développés ensuite, d'une part les différentes langues naturelles, et d'autre part les multiples genres de textes qui y sont attestables. Sans toutefois que soient conceptualisées les conditions et les processus ayant entraîné ces différenciations successives; conditions et processus qui restent dès lors, malgré la théorie récente des «principes et paramètres» (cf. Chomsky, 1985), largement mystérieux.

C'est manifestement en articulation à ce positionnement épistémologique dominant que se sont élaborées les démarches et méthodes **traditionnelles** d'enseignement des langues. Celles-ci préconisent en effet d'abord une approche «grammaticale» (au sens de «grammaire de phrases»), destinée à doter les élèves d'une connaissance explicite des principales catégories et structures du système de la langue. Et c'est sur cette base que les élèves sont censés développer ensuite une maîtrise textuelle, dans ses aspects de production aussi bien que de compréhension-interprétation.

Ce schéma postule donc que les connaissances relatives au système, telles qu'elles sont construites dans le cadre de l'enseignement grammatical, peuvent être efficacement exploitées dans le cadre des activités de production et d'interprétation des textes. Or l'observation des pratiques traditionnelles d'enseignement des textes et de la littérature montre qu'en réalité ce sont d'autres concepts, issus notamment de la tradition rhétorique, qui sont exploités à cet effet. Et de manière plus générale, l'utilité des connaissances grammaticales traditionnelles pour le développement de la maîtrise textuelle n'a, à ce jour, jamais pu être démontrée.

Cependant, comme l'a montré de Mauro (1969), depuis la Renaissance, la conception traditionnelle de la langue a fait l'objet de multiples contestations, s'alimentant à des arguments philosophiques, linguistiques et psychologiques décisifs. Dans la perspective tracée par Spinoza, Hegel (1807), Marx et Engels (1846), en prenant appui sur les propositions de Bakhtine, Habermas et Ricœur notamment, et en tenant compte aussi des dimensions **socio-langagières** du développement humain mises en évidence notamment par la psychologie vygotskyenne, il est possible de proposer un **renversement complet** du schéma classique des rapports entre connaissance et langage.

Cette autre conception pose que, dans l'espèce humaine, les **activités collectives médiatisées par les pratiques langagières sont**, phylogénétiquement et ontogénétiquement, **premières**. C'est dans le cadre des évaluations socio-langagières de l'activité que sont délimitées des actions imputables à des humains singuliers. C'est ensuite par l'appropriation et l'intériorisation des propriétés socio-sémiotiques de ces évaluations que se construisent des agents ou **personnes** se sachant susceptibles d'agir et de se représenter les contextes d'action et de se représenter le monde. Dans ce cadre, d'une part, la pensée consciente émerge comme un produit de l'action et



des discours; et d'autre part, ce qui nous intéressera plus ici, le «système d'une langue» ne peut plus être considéré comme un donné premier, précédant les pratiques langagières et les conditionnant, mais il doit être considéré comme un construit cognitif, élaboré secondairement par abstraction-généralisation à partir des régularités linguistiques observables dans les différentes pratiques langagières. Les textes sont donc premiers et le système de langue n'est qu'un *construct* second, à l'élaboration duquel se sont livrés (avec plus ou moins de bonheur) des générations de grammairiens.

Sur fond de ce renversement radical de perspective théorique, que nous avons discuté en détail ailleurs (Bronckart, 1985, 1992, 1995a, 1995b, 1996; Bronckart & al., 1996), pourrait être envisagée une démarche didactique **idéale**, qui consisterait à faire débiter l'enseignement de la langue par des activités de lecture et de production de textes, et à articuler ensuite à cette démarche initiale des activités d'inférence puis de codification des régularités observables dans le corpus de textes ainsi mobilisé.

L'inférence et la codification porteraient d'abord sur les régularités d'organisation et de fonctionnement des genres de texte et des types de discours, dans leurs relations aux contextes de production. On observerait dans ce cadre que si certaines des règles d'organisation des textes ont un aspect relativement contraignant, la plupart d'entre elles proposent plusieurs solutions acceptables (cf., par exemple, les possibilités d'usage du *passé simple*, du *passé composé* ou du *présent* comme temps de base d'un récit), et que leur maîtrise présente dès lors toujours un caractère fonctionnel ou contextualisé.

L'activité d'inférence et de codification ne porterait qu'ensuite, au titre de généralisation, sur les régularités majeures d'organisation des phrases et sur les catégories d'unités disponibles dans la langue, en l'occurrence sur les quelques règles contraignantes de la morphosyntaxe, dont la connaissance est notamment nécessaire pour la maîtrise de l'orthographe grammaticale.

Ce schéma «idéal» d'enseignement de la langue n'est cependant **pas applicable** comme tel.

Tout d'abord parce qu'une réforme didactique ne peut consister en une «application», ou en une transposition directe, de conceptions théoriques (quelle que soit leur pertinence intrinsèque) au champ pratique. Comme nous l'avons soutenu ailleurs (cf. Bronckart & Schneuwly, 1991), toute intervention didactique implique d'abord une prise en compte de la «situation» de l'enseignement d'une matière, c'est-à-dire de l'histoire dont elle procède aussi bien que des contraintes actuelles du système scolaire en lequel elle s'insère. Et l'histoire des pratiques scolaires, tout comme les représentations des agents du système (parents, enseignants, élèves) restent profondément marquées par la conception représentationnelle du langage évoquée plus haut. Aucune réforme didactique ne peut prétendre modifier d'emblée et radicalement cet état des choses; elle doit, au contraire, en tenir compte et s'y articuler.



Toute proposition de rénovation didactique doit en outre tenir compte des processus d'apprentissage et de développement mis en œuvre par l'enfant-élève. L'acquisition de connaissances ou de pratiques nouvelles ne procède ni de la simple reproduction guidée (et fractionnée) des modèles sociaux, comme le pensait la tradition scolastique et comme le pensent certains behavioristes, ni de la simple auto-réalisation d'une puissance cognitive toujours déjà là (d'une compétence innée), comme le pensent les néo-nativistes d'inspiration chomskyenne. Au contraire, comme l'ont montré les travaux de Piaget (1936) aussi bien que de Vygotsky (1926, 1934), le développement de connaissances et de pratiques nouvelles exige certes d'abord le contact avec les modèles à acquérir, mais il procède aussi, chez l'apprenant, par généralisation et par conceptualisation, c'est-à-dire par construction de systèmes de représentations successifs.

Face à l'«impossibilité sociale» de modifier radicalement le statut et la prééminence de l'enseignement grammatical, et en tenant compte des processus d'apprentissage effectivement mis en œuvre par les élèves, il nous semble que l'enseignement de la langue ne peut qu'évoluer dans la direction d'un **compromis**, et en l'occurrence, se déployer selon deux axes parallèles. D'une part, sur la base d'un corpus de phrases ou d'énoncés «triés», la mise en place d'activités d'inférence et de codification aboutissant à une maîtrise des principales notions et règles du système de la langue (par exemple, les règles de structuration des groupes, les caractéristiques de la conjugaison du verbe, les règles de construction de la phrase passive, etc.). D'autre part et simultanément, sur la base d'un corpus de textes empiriques eux aussi «triés», la mise en place d'activités de sensibilisation aux conditions de fonctionnement des textes dans leur contexte communicatif, débouchant, localement, sur la conceptualisation de quelques règles de planification et de textualisation (par exemple, les conditions d'emplois des temps des verbes, des organisateurs, des unités anaphoriques, etc.).

Le compromis auquel nous adhérons est donc celui d'un passage progressif d'une didactique du «système» de la syntaxe phrastique à une didactique du fonctionnement textuel.

C'est donc la maîtrise du fonctionnement textuel qui constitue l'objectif ultime de la didactique de la langue, objectif à atteindre en fin de scolarité obligatoire. Et sans entrer ici dans le détail de la méthodologie adaptée à cet objectif (que nous concevons comme la construction de **séquences didactiques**, à laquelle plusieurs de nos collègues travaillent à Barcelone et à Genève notamment [cf. Camps, 1994; Dolz, 1995a, 1995b; Dolz & Pasquier, 1996; Schneuwly & Bain, 1994]), on peut *in abstracto*, c'est-à-dire en laissant de côté les problèmes de progression pédagogique, organiser le programme de didactique du fonctionnement textuel en quatre rubriques générales.

La première concerne le statut de **production sociale** de toute forme de texte, ou encore le problème du rapport des textes à l'activité humaine en général. Ce premier chapitre de la didactique du texte implique le développement de réflexions et d'activités autour des thèmes suivants.



– L'analyse des **actions langagières**, comme **unités psychologiques synchroniques**, fédérant les représentations dont dispose un agent à propos du contexte de son action, dans ses aspects physiques, sociaux et subjectifs, ainsi qu'à propos du contenu thématique à transmettre.

– La mise en évidence du fait qu'une action langagière ne peut se réaliser effectivement, dans le cadre d'une formation sociale donnée, que par l'exploitation des *formes communicatives* en usage dans cette même formation; en d'autres termes, qu'elle requiert l'emprunt à ces **construits historiques** que constituent les **genres de textes**. Disponibles dans l'*intertexte*, ces genres s'adaptent en permanence à l'évolution des enjeux socio-communicatifs, et ils sont dès lors porteurs de multiples **indexations sociales** (valeurs de pertinence et d'appropriété eu égard aux situations d'action).

– La mise en évidence donc de ce que la personne qui entreprend une action langagière doit nécessairement mettre en **interface** la connaissance de sa situation d'action et celle des genres de textes, tels qu'ils sont indexés dans l'intertexte et tels qu'ils mobilisent les ressources et les pré-construits particuliers d'une langue naturelle. Ce processus débouche sur la production d'un **texte empirique**, qui, de la sorte, est toujours un produit de la **dialectique** qui s'est instaurée entre représentations synchroniques des contextes d'action et représentations relatives à ces instruments historiques que constituent les langues et les genres de texte. Et cette dialectique est **productive** de ces propriétés nouvelles et irréductiblement singulières, qui définissent le **style** propre à chaque texte empirique.

Les autres chapitres de la didactique du texte concernent son **architecture interne**, ou encore les trois niveaux d'organisation de ce que nous qualifions de «feuilleton textuel»: l'*infrastructure générale du texte*, les *mécanismes de textualisation* et les *mécanismes de prise en charge énonciative*.

Le niveau le plus «profond» de l'**infrastructure textuelle** est constitué par le plan général du texte, par les types de discours qu'il comporte, par les modalités d'articulation de ces types de discours, ainsi que par les séquences qui y apparaissent éventuellement. Le plan général concerne l'organisation d'ensemble du contenu thématique du texte; il se donne à voir dans le processus de lecture et peut être codifié dans un *résumé*. La notion de type de discours désigne les sortes de segments que comporte le texte: segments de *discours interactif*, de *récit interactif*, de *discours théorique* ou encore de *narration*, tels que nous les avons décrits et théorisés dans *Le fonctionnement des discours* (1985). La notion de séquence (cf. Adam, 1992) désigne des modes de planifications, plus conventionnelles ou plus spécifiquement langagières, qui se déploient à l'intérieur du plan général de texte (séquences narrative, explicative, argumentative, etc.). Le statut de ces séquences, leur espace de validité (locale ou globale) et les relations d'interdépendance qu'elles entretiennent avec les types de discours constituent, comme l'a montré Adam (1992), un domaine de réflexion particulièrement intéressant.



Les **mécanismes de textualisation** se superposent à l'infrastructure textuelle; en créant des *séries isotopiques*, ils contribuent à l'établissement de la *cohérence thématique* du texte. Ils en explicitent, ou en «marquent», à l'intention du destinataire, les grandes articulations hiérarchiques, logiques et/ou temporelles. En accord avec la plupart des travaux en ce domaine, nous y distinguons trois sous-ensembles.

– Les mécanismes de connexion contribuent au marquage des articulations de la progression thématique; ils sont réalisés par des *organiseurs textuels* qui peuvent s'appliquer au plan général du texte, aux transitions entre types de discours, aux transitions entre phases d'une séquence, ou encore aux articulations plus locales entre phrases et/ou propositions.

– Les mécanismes de cohésion nominale ont pour fonction d'une part d'introduire les thèmes et/ou les personnages nouveaux, et d'autre part d'assurer leur «reprise» ou leur «relais» dans la suite du texte. Les unités réalisant ces mécanismes sont qualifiées d'*anaphores*; il peut s'agir de pronoms, personnels ou relatifs, de démonstratifs, de possessifs, mais aussi de certains syntagmes nominaux, qui le plus souvent incluent un déterminant spécifique.

– Les mécanismes de cohésion verbale enfin assurent l'organisation temporelle et/ou hiérarchique des procès (états, événements ou actions) verbalisés dans le texte, et ils sont essentiellement réalisés par les *temps des verbes*, marques morphologiques qui sont cependant en interaction avec les types de lexèmes verbaux auxquels elles s'appliquent, ainsi qu'avec d'autres unités à valeur temporelle (adverbes et organisateurs textuels notamment).

Les mécanismes de **prise en charge énonciative** contribuent enfin, à un troisième niveau, à la confection de la cohérence pragmatique, ou interactive, du texte. Moins dépendants de la linéarité et de la progression textuelles, et visant plus directement le destinataire, ils rendent explicites les *voix* qui s'expriment dans le texte et les rapports que ces dernières entretiennent avec l'agent producteur. Et ils formulent, au travers des ressources subtiles et linguistiquement très diverses de la *modalisation*, les prises de position, les évaluations et les appréciations que ces voix introduisent à l'égard du contenu thématique.

De la didactique du texte à la didactique de la littérature

Il me reste maintenant à démontrer comment chacun de ces quatre chapitres de la didactique du texte peut constituer un lieu d'articulation ou encore un lieu d'enclenchement d'une didactique de la littérature articulée aux objectifs que nous avons retenus.



Le statut de production sociale des textes

Dans le cadre de cette phase ultime de la didactique de la langue que constitue la didactique du fonctionnement textuel, on peut analyser les mécanismes que nous qualifions d'**interfaçage** entre représentations des situations d'actions langagières et connaissances des genres de textes appropriés à ces situations. Au travers d'activités diverses (de comparaison de corpus de textes, d'analyse des différents aspects de leur contexte, d'écriture en réponse à une consigne précise, etc.), on peut ouvrir la problématique de l'**indexation sociale** des genres textuels, la discuter de manière globale, en termes d'efficacité et d'appropriété à des situations de communication. Mais cette problématique débouche inéluctablement sur celle des **valeurs** qui traversent et orientent ces indexations sociales.

Le pas à franchir pour entrer dans la didactique de la littérature consiste alors à resaisir cette problématique de l'indexation sous la forme spécifique qu'elle prend dans *le champ littéraire*, c'est-à-dire dans l'articulation entre des œuvres désignées dans ce champ et les commentaires évaluatifs qui y sont proposés. La didactique de la littérature commence lorsque le débat se porte sur les critères et les arguments qui conduisent à l'attribution du statut de littérarité; arguments d'ordre historique, sociologique, psychologique, philosophique et linguistique, qui sont sans doute tous contestables, mais qui sont en même temps tous intrinsèquement légitimes. Et c'est la confrontation à ces arguments, c'est l'entrée de l'élève dans le débat interprétatif, qui est la condition même (nécessaire, mais certes non suffisante) de la réalisation des cinq grands objectifs éducatifs évoqués plus haut.

L'infrastructure textuelle

Je me limiterai à une seule des dimensions de cette infrastructure, celle des types de discours, que nous avons beaucoup étudiés.

Dans le cadre de la didactique du fonctionnement textuel, on peut examiner les différents segments qui composent un texte relevant d'un genre donné. Dans un *conte* par exemple, on peut distinguer les segments proprement narratifs, des segments de discours interactifs ou dialogués ou encore des segments de commentaires théoriques. Au travers d'activités diverses (lectures, productions, exercices organisés selon la logique d'une séquence didactique, etc.), on peut mettre en évidence les unités linguistiques spécifiques ou discriminatives de ces types de discours, et on peut examiner aussi les modalités d'articulation de ces types dans le texte qui les englobe.

Le pas à franchir pour entrer dans la didactique de la littérature est d'examiner les mécanismes qui sous-tendent les distributions différentielles d'unités linguistiques, en l'occurrence les décisions qui président à la création de mondes discursifs, en soi radicalement «autres» que le monde ordinaire des auteurs et des lecteurs: – monde narratif, dont les coordonnées temporelles et spatiales radicalement disjointes de



l'espace-temps de la production textuelle rendent notamment possible la fiction; – monde théorique et son parti pris d'universalité et d'a-temporalité; – monde du discours interactif qui, comme les dialogues de théâtre le montrent parfaitement, épure et idéalise les formes d'interaction verbale à l'œuvre dans le monde ordinaire. L'analyse linguistique des types discursifs peut ainsi déboucher, dans la continuité, sur une analyse des processus de désancrage-réancrage qui, nous l'avons vu, contribuent à la définition du littéraire; désancrage-réancrage qui posent les connaissances, les événements ou les actions dans des coordonnées autres que celles du monde ordinaire, pour mieux les faire voir.

Les mécanismes de textualisation

L'exemple le plus parlant est ici celui de la cohésion verbale, ou encore celui de la fonction effective que remplissent les temps des verbes dans l'organisation textuelle.

A prendre l'exemple de la temporalité des narrations, on peut, dans le cadre de la didactique du fonctionnement textuel, et au travers de toutes les sortes d'activités que l'on voudra, s'intéresser au rôle que jouent les différents temps dans le marquage des phases du schéma narratif, dans la distinction d'un avant-plan et d'un arrière-plan (par la distribution des passés simples et imparfaits en français) ou encore dans l'établissement de mises en relations temporelles locales (par exemple, l'expression de l'antériorité d'une action par rapport à une autre, marquée par l'opposition temps simple/temps composé, associée à un enchâssement syntaxique).

Le pas à franchir pour entrer dans la didactique de la littérature consiste, à nouveau, à se pencher sur les mécanismes qui sous-tendent effectivement ces marquages. A montrer que la conception traditionnelle des relations temporelles (un rapport de simultanéité, d'antériorité ou de postériorité qui serait posé entre le moment de la production et le moment du procès verbalisé dans un texte) n'a aucune pertinence pour comprendre les mécanismes en jeu dans la narration. Dans ce type de discours, le rapport à l'espace-temps de production a été posé une fois pour toutes, et il est définitoire ou constitutif du type même. Dès lors, l'organisation temporelle d'une narration (comme d'ailleurs d'un discours théorique) a une finalité exclusivement interne. En l'occurrence, l'analyse des conditions d'emploi des temps renvoie toujours, inéluctablement, au rapport qui est posé entre le décours de l'activité narrative elle-même (ou activité du narrateur) et le décours de la diégèse ou du narré. Rapport qui peut présenter une allure tout à fait mécanique, comme dans les polars de bas de gamme que nous avons analysés avec I. Plazaola Giger (1993), mais rapport qui peut être aussi le lieu d'interventions de l'auteur, de détournements et de déplacements des rapports entre activité narrative et narré, dont les ouvrages de Simenon, notamment, constituent un exemple particulièrement clair. Et ce sont précisément les interventions de cet ordre qui peuvent conduire à estimer que ces polars-là sont aussi de la littérature.



Les mécanismes de prise en charge énonciative

En ce chapitre, à part peut-être la mise en évidence des multiples formes de réalisation des modalisations, il me semble que l'on entre très rapidement dans des problématiques qui sont au cœur même de la littérature.

Les activités d'identification des *voix* s'exprimant dans un texte, les tentatives de mise en évidence des mécanismes de gestion de ces voix par les instances formelles que constituent le *narrateur* ou l'*énonciateur* font entrer de plain-pied dans les enjeux les plus profonds de la littérature. Parce que ces instances, les voix qu'elles distribuent et les modalisations que ces voix régissent, ne sont en définitive rien d'autre que des traces du processus interprétatif ou évaluatif lui-même: évaluation du statut des thèmes mis en discours (certain, probable, heureux, etc.); évaluation de l'origine et des mobiles de cette évaluation première; évaluation du statut et de la légitimité même de cette seconde évaluation, et ainsi de suite, dans cette mise en abysse permanente en laquelle se déploie le travail interprétatif de l'humanité sur elle-même.

Mise en abysse d'une complexité sans commune mesure avec les quelques propositions que je viens de formuler. Complexité qui me conduit donc, sagement, à clore cette contribution.



Références bibliographiques

- Adam, J.-M. (1992). *Les textes: types et prototypes*. Paris: Nathan.
- Bakhtine, M. (1979). *Esthétique de la création verbale*. Paris: Gallimard.
- Balibar, R. (1974). *Les français fictifs*. Paris: Hachette.
- Bourdieu, P. (1977). La production de la croyance: contribution à une économie des biens symboliques. *Actes de la recherche en sciences sociales*, 13, 3-43.
- Bourdieu, P. (1980). *Le sens pratique*. Paris: Minit.
- Bronckart, J.-P. (1985). Vygotsky; une œuvre en devenir. In: B. Schneuwly et J.-P. Bronckart (Ed.), *Vygotsky aujourd'hui* (pp. 7-21). Paris: Delachaux et Niestlé.
- Bronckart, J.-P. (1992). El discurso como accion. Por un nuevo paradigma psicolinguístico. *Anuario de Psicología*, 54, 3-48.
- Bronckart, J.-P. (1995a). Theories of action, speech, natural language and discourse. In: J. V. Wertsch & al. (Ed.), *Sociocultural Studies of Mind* (pp. 75-91). Cambridge: Cambridge University Press.
- Bronckart, J.-P. (1995b). Du behaviorisme à l'interactionisme social. In: H. Lejeune & al. (Ed.), *Des animaux et des hommes, Hommage à Marc Richelle* (pp. 255-292). Paris: PUF.
- Bronckart, J.-P. (1996). Units of analysis in Psychology and their Interpretation; Social interactionism or Logical Interactionism. In A. Tryphon & J. Vonèche (Ed.), *Piaget-Vygotsky: The Social Genesis of Thought* (pp. 85-106). New York: Erlbaum, Taylor et Francis.
- Bronckart, J.-P., Bain, D., Schneuwly, B., Davaud, C. & Pasquier, A. (1985). *Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*. Paris: Delachaux et Niestlé.
- Bronckart, J.-P., Clémence, A., Schneuwly, B. & Schurmans, M.-N. (1996). Manifesto. Reshaping humanities and social sciences: A Vygotskian perspective. *Swiss Journal of Psychology*, 55, 74-83.
- Bronckart, J.-P. & Schneuwly, B. (1991). La didactique du français langue maternelle; l'émergence d'une utopie indispensable. *Education et recherche*, 13, 8-26.
- Camps, A. (1994). *L'ensenyament de la composició escrita*. Barcelona: Barcanuova.
- Chomsky, N. (1970). *Le langage et la pensée*. Paris: Payot.
- Chomsky, N. (1985). *Règles et représentations*. Paris: Flammarion.
- Dolz, J. (1995a). L'apprentissage des capacités argumentatives: étude des effets d'un enseignement systématique et intensif du discours argumentatif chez des enfants de 11-12 ans. *Bulletin Suisse de Linguistique appliquée*, 61, 137-169.
- Dolz, J. (1995b). Escribir textos argumentativos para mejorar su comprensión. *Comunicacion, Lengua y Educacion*, 26, 65-77.
- Dolz, J. & Pasquier, A. (1996). Un decalogo para enseñar a escribir. *Cultura y Educacion*, 2, 31-41.
- Habermas, J. (1987). *Théorie de l'agir communicationnel*. Paris: Fayard.
- Hegel, F. (1807). *Phénoménologie de l'Esprit*. Traduction française (1947). Paris: Aubier.
- Hjelmslev, L. (1968). *Prolégomènes à une théorie du langage*. Paris: Editions de Minit.



- Marghescou, M. (1974). *Le concept de littérarité. Essai sur les possibilités théoriques d'une science de la littérature*. La Haye: Mouton.
- Marx, K. & Engels, F. (1846). *L'idéologie allemande*. Traduction française (1972). Paris: Editions sociales.
- Mauro, T. (de) (1969). *Une introduction à la sémantique*. Paris: Payot.
- Pennac, D. (1992). *Comme un roman*. Paris: Gallimard.
- Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Neuchâtel: Delachaux et Niestlé.
- Plazaola Giger, I. & Bronckart, J.-P. (1993). Le temps du polar. *Langue française*, 97, 14-42.
- Reuter, Y. (1990). Définir les biens littéraires, *Pratiques*, 67, 5-14.
- Reuter, Y. (1992). Enseigner la littérature? *Recherches*, 16: *Apprendre la littérature?* 55-70. Lille.
- Reuter, Y. (1996). Pourquoi enseigner la littérature? In: *Actes du Colloque «Les représentations de la littérature dans les institutions d'enseignement et les pratiques didactiques»* (pp. 45-54). Tours: Université de Tours.
- Ricœur, P. (1983). *Temps et récit*, t. I. Paris: Seuil.
- Ricœur, P. (1984). *Temps et récit*, t. II: *La configuration dans le récit de fiction*. Paris: Seuil.
- Ricœur, P. (1985). *Temps et récit*, t. III: *Le temps raconté*. Paris: Seuil.
- Ricœur, P. (1986). *Du texte à l'action; essais d'herméneutique II*. Paris: Seuil.
- Rivarol (1784). *Discours sur l'universalité de la langue française*. Réédition (1936). Paris: Larousse.
- Saussure, F. (de) (1916). *Cours de linguistique générale*. Paris: Payot.
- Schneuwly, B. (1994). Genres et types de discours: considérations psychologiques et ontogénétiques. In: Y. Reuter (Ed.), *Les interactions lecture-écriture* (pp. 155-173). Bern: P. Lang.
- Schneuwly, B. & Bain, D. (1994). Mecanismos de regulacio de las activitats textuales. *Articles*, 2, 87-104.
- Spinoza, B. (1975). *Traité de la réforme de l'entendement*. Traduction française (1964). Paris: Flammarion.
- Vygotsky, L. S. (1926). La signification historique de la crise de la psychologie. In: (1982), *Œuvres choisies, Tome I*. Moscou: Editions pédagogiques.
- Vygotsky, L. S. (1934). *Myschlenie y rech'*. Traduction française, F. Sève (1985). *Le langage et la pensée*. Paris: Editions sociales.
- Weber, M. (1971). *Economie et société*. Paris: Plon.



Didactique des sciences: faut-il dépasser le constructivisme?

André Giordan

*L'homme est une machine à apprendre...
F. Jacob*

Vingt ans après sa mort, les travaux de Jean Piaget demeurent en pleine actualité. Les méthodes dites «constructivistes» qu'il a contribué à promouvoir se répandent dans l'enseignement. Aux Etats-Unis, sous l'impulsion du Prix Nobel américain de physique Lederman, des méthodes actives ont été promues dans les classes primaires de Chicago. En France, le Prix Nobel français Georges Charpak et l'astrophysicien Pierre Léna tentent de mettre en place un ensemble d'activités scientifiques dans les écoles françaises sur ce modèle.

Dans les musées, nombre de lieux d'investigations pour jeunes ont été construits en référence à Piaget. Parmi les plus achevés, citons le *Children Museum* de Boston, l'*Exploratorium* de San Francisco ou encore la *Cité des Enfants* de la Cité des sciences et de l'industrie de la Villette.

Toutes ces pratiques éducatives ou muséales partent des besoins spontanés, des questions et des intérêts «naturels» des apprenants. Elles prônent leur libre expression, leur créativité et leur savoir-être. Elles mettent en avant une certaine découverte autonome de l'enfant ou encore l'importance des tâtonnements dans l'acte d'apprendre.

L'individu – enfant ou adulte, puisque ce type de formation a des prolongements dans la formation continue ou à l'université comme au Québec et aux Etats-Unis – ne se contente plus de recevoir des données brutes, il les sélectionne et les met en forme.

Une construction du savoir s'opère ainsi au travers d'actions sur des objets ou des phénomènes, la résolution de problèmes ou encore la mise en place de projets. Les observations, les manipulations, les enquêtes et les petites expériences sont privilégiées. Chaque fois, l'expression des représentations des apprenants sur les phénomènes, le travail de groupe, facilitent l'acquisition du savoir ou développent les compétences.



A travers ses multiples travaux, Piaget a eu l'immense mérite de montrer qu'apprendre n'est pas le résultat d'un simple processus de «transmission-réception». Le savoir ne peut être comparé à de simples empreintes que des stimulations sensorielles laisseraient dans l'esprit de l'élève, à la manière de la lumière sur une pellicule photographique.

Pour lui, apprendre est le résultat d'une séquence à la fois déterminée génétiquement et dépendante de l'activité d'un sujet sur des objets. De fait, le savoir se construit au travers de l'équilibration de schèmes mentaux déjà en place. Deux mécanismes y contribuent principalement: ceux d'*assimilation* et d'*accommodation*, empruntés à la biologie de l'évolution.

Tout organisme assimile ce qu'il prend de l'extérieur à ses propres structures, y compris les informations récupérées par ses perceptions. En retour, ce processus s'accompagne d'une accommodation, c'est-à-dire d'une modification des organes sur un plan biologique ou des instruments intellectuels sur le plan cognitif. Si le sujet veut assimiler un savoir, il doit être capable d'accommoder en permanence son mode de pensée aux exigences de la situation.

Une longue tradition

En fait, la pensée de Piaget s'inscrit dans une longue tradition de pensée. Celle-ci remonte à la fin du XVIII^e siècle, avec principalement Emmanuel Kant. Dans sa *Critique de la raison pure* (1781), il soutient que le savoir dépend des sens, tout comme le proclamait le philosophe anglais John Locke, contrairement aux Rationalistes qui ne juraient que par une raison innée. Toutefois, il n'évacue pas totalement le fonctionnement de la raison. Seule cette dernière possède les conditions requises pour interpréter ce que nous percevons du monde.

Repris par la psychologie américaine de la fin du XIX^e et du début du XX^e siècle avec John Dewey, Herbert Spencer et Williams James, puis par un autre genevois, Edouard Claparède, un rôle très important est accordé au «sujet», et en particulier à ses «ressorts intérieurs» dans le développement cognitif. Pour tous, comme le proclame Spencer, «*l'éducation doit se conformer à la marche naturelle de l'évolution mentale*».

Des écoles «nouvelles» se créent alors un peu partout en Europe sous l'influence de psychologues, Binet en France, Marbe en Allemagne; elles proposent de soumettre l'éducation «à l'observation précise de la pensée elle-même».

Sur le plan pédagogique, de nombreux courants pédagogiques s'en sont inspirés plus ou moins directement. De Montessori à Freinet en passant par Audemars et Lafendel, Decroly, Ferrière, le Groupe Français de l'Ecole nouvelle (GFEN), tous se centrent



sur l'apprenant. Tous militent contre un apprentissage par «émission-mémorisation» qui privilégie l'enseignement au dépend de l'apprentissage.

Avec l'apport spécifique de Piaget – les stades du développement de l'enfant et une méthodologie à toute épreuve, la démarche clinique –, l'école genevoise essaime un peu partout dans le monde.

Daniel Bain a mené l'essentiel de son activité de chercheur en sciences de l'éducation au Centre de recherches psychopédagogiques (CRPP) du Cycle d'orientation de Genève. Discret, mais très efficace, Daniel Bain a toujours voulu être un chercheur de base, au plus près des multiples questions du terrain. Cela ne l'a pas empêché de réaliser des contributions importantes, dont certaines ont eu un impact notable auprès des enseignants et des décideurs. Pour mener à bien ses investigations, il a su générer et favoriser des équipes d'enseignants, tout en maintenant un lien fort avec la hiérarchie scolaire.

Le Laboratoire de Didactique et Epistémologie des Sciences (LDES) et moi-même souhaitons lui rendre hommage d'abord en tant que chercheur. Dans nos domaines, il s'est questionné de façon épistémologique sur les problèmes psychopédagogiques de la mesure (erreurs, biais, fautes) et sur leur généralisabilité.

Dans le cadre plus limité de la didactique des sciences, il a mis au point et promu tout à tour l'observation scientifique expérimentale et la modélisation; il a contribué à mettre à plat les conceptions et les obstacles des élèves dans les processus d'élaboration de savoirs.

Mais le LDES souhaite surtout le remercier pour le rôle inestimable qu'il a joué dans le démarrage du Laboratoire. Lors de mon arrivée à Genève, il a grandement favorisé les contacts avec les équipes de terrain. Au cours de ces vingt années, ces interactions se sont renforcées. Nombre de travaux aujourd'hui connus et souvent cités, réalisés par les collaborateurs successifs du Laboratoire, ont eu pour point de départ: Daniel Bain.

Progressivement, l'approche «constructiviste» qui étudie les processus mentaux gagne du terrain, au point de supplanter totalement le modèle behavioriste qui s'intéressait aux comportements d'apprentissage observables. Au cours du temps, ce mouvement est d'ailleurs devenu fort complexe, avec de multiples variantes pour traiter des relations entre l'objet et le sujet. Par exemple, Gagné et Jerome Seymour Bruner, deux psychologues américains, vont mettre l'accent sur les «associations» à établir entre les informations externes et la structure de pensée. Pour eux, toute perception est d'abord une catégorisation. Ausubel, également américain, est plus explicite: il parle d'ancrage ou encore de «ponts cognitifs» (*cognitive bridge*) à réaliser entre les informations et l'élève.

Pourtant, au moment où ce courant de pensée se développe, prend de l'ampleur, nombre de limites se font jour. L'éducation scientifique est particulièrement en avant



sur ce plan, avec le développement des recherches didactiques depuis vingt ans. Par exemple, nos travaux montrent que les modèles «constructivistes» restent plutôt frustes pour décrire les multiples mécanismes de l'apprendre. Tout ne dépend pas des seules structures cognitives générales. Des étudiants, des chercheurs qui ont accès au formalisme logico-mathématique, le stade le plus avancé de Piaget, peuvent raisonner à l'égal d'enfants de 6-7 ans sur des contenus inhabituels.

Plus les situations sont éloignées des savoirs maîtrisés d'ailleurs, plus les individus utilisent des stratégies de raisonnement primitives. Ce qui est en cause, ce n'est pas seulement un mode opératoire, mais ce que nous appelons une «conception» de la situation. Interviennent à la fois un type de questionnement, un cadre de références, une façon de raisonner ou de produire du sens... En se limitant à décrire des fonctionnements généraux et des états d'équilibre (les «stades»), les modèles constructivistes ne rendent pas compte du traitement de situations spécifiques par les apprenants. Ils ne permettent pas d'interpréter, à eux seuls, les inférences que ceux-ci peuvent faire à partir des informations dont ils disposent.

Cette dernière critique a conduit des théoriciens à formuler des hypothèses annexes pour maintenir la théorie de la construction propre à Piaget. Par exemple, Jerry Fodor suppose des systèmes de traitement indépendants, appelés «modules», à l'origine d'études distinctes sur la perception, la mémoire ou le langage. De leur côté, Newell et Simon ont réduit le fait de penser à «traiter de l'information», et par là à manipuler des symboles.

En jouant sur l'analogie «cerveau-ordinateur», l'arrivée de l'informatique a conduit à un autre mouvement de pensée: *l'intelligence artificielle*. Le jeu d'échecs a été la grande application qui l'a popularisée. Les «représentations mentales» – un moment oubliées – sont remises au centre des débats, relayées par le *connexionnisme* qui réintroduit un changement de niveau dans la description du système cognitif. Les états mentaux redeviennent, comme au milieu du XIX^e siècle, les propriétés émergentes du système neuronal. Un lien peut être établi à nouveau avec la neurologie et les psychophysiologies.

Les limites du constructivisme

En fait, rien n'est immédiatement accessible dans l'apprendre. L'appropriation d'un savoir ne se réalise pas uniquement de façon automatique par assimilation et accommodation ou encore par abstraction «réfléchissante», comme le supposent Piaget et son école. C'est une vue trop optimiste ou idéalisée. Pour les apprentissages de concepts ou de démarches un tant soit peu complexes, une nouvelle information s'inscrit rarement dans la ligne des savoirs maîtrisés. Au contraire, ceux-ci constituent autant d'obstacles sur les plans cognitif et émotionnel.

Prenons un exemple: la nutrition des êtres vivants. Quand un élève apprend le comportement alimentaire d'un animal proche, apprendre consiste seulement à relier du



nouveau à ce qu'il sait ou croit, éventuellement à relier du nouveau à ce qu'il fait déjà. L'apprenant possède déjà un cadre de références assuré sur la nutrition. Il dispose de procédures de raisonnement ou d'activités. Il ne fait que compléter son savoir ou ses possibilités d'action. Dans un tel cas, l'information peut lui être fournie sans problème par simple transmission. Il peut même l'enregistrer directement. Il possède tous les éléments et tous les moyens cognitifs pour la décoder et la mobiliser. Une seule condition est fortement limitante, c'est la motivation.

S'il s'agit maintenant de la nutrition d'un animal plus lointain, un oursin, un moustique, une méduse ou une paramécie, l'apprentissage devient déjà moins évident. «Où est la bouche?», «Avaler un liquide, est-ce que c'est manger?», «Par où sortent les excréments?», «Que veut dire digérer?». Une ou plusieurs de ces questions peuvent perturber l'individu. Pour apprendre, il va devoir assimiler des informations nouvelles, moins évidentes. Eventuellement, il accommode son cadre de référence si celui-ci n'apparaît plus totalement pertinent. Par ce cheminement, il étend et restructure son équipement cognitif. Dans ce cas le modèle piagétien est opérant.

Prenons maintenant un apprentissage plus complexe, la nutrition des plantes. Ses savoirs antérieurs deviennent totalement incompatibles avec le message à apprendre. L'assimilation immédiate est impossible. Les savoirs en place font obstacle: pour les apprenants, «la plante se nourrit dans le sol par les racines». La lumière reste quelque chose de mystérieux pour l'apprenant, parfois «un fortifiant». Elle «agit par sa chaleur». Quant au gaz carbonique, il est un «produit nocif» envisagé dans une respiration.

Un important travail de déconstruction est à réaliser. Les arguments convaincants ne sont pas immédiatement évidents. De plus, des interférences apparaissent: les élèves doivent se décentrer des racines pour envisager le rôle des feuilles. Mais cette décentration ne peut pas être totale, les racines jouent un rôle indispensable pour l'absorption de l'eau, des sels minéraux et même du gaz carbonique...

L'apprentissage devient encore plus complexe quand il s'agit de questions de génétique, de physique quantique, de génétique des populations, de régulation ou encore d'environnement et de santé. Dans ces derniers cas interviennent, en plus, des apprentissages d'attitudes ou de valeurs. Ces dernières sont encore moins faciles à transformer: savoir ne signifie pas modifier automatiquement son comportement...

Une déconstruction des conceptions initiales de l'apprenant devient ainsi une étape préalable indispensable. Or, contrairement à ce que suggérerait naïvement Gaston Bachelard, cette démarche est de fait pratiquement impossible. L'apprenant ne se laisse pas facilement déposséder de ses opinions et de ses croyances qui se révèlent être autant de compétences. Construction et déconstruction ne peuvent être que des processus parallèles, de fait interactifs, voire antagonistes. Le nouveau savoir ne s'installe véritablement que quand l'antécédent apparaît périmé. Entre-temps, le savoir antérieur, seul outil à disposition de l'apprenant, a servi de cadre interprétatif.



Par ailleurs, il faut ajouter que les tenants du constructivisme ont isolé trop longtemps l'individu apprenant. Or l'expérience de chacun ne se construit pas seulement avec des objets mais dans un environnement. En mettant l'accent sur les seules capacités cognitives, ils minimisent la place et le rôle du milieu.

Certains constructivistes, au travers de leur expérimentation, en vinrent même à ignorer que le développement se réalise au sein d'une société¹². Le milieu culturel contribue pourtant à donner du sens aux situations. Quant à la sphère affectivo-émotionnelle, si elle n'est niée par personne, elle n'a pas non plus été prise en compte, faute de modèle explicitant les liens entre le cognitif et l'affectif¹³. Pourtant les sentiments, les désirs, les passions éventuels jouent un rôle stratégique dans l'acte d'apprendre.

Enfin, les différents constructivistes paraissent largement muets sur les contextes et les conditions qui favorisent l'apprendre, ce qui est frustrant quand on se préoccupe d'éducation ou de médiation. Cela est tout à fait normal, ce n'est pas leur préoccupation... Il faut ajouter que ce type de question appliquée ne permet pas de «belles» recherches. Au mieux, ils avancent les idées de «maturation», c'est-à-dire de développement naturel, ou d'«équilibre». Encore faut-il que le sujet puisse la réaliser; et pour cela qu'il y trouve un intérêt. Pour combler cette dernière lacune, les post-piagétiens envisagent maintenant la «co-action» ou encore le «conflit cognitif». De tels projets sont encore pauvres pour inférer des situations ou des ressources face à la complexité des enjeux d'une *éducation intégrée*. Le monde extérieur n'enseigne pas directement à l'individu ce qu'il est censé apprendre. L'individu doit inventer du sens à partir de l'environnement qu'il rencontre et en tenant compte de son histoire. Un processus de médiation devient un des paramètres indispensables. Autant d'arguments qui imposent aujourd'hui d'aller au-delà du constructivisme...

On apprend au travers de ce qu'on est et à partir de ce qu'on connaît déjà

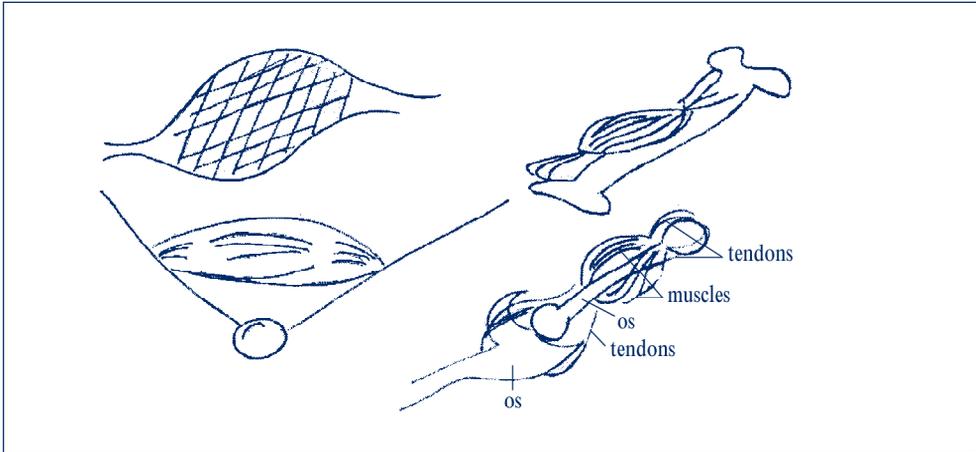
En fait, l'apprendre est une fonction complexe, non réductible à un seul modèle. Par nombre d'aspects, elle présente même de multiples composantes paradoxales. Par exemple, l'individu comprend, apprend l'environnement au travers de ses conceptions. Ces dernières sont les seuls outils qu'il maîtrise, c'est au travers d'elles qu'il décode la réalité et les informations qu'il reçoit (voir schéma ci-après). En même temps, elles sont ses «prisons» intellectuelles qui l'enferment dans une façon de comprendre le monde: une causalité linéaire, une logique classique par exemple. Pour apprendre dans une période de mutation sociale, il devra aller à l'encontre de celles-ci; mais il ne le pourra qu'en faisant avec.

¹² Cet aspect a pourtant été souligné par les psychologues Vygotsky, Wallon et plus récemment Bruner. L'enfant apprend à agir sur son environnement et à utiliser les systèmes de signification grâce aux interactions avec autrui et à la médiation sociale. Vygotsky affirme même que les capacités apparaissent d'abord en situation interindividuelle avant d'être intériorisées. Mais ces points ont été fort peu travaillés.

¹³ En fait, faute de rencontres entre les disciplines respectives retranchées dans leur territoire.



Schéma 1: Différentes conceptions d'adultes sur les muscles et leur fonctionnement



De coûteux détours sont alors indispensables pour réaliser une appropriation de l'expérience sociale¹⁴. On n'entre pas directement dans la façon de penser de l'autre. Lors d'usage de notions simples de physique classique, sous les mêmes mots (force, énergie, travail, accélération, puissance, etc.), l'expert et l'apprenant n'opèrent pas sur les mêmes réseaux de signification. Leurs expériences n'ont pas la même signification. La force qui s'applique au centre de gravité est une invention de l'esprit qui arrange bien le scientifique ou l'ingénieur. C'est une bonne idée explicative, mais il faut bien maîtriser les conditions d'application «pour que ça marche». L'objet doit être supposé cohérent; s'il est déliquescant, le concept «centre de gravité» devient inutile.

Pourtant l'enseignement, la médiation prennent tout leur sens à cet instant. Car, autre paradoxe, si l'individu doit apprendre seul, et personne ne peut le faire à sa place, l'apprenant a peu de chance de «découvrir» seul l'ensemble des éléments pouvant modifier ses questions, ses concepts ou son rapport aux savoirs. Le sens que nous attribuons aux connaissances ne peut se transmettre directement. Seuls les apprenants peuvent élaborer leurs significations propres, compatibles avec ce qu'ils sont au travers de leur expérience propre. Toutefois le médiateur peut faciliter cette production de sens en filtrant les multiples informations, en amplifiant ou réduisant l'apport des stimuli extérieurs. Il peut faciliter la comparaison, les mises en relation (temporelle, spatiale, causale) ou encore inciter à l'organisation. Si le médiateur fait défaut, ou si les médiations sont trop pauvres, parcellaires, inadaptées, incompréhensibles, l'individu ne développe pas de démarche adéquate ou ne sait pas tirer partie des données de l'environnement.

¹⁴ A propos de la lumière, l'élève du secondaire devra passer d'une vision granulaire à une conception ondulatoire. A l'université, il devra intégrer les deux conceptions.



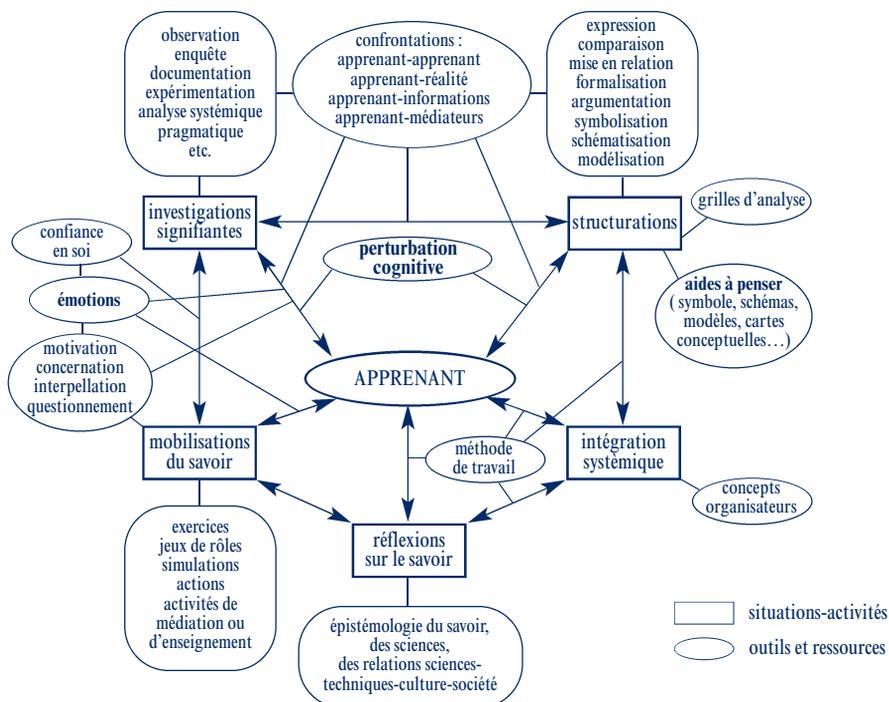
Le modèle allostérique

Enseigner, médiatiser tient plutôt d'une alchimie complexe. Le «modèle allostérique» que nous tentons de corroborer permet de catégoriser et de mettre en relation, et par là d'inférer et de prévoir le système de paramètres indispensables à l'apprendre (voir encadré). Tous sont autant de facteurs limitant l'apprentissage; celui-ci est impossible quand l'un manque à l'appel.

Environnement didactique

Une avancée dans l'apprendre ne se retrouve pas seulement dans l'environnement, comme le suggèrent les behavioristes. Elle n'est pas non plus uniquement dans l'individu, comme le proclament les constructivistes, mais dans l'interaction entre ces divers éléments. Au niveau neurologique, le cerveau ne peut modifier ses synapses ou produire des neuromédiateurs pour installer les réseaux neuroniques propres à la pensée qu'en interaction avec un contexte social.

Le modèle allostérique permet de prévoir les situations, les outils et les ressources qui facilitent l'apprendre.



L'environnement didactique favorisant l'apprendre



Tous les divers paramètres présentés ci-dessus sont à prendre en compte. On élabore par soi-même, mais pour cela il est nécessaire de confronter ses conceptions à des informations recueillies lors d'investigations multiples et variées (pragmatique, enquêtes, expérimentations, analyse systémique). Surtout, il est important de se confronter à l'expérience des autres (documentation, apports du médiateur, etc.). L'argumentation permet de prendre du recul par rapport à ses propres idées, elle fait prendre conscience de leurs limites. Par ailleurs elle favorise de nouvelles investigations pour trouver de nouveaux arguments.

En plus des activités, l'apprenant doit pouvoir disposer d'outils et de ressources (analogies, métaphores, etc.) qui l'aident à penser. Les symboles, les schémas, les modèles facilitent la structuration du savoir et sa manipulation. L'enseignant, le médiateur doit fournir des ressources qui soient lisibles, compréhensibles, adaptées au public. Par exemple, le symbolisme des chiffres romains ne permet pas la multiplication!

Un réseau de concepts organisateurs (intégration systémique) peut également servir de «colonne vertébrale» pour fédérer les multiples savoirs. Les élèves et le grand public se perdent actuellement dans la masse des informations qu'ils reçoivent sans pouvoir les situer.

Enfin, une réflexion sur le savoir lui-même, son statut, sa place dans la société, ses mécanismes, etc. est également un temps important. Nombre d'individus ne butent pas sur le savoir lui-même, mais sur l'image qu'ils se font du savoir. La moindre équation rebute le grand public parce qu'il s'est construit une image compliquée de ce savoir. Par ailleurs, les élèves n'ont pas compris les ressorts d'une démarche expérimentale par rapport à une démarche sociale ou juridique. De même, nombre d'enseignants butent sur les questions de causalité, de logique, de système parce qu'ils n'ont jamais eu l'occasion de les éprouver.

On apprend progressivement; il faut revenir plusieurs fois sur un contenu, l'aborder par différents horizons, l'affiner au contact de la réalité en mobilisant les premiers savoirs dans des actions pour tester leur efficacité. Seuls les savoirs réellement «utiles» sont retenus. Les exercices, à condition qu'ils ne soient pas seulement d'application, les simulations, les jeux de rôles sont des ressources importantes. L'enseignement à des pairs est également une forme de mobilisation. C'est en devant décomposer un savoir pour l'enseigner que l'on comprend. Pourquoi ne pas le valoriser dans l'apprendre? Chaque fois, il faut se donner les situations les plus diverses et les exploiter le plus complètement. C'est en dizaines d'heures (en milliers pour les savoirs les plus élaborés) que se compte le temps nécessaire à un apprentissage.



De même, comme tous sont en interaction, un subtil cocktail est à réaliser par l'enseignement ou la médiation. Les arguments qui convainquent sont très divers suivant les individus. Dans un cours de diététique pour diabétiques, vous pouvez souhaiter faire apprendre qu'il y a des glucides dans les poires (12% sur les tables diététiques). Si ensuite, pour des raisons de facilité de calcul, vous proposez de prendre 10%, l'ingénieur ou le maçon acceptent sans aucun problème une telle approximation. Par contre, un banquier qui a fait sa fortune en jouant sur les décimales dans les taux d'intérêt ne vous trouvera pas sérieux et pourra se trouver bloqué...

Une régulation entre ces divers paramètres par apprenant est encore à envisager. On oublie à la fois ce qui est inutile, mais également ce qui est trop intense: un traumatisme qui ébranle notre équilibre, la connaissance d'une maladie incurable crée une dissonance trop forte qui évacue toute acquisition. Pour apprendre il faut être perturbé sur ses certitudes; si on l'est trop, on en devient paralysé. Toute perturbation doit aller de pair avec une confiance en soi ou un accompagnement. On accepte d'autant mieux une perturbation cognitive qu'on a une assurance d'accompagnement suivi, etc.

Apprendre, une émergence

Apprendre devrait être ainsi envisagé aujourd'hui comme une émergence née de la rencontre de potentialités neuroniques, d'une histoire individuelle mémorisée et d'un environnement direct ou médiatisé qui permet son actualisation permanente. Sur le monde qui l'entoure, chaque individu possède des croyances très précises et met en œuvre des démarches opérationnelles. S'il n'en a pas sur la question abordée, il manipule d'autres idées sur la matière, l'énergie, la causalité qui interfèrent fortement avec la question abordée. Ce système de pensée mis en branle à propos d'un projet – que nous nommons *conception* – oriente la façon dont l'apprenant décode les informations et formule ses nouvelles idées...

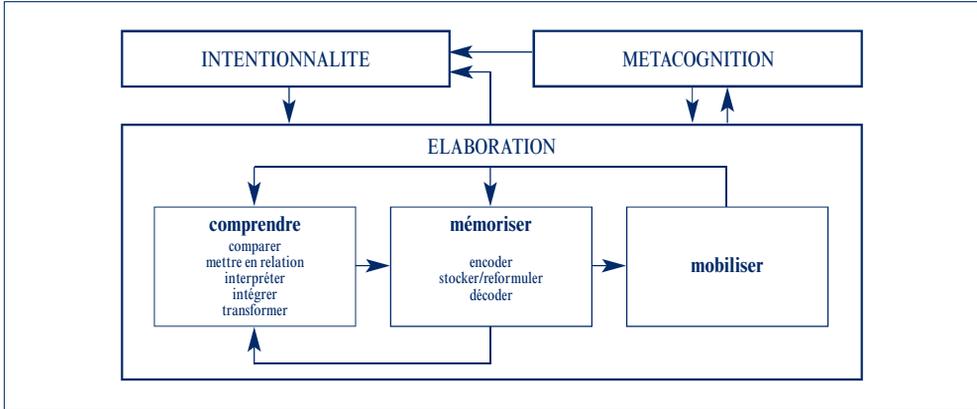
Apprendre n'est donc pas ajouter de nouvelles informations. Appréhender un nouveau savoir, c'est l'intégrer dans une structure de pensée déjà «en place» qui, oh! paradoxe, les rejette. C'est au travers de savoirs propres, antérieurs à la situation éducative (ou culturelle) mais qu'il est capable de mobiliser dans celle-ci, que l'apprenant est capable de récolter, trier et décoder les données nouvelles, éventuellement de les confronter.

Cette intégration relève d'un processus d'organisation (réorganisation) et de régulation d'éléments préalables en interaction avec des données nouvelles qui aboutira à leur métamorphose éventuelle. Toutefois, l'émergence de nouveaux savoirs n'est possible que si l'apprenant saisit ce qu'il peut en faire (intentionnalité), s'il parvient à modifier sa structure mentale quitte à la reformuler complètement (élaboration) et si ces nouveaux savoirs lui apportent un «plus» dont il peut prendre conscience (métacognition) sur le plan de l'explication, de la prévision ou de l'action. L'affectif, le cognitif et le sens se trouvent ainsi intimement liés, en régulations multiples. Et



tous trois sont régulés par des facteurs sociaux; l'apprentissage dépend fortement d'un contexte, il se réalise toujours dans un environnement socioculturel.

Schéma 2: Principales dimensions de l'acte d'apprendre



Au point de départ, apprendre nécessite toujours une intention: un projet, même implicite. Il est motivé par un besoin, un désir, un manque: une question qui interpelle l'apprenant par exemple. Il peut être direct: progresser vers ce que l'on veut pouvoir faire, répondre à une nécessité, ou indirect: avoir une bonne note, passer un examen, (se) faire plaisir, donner une certaine image de soi, etc. Tout dépend de ce qu'on a décidé d'être, de faire ou de savoir. Pour aboutir, l'apprenant se donne les moyens ou cherche les situations nécessaires. En tout cas, il n'y a pas d'apprendre sans affect, c'est le «moteur» de l'ensemble du processus. Par ailleurs, les émotions, le plaisir, le désir, sont favorables à la mémorisation. Ils apportent un poids aux informations, facilitent leur sélection, leur confèrent une importance; sans doute en liaison avec la production de neuromédiateurs lors de l'établissement des synapses.

S'il est indispensable, l'aspect émotionnel n'est pas suffisant. On peut désirer fortement tout connaître sur la mécanique quantique, encore faut-il pouvoir entreprendre les démarches cognitives correspondantes. L'apprendre procède alors d'un processus d'élaboration d'un individu confrontant les informations nouvelles et ses conceptions mobilisées, et produisant de nouvelles significations plus aptes à répondre à ses interrogations. Le changement s'opère de façon discontinue et dans une sorte de mini-crisis successives qui peuvent être parfois une crise d'identité, tant un individu a pu s'investir dans ses actes. Heureusement, le savoir progresse contre les séductions de l'apparence ou de l'évidence par optimisations successives. Lorsqu'il y a compréhension d'un nouveau modèle, la structure mentale s'est métamorphosée. Le cadre de questionnement s'est progressivement reformulé, la grille de références largement réélaboree. On ne traite plus les mêmes questions quand on passe de la génétique mendélienne à la génétique des populations. Les mêmes mots n'ont plus le même sens dans la vision macroscopique de la matière, la théorie atomico-moléculaire ou dans la théorie quantique.

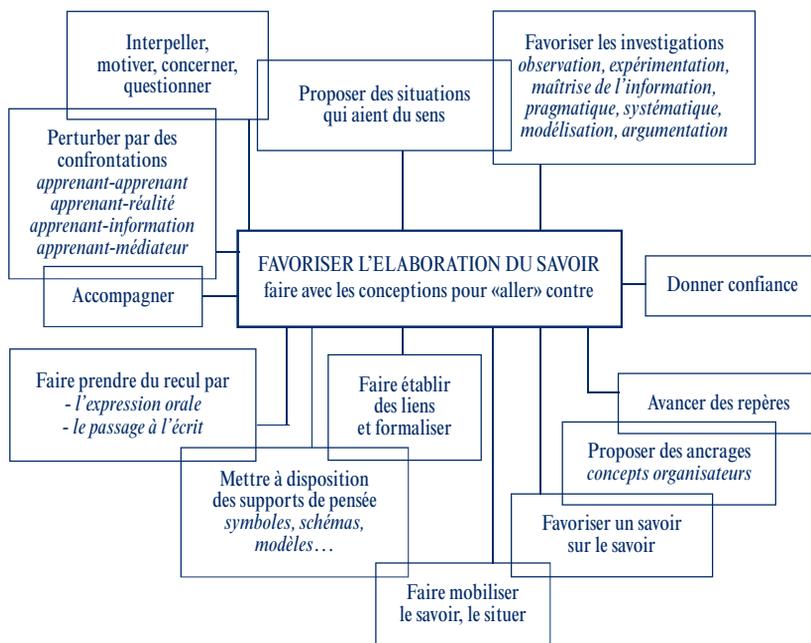


L'enseignant, un organisateur des conditions de l'apprendre

A travers les résultats de ces travaux, on peut voir combien le métier d'enseignant devient un métier sinon impossible, du moins très complexe! Seul un professionnel possédant des outils et des ressources peut l'exercer. Mais lui-même doit également transformer ses propres conceptions sur ce que signifie «enseigner», ainsi que sur la fonction même de l'enseignant.

Tout d'abord, il doit renoncer à trouver une panacée. En matière d'éducation scientifique le vaccin et le médicament n'existent pas: au mieux, c'est une multithérapie qu'il faut dispenser. L'enseignant doit savoir mettre en scène un cocktail de paramètres. Ils sont nombreux et doivent pouvoir entrer en interaction. Il doit aussi savoir les doser de façon à ce qu'ils perturbent sans totalement déstabiliser, et accompagnent sans tout à fait prendre en charge.

Enfin, il doit également savoir prendre en compte des paradoxes. Par exemple, la manière d'enseigner doit prendre appui sur l'élève, tout en allant à l'encontre de ce qu'il pense. De même, s'il faut favoriser les conditions d'une autodidaxie, il faut en même temps permettre à l'apprenant de se confronter aux situations qui l'interpellent ou sont porteuses de sens pour lui. Savoir jouer sur le plaisir tout en valorisant l'effort est également un point à ne pas oublier.



Les rôles de l'enseignant



Jusqu'à présent, l'enseignant était un simple distributeur du savoir. Il avait fait son «boulot» quand il avait dit ou montré. Dans cette nouvelle acceptation de la profession, l'enseignant devient l'organisateur des conditions de l'apprentissage. Ses tâches se situent plutôt en amont. Elles consistent à interpeller l'élève de manière à ce qu'il se sente concerné et de ce fait qu'il ait envie d'apprendre. Elles sont également d'encourager à l'effort que nécessite tout apprentissage. Partir de l'apprenant, ce n'est pas y rester. L'enseignant doit constamment proposer un projet éducatif qui ait du sens pour son public. En particulier, c'est l'enseignant qui peut créer ou amplifier les conditions d'un questionnement. Le vécu quotidien et toutes les questions qui s'y rattachent sont une source d'investigation à exploiter.

Pour tout ce qui concerne les notions proprement dites, il peut être relayé par des livres, des fiches, des films, ainsi que par l'utilisation des multimédias qui, souvent, attirent plus facilement les élèves. Par contre, lui seul peut permettre à ces derniers de prendre du recul ou de se situer. En avançant des repères, il facilite des états de questions, joue le rôle de référent que l'on vient consulter pour se situer dans un flot de données. Il peut également provoquer une réflexion sur les savoirs et sur leur place dans la société, éléments fondamentaux pour les élèves en difficulté. Comment ces élèves-là perçoivent-ils le savoir? Que considèrent-ils comme savoir et pourquoi? A quels savoirs et/ou connaissances accordent-ils de l'importance, et pourquoi? Toutes ces questions d'ordre «philosophique» devraient être abordées dès la maternelle, car elles permettent le développement d'un esprit d'analyse critique, ainsi qu'une construction de la personnalité.

Ainsi, l'enseignant ne doit pas cesser d'être un transmetteur. Mais ce qu'il a de plus important à transmettre, c'est un désir, une passion, celle d'apprendre.

Tous ces mécanismes sont différenciés selon les contenus. Ils passent par des phases de rectifications, de mutations ou éventuellement d'interférences entre conceptions mobilisées et informations filtrées. Une autre configuration se stabilise quand elle apparaît plus apte à résoudre les questions qui ont sollicité la démarche. La modification de conception est plus facile si un autre équilibre pointe à l'horizon. Un autre mode de fonctionnement plus pertinent s'est mis en place, l'individu a pu en tester son opérationnalité.

Une fois formulée, cette expérience cognitive n'est pas simplement stockée, elle doit être constamment mobilisable et mobilisée pour la suite. Heureusement, le cerveau déchire en permanence ses souvenirs pour les réorganiser en temps réel. C'est cette organisation qui investit en retour la manière d'aborder la situation nouvelle¹⁵.

¹⁵ Plus nous disposons de procédures de traitement d'informations variées, pertinentes, plus nous pouvons consacrer de ressources à l'interprétation d'une situation. L'apprenant dispose d'une pluralité d'approches de son environnement qu'il met en œuvre de façon différenciée suivant les situations et les contenus (conditions présentes). Plus la situation est incongrue, déstabilisante, originale, plus il utilise une façon de faire inférieure à ses potentialités.



L'élaboration d'un nouveau savoir ne passe pas nécessairement par la destruction des savoirs antérieurs. Le plus souvent, il faut plutôt y voir une neutralisation ou une substitution; différentes formulations peuvent cohabiter.

Enfin, la maîtrise de l'univers de l'apprendre exige encore que l'apprenant donne une signification au savoir élaboré (métasavoir). Pour parfaire un apprentissage, il importe que ce dernier prenne conscience de sa structure, de son importance et de ce qu'on peut en faire. Même inconsciemment, l'apprenant situe ses savoirs par rapport à son projet. Il les adapte à sa propre manière de faire. Cela contribue à développer une attitude favorable à l'apprendre. Comme le tout est interconnecté, le métasavoir peut constituer en retour un projet ou un plaisir.

Ajoutons encore qu'il s'agit toujours de développer un savoir biodégradable. Tout savoir qui s'installe – même le plus performant – devient à la longue dogmatique. Il conduit à une certaine rigidité mentale. Or la situation actuelle est riche d'incertitudes. La savoir doit pouvoir s'adapter en permanence pour faire face à l'approximatif, à l'incomplet, à l'imprécis et à l'imprévu. Une telle résistance n'est pas l'apanage des élèves. On peut constater combien chez les chercheurs eux-mêmes les mêmes mécanismes sont à l'œuvre. Les conceptions nouvelles ne triomphent pas, ce sont les défenseurs des anciennes qui meurent!



Références bibliographiques

Pour en savoir plus sur le modèle allostérique (par ordre de niveaux de lecture):

Giordan, A. (1998). *Apprendre!* Paris: Belin.

De Vecchi, G. & Giordan, A. (1989). *L'enseignement scientifique: comment faire pour que «ça marche»?* Nice: Z'Editions.

Giordan, A. & De Vecchi, G. (1994). *Les origines du savoir* (2^e édition). Lausanne et Paris: Delachaux et Niestlé.

Giordan, A., Girault, Y. & Clément, P. (éds.) (1994). *Conceptions et connaissance*. Bern: Peter Lang.

Giordan, A. & Girault, Y. (éds.) (1996). *New learning models*. Nice: Z'Editions.



Faut-il généraliser?

Jean Cardinet

Dans la vie courante et dans le domaine scientifique

Danger d'inférences incontrôlées

La pensée humaine repose sur l'emploi de concepts, et de relations entre concepts. Etant abstraits, ces concepts et relations peuvent s'appliquer à une infinité de situations concrètes et ils fondent ainsi nos prévisions.

Pourtant, l'extension des concepts et la véracité des relations ne peuvent pas être contrôlées pour tous les cas où on est amené à les utiliser. Chaque être humain procède à des généralisations, à partir de l'échantillon d'observations que lui a fourni son expérience. Si la loi qu'il en infère ne semble pas contredite après un certain nombre d'essais, l'homme intègre cette loi à l'ensemble de ses représentations de la réalité.

Ce mécanisme psychologique d'inférence ne satisfait pas les logiciens. Il comporte en effet un «saut», un passage à risque, lorsqu'on généralise de l'échantillon à la population. En effet, cette inférence ne peut pas faire l'objet d'une preuve, puisque la population de tous les cas possibles est infinie.

De fait, on connaît les nombreuses erreurs scientifiques auxquelles ces généralisations abusives ont donné lieu: par exemple, que «les corps tombent plus ou moins vite selon leur poids».

Les erreurs dues à ce mécanisme de généralisation sont encore plus fréquentes dans la vie courante. Certaines personnes décrètent, après une seule expérience malheureuse, que tous les ... (hommes) sont malhonnêtes, ou que tous les ... (légumes) sont immangeables!

Nécessité de connaissances générales

Pourtant, si la démarche de généralisation est universelle et même déjà présente dans tout le règne animal, c'est qu'elle a une fonction d'adaptation évidente: elle est la base du conditionnement et des premiers apprentissages. «Chat échaudé craint l'eau froide».



N'en déplaise aux logiciens, il nous faut donc généraliser. Mais il faut le faire avec prudence, dira le sens commun. Qu'entend-on par là? Il faut veiller à ne pas courir un trop grand risque de se tromper, en exigeant un nombre suffisant de confirmations de l'affirmation qu'on porte.

C'est cette intuition de base que les statisticiens ont développée au milieu du XX^e siècle, en tenant compte des risques d'erreur dans toutes les statistiques inférentielles.

Principe de l'inférence statistique

Un théorème fondamental dans ce domaine est celui de la variance de la moyenne des échantillons tirés aléatoirement d'une population: cette variance est inversement proportionnelle à la taille des échantillons.

On sait donc comment se distribuent les valeurs moyennes observées autour de la moyenne théorique de la population. On peut alors déterminer une valeur limite qui ne sera dépassée, par exemple, que dans cinq cas sur cent.

Si l'on observe un échantillon dont la moyenne se situe au-delà de cette limite, il y aura peu de chances (moins de 5%) qu'il provienne de la population étudiée. On se trompera donc dans moins de 5% des cas en décidant qu'il est d'une autre origine.

Application: contrôle statistique et études de généralisabilité

On arrive ainsi, non pas à porter des jugements certains, mais à connaître ses risques d'erreur et à pouvoir les contrôler en partie.

On peut en effet réduire les risques en augmentant la taille des échantillons. Mais on augmente alors les coûts... Un compromis est nécessaire, à déterminer dans chaque situation.

Dans les sciences de l'éducation, la logique du contrôle statistique des risques est appliquée de façon très générale. Nous allons voir que la théorie de la généralisabilité en est directement héritière, parce qu'elle vise essentiellement à contrôler l'importance de nos erreurs de mesure, en déterminant quel nombre d'observations est nécessaire dans chaque cas pour réduire notre marge d'erreur à un niveau acceptable.

Tradition et nouveauté

L'ouvrage dans lequel s'insère ce texte vise à faire le point sur un certain nombre de domaines, dont celui de la théorie de la généralisabilité. Ce genre de bilan nécessite de prendre en compte l'ensemble des informations à disposition. On ne pourra donc



éviter de répéter certains aspects de cette théorie qui sont bien connus des spécialistes. Nous verrons notamment d'abord comment elle a permis à l'origine de passer des échelles «relatives» de la psychologie aux échelles dites «absolues» du domaine pédagogique, et nous montrerons les problèmes relatifs aux examens qu'elle a permis de mieux mettre en lumière.

Mais nous verrons aussi que la théorie de la généralisabilité ouvre aujourd'hui toute une série de perspectives nouvelles pour la mise en évidence des sources de variance indésirables et la réduction de ces erreurs. Surtout, elle paraît introduire une voie d'approche nouvelle (parce qu'individualisée) pour mesurer l'apprentissage de façon plus assurée.

En psychologie

Echelles relatives

C'est sans doute Galton qui a trouvé, à la fin du XIX^e siècle, le moyen de quantifier les «facultés» humaines, en comparant les performances des individus à la distribution des résultats d'une large population de référence. Un score était jugé bon, moyen ou faible d'après sa position relative par rapport à l'ensemble des autres scores observés.

Il devenait indifférent, dans ces conditions, d'utiliser telle ou telle question, ou de placer les sujets dans telle ou telle situation, pourvu que les conditions de l'examen soient maintenues constantes pour tous. On pouvait utiliser n'importe quel genre d'épreuves, mais il fallait veiller à ce que leur passation et leur cotation soient strictement standardisées.

Au départ, fidélité du test

Contrôler la qualité de la mesure impliquait dans ce cas de vérifier si le rang donné au sujet après un premier examen demeurerait stable lorsqu'on le soumettait à un second examen. Cette mesure relative avait l'avantage de permettre l'utilisation d'instruments différents. Simplement, le rang que le sujet obtenait au second «test» devait être proche du rang qu'il avait obtenu au premier pour que le psychologue puisse prétendre mesurer la même «faculté».

Des méthodes statistiques ont très vite été utilisées pour ces contrôles, en particulier le coefficient de corrélation, calculé entre mesures différentes. On obtenait ainsi un «coefficient de fidélité», censé d'abord vérifier la qualité métrique du test, envisagée comme une propriété qui lui serait intrinsèque.



Mais on s'est vite aperçu que, pour chaque méthode de comparaison, le coefficient de fidélité prenait une autre valeur. On s'est rendu compte alors que l'on mesurait en fait des sources d'erreur indépendantes. Le coefficient «test-retest» mesurait l'effet des occasions. Le coefficient entre épreuves parallèles mesurait l'effet des problèmes (ou items) utilisés. Le coefficient «interjuges» mesurait la source d'erreur liée aux cotateurs, etc.

Ensuite, généralisabilité à diverses populations de référence

C'est pour unifier ce domaine et rassembler le traitement de toutes ces sources d'erreurs que Cronbach et ses collaborateurs ont développé la théorie de la généralisabilité (Cronbach, Gleser, Nanda & Rajaratnam, 1972).

On disait avant eux que la valeur métrique du test dépendait de la corrélation entre le résultat observé et la note vraie, mais la définition qu'on donnait à cette note vraie était des plus vagues. Cronbach au contraire, dès les années 50, la considère comme la moyenne de tous les résultats possibles pour ce sujet, lorsqu'on tire au hasard les conditions d'observation dans la population de toutes les conditions admissibles. Minimale en apparence, ce changement a des conséquences majeures sur le plan théorique.

D'abord, on ramène ainsi les formules de la psychométrie dans le giron de la théorie statistique, plus particulièrement dans celui de l'analyse de la variance. Les erreurs de mesure ne sont plus que des fluctuations d'échantillonnage, celles que les techniques statistiques ont appris à contrôler. Du coup, toute la méthodologie des plans expérimentaux peut venir enrichir nos méthodes de mesure.

Ensuite, on déduit de cette définition que la fidélité d'un test n'est pas une propriété de l'instrument lui-même, mais dépend de la population parente à laquelle on veut généraliser les résultats observés. Si la population à laquelle on se réfère est homogène, les fluctuations d'échantillonnage auront moins d'ampleur que si cette population est hétérogène. On comprend alors qu'une multitude de coefficients de fidélité puissent coexister pour un même test, et que ce soit à l'utilisateur de préciser la nature de l'inférence qu'il souhaite effectuer. Il peut inférer avec sécurité du résultat à un test à la population de résultats à un autre test parallèle. Il risque des erreurs bien plus graves s'il veut tirer du premier test des conclusions concernant, par exemple, les aptitudes de l'individu dans l'ensemble des situations de la vie courante.

Application: contrôle de la variance d'échantillonnage

Enfin, puisque ce sont les sources de variance apportées par les diverses facettes des conditions d'observations qui sont à l'origine des fluctuations d'échantillonnage, et puisque les facteurs indépendants additionnent simplement leurs effets, les théo-



rèmes de l'analyse de la variance suffisent pour estimer la fidélité de n'importe quel dispositif de mesure. Il devient même possible d'optimiser la mesure, en prenant un plus grand nombre d'observations pour les facettes les plus perturbatrices.

Il suffit de connaître les variances des facteurs que l'on échantillonne et de les additionner, puisque leurs effets s'ajoutent. Par exemple, on peut combiner l'effet de l'échantillonnage des questions et celui de l'échantillonnage des occasions, ou de n'importe quel autre facteur perturbant les mesures dans l'univers de conditions considéré, pour calculer la fidélité globale qui en résulte. De même, on peut estimer d'avance la fidélité de n'importe quel autre dispositif que l'on obtiendrait en modifiant l'échantillonnage des facettes prises en compte dans l'étude. Mieux encore, ce modèle libère les psychologues de la nécessité de construire des tests strictement parallèles, et même il permet de traiter des échelles de mesure «absolues», telles qu'on en utilise en éducation.

En éducation

Echelles absolues

Le fait qu'un élève soit meilleur ou moins bon que ses camarades ne nous dit pas quel est son niveau réel de connaissance. Or c'est ce niveau qui compte dans la pratique. En effet, bien qu'au pays des aveugles, les borgnes soient rois, les exigences de la vie réelle ne peuvent se satisfaire de ce relativisme. Le mécanicien doit savoir réparer les voitures, le chirurgien doit savoir opérer, le pilote n'a pas non plus droit à l'erreur. Même les connaissances en orthographe sont, pour une future secrétaire, définies sur une échelle quasi-absolue, indépendamment du niveau moyen de ses condisciples: on n'admettra pas plus de X fautes par page, par exemple.

Généralisabilité de la mesure «absolue»

Si les échelles de mesure relatives inventées par Galton ont l'avantage de la flexibilité, en pouvant s'appliquer à des situations quelconques, pourvu qu'elles soient standardisées, les tests qui s'en inspirent ne répondent pourtant pas bien aux exigences des éducateurs chargés de faire passer des examens. Les mesures relatives sont en effet plus appropriées dans les situations de concours, où l'on veut simplement classer les candidats par ordre de mérite. Les examens, par contre, sont censés contrôler un niveau de connaissances par rapport à un seuil fixé, déterminé par le programme, et non par la distribution des résultats observés.

La théorie de la généralisabilité rend bien compte de cette différence. Dans un concours, le facteur de difficulté des questions n'intervient pas dans le classement des candidats, puisqu'ils sont tous confrontés aux mêmes exigences (un item réussi



ou raté par tous ne modifie pas les positions relatives des individus les uns par rapport aux autres.) Le classement n'est affecté que si certaines questions font davantage problème pour les uns, alors que d'autres questions perturbent davantage les autres. En termes d'analyse de la variance, c'est l'interaction *Candidats X Questions* qui est la seule source d'erreur.

Dans un examen, au contraire, il importe beaucoup de contrôler la difficulté moyenne des questions, car on peut faire varier à volonté le nombre de réussites ou d'échecs selon le niveau d'exigence des questions que l'on pose. L'interaction *Candidats X Questions* existe encore comme source d'erreurs, mais la variabilité introduite par le choix de questions s'y ajoute comme source de perturbation majeure. C'est elle qui fait varier le niveau effectif d'exigence, alors que l'on voudrait que ce niveau soit déterminé de façon «absolue» (par exemple comme la maîtrise de 60% de l'ensemble de toutes les questions du programme).

Difficulté de contrôler l'échantillonnage des questions

Ce problème est fondamental en éducation. Certes, les enseignants sont bien conscients de la part d'arbitraire qui subsiste, malgré leur souci d'objectivité, dans leur appréciation du niveau de performance exigible, mais ils ne soupçonnent généralement pas la gravité de la situation. Ils pensent d'abord aux incertitudes que comporte la cotation des travaux qu'ils rendent à leurs élèves. Ils se doutent en effet que le même travail serait noté différemment par un collègue, voire par eux-mêmes quelques mois plus tard. Mais aussi pertinentes que soient ces craintes, le vrai problème n'est pas là. Il est dans le choix du contenu et de la forme des questions qu'ils posent.

Les programmes notionnels, tels qu'ils sont rédigés généralement, ne précisent pas comment les élèves sont censés manifester leurs connaissances. Les enseignants peuvent aussi bien leur demander de réciter un cours que de résoudre des exercices d'application. Or la difficulté peut en être totalement différente, surtout si les élèves n'y ont pas été préparés spécialement.

Même pour des questions à choix multiples, qui sembleraient garantir une mesure objective en standardisant les conditions de réponse, il reste un grand nombre de sources d'erreurs. Les enseignants qui préparent de telles questions ne peuvent pas prévoir les difficultés qu'elles présenteront pour leurs élèves, tant il y a de sources d'influence qui interviennent: longueur ou complexité variable des phrases, habillage du problème avec des situations plus ou moins familières aux élèves, présence d'indices verbaux ou graphiques facilitateurs ou trompeurs, nombre d'éléments à prendre en compte, nombre d'étapes dans la recherche de la solution, etc.

Sans doute le recours à des questions tirées de banques d'items étalonnés apporte-t-il un début de réponse à cette difficulté, mais ce n'est pourtant pas la solution miracle: selon le style d'enseignement qu'ils auront reçu, les élèves se verront péna-



lisés ou favorisés par le choix de telle forme de questionnement plutôt que de telle autre. De leur côté, les questions sélectionnées suivant les procédures de la *Théorie de Réponse aux Items* risquent de demander une part importante de raisonnement, ce qui n'est pas nécessairement le trait que l'on cherche à mesurer, et ce qui pénalisera forcément certains élèves.

Les sources de difficultés, qu'elles affectent tous les élèves ou certains plus que d'autres, sont donc innombrables, et ainsi il n'est pas possible de contrôler a priori le niveau de difficulté d'une question.

Résultat: mise en doute des examens

Il découle de la discussion précédente qu'il n'est guère possible de contrôler le niveau d'exigence d'un examen.

Lorsqu'une banque d'items a été constituée, composée de façon structurée, de façon à assurer à tous les points du programme une représentation équilibrée, et à couvrir les différents niveaux d'objectifs éducatifs, la théorie de la généralisabilité permet de prédire la marge d'erreur à attendre, pour des échantillons de N questions choisies au hasard dans cet univers de toutes les questions possibles. Si l'on inverse la formule, on peut déterminer N , la taille qu'il faut donner à l'échantillon, pour limiter la marge d'erreur à ce que l'on peut accepter pour l'estimation de la moyenne d'un candidat.

Malheureusement, la variabilité des difficultés des questions est telle qu'il faudrait un examen excessivement long pour compenser les fluctuations aléatoires et stabiliser le niveau de difficulté moyen des différents échantillons que l'on tirerait ainsi au hasard. Cette source d'erreur majeure, s'ajoutant (puisque les variances d'erreur s'additionnent) à toutes celles que les travaux de docimologie ont mis en évidence, il est souvent difficile dans la pratique de mesurer le niveau de connaissance des élèves sur une échelle absolue.

On objectera sans doute que les examens fonctionnent actuellement, sans poser trop de problèmes apparemment. Cela est vrai, parce que les enseignants introduisent spontanément des correctifs, pour faire fonctionner le système tel qu'il est, sans trop s'inquiéter de sa cohérence.

Tout d'abord, ils se sont créé un sous-ensemble de problèmes-types, ce qui restreint considérablement l'univers des questions admissibles sur la base des programmes, et par conséquent la variabilité des niveaux de difficulté. Dans cet univers homogénéisé, les échantillons de questions qu'ils choisissent ne peuvent plus guère atteindre des niveaux de difficulté extrêmes. Mais on en voit la conséquence: exercices d'apprentissage et travaux de contrôle fonctionnent en boucle fermée. L'école se crée son milieu propre et récompense les élèves qui s'y adaptent, mais elle reste loin de pré-



parer à la diversité des problèmes de la vie réelle. Les transferts indispensables restent à faire après la fin de l'enseignement.

Ensuite, les jurys d'examen se réservent le droit d'adapter les critères de correction ou les barèmes pour que les résultats de l'examen restent dans les normes habituelles. L'examen est ainsi plus ou moins transformé en concours, non par une volonté sélective consciente, mais pour égaliser les chances de réussite d'une année à l'autre, en corrigeant les fluctuations aléatoires gênantes.

Egalité ni possible ni souhaitable

L'objectivité dans les examens est donc un mythe, si l'on entend par là un idéal inatteignable. Notons toutefois qu'il n'y a pas lieu de s'indigner si la réalité scolaire ne correspond pas aux modèles pédagogiques et statistiques que nous voudrions lui imposer. Le fonctionnement quotidien des écoles doit répondre à d'autres logiques aussi que celle de l'exactitude de la mesure.

Les enseignants ont bien raison, par exemple, de tenir compte du niveau effectif de leurs élèves plutôt que d'appliquer mécaniquement le programme officiel. Ils doivent en conséquence évaluer leurs élèves par rapport à ce qu'ils ont enseigné, plutôt que par rapport à des exigences définies dans l'abstrait pour l'ensemble de la population scolaire.

Ainsi, des études statistiques ont montré que certaines classes ont deux ans d'avance sur d'autres. Insister sur l'égalité absolue des exigences pourrait, dans ces conditions, conduire à des catastrophes, en occultant la diversité réelle et inévitable des milieux sociaux et de leurs cultures.

Mais lucidité et imagination nécessaires

Tenir compte de la diversité des valeurs qui sont à respecter à la fois, sans vouloir absolutiser telle ou telle conception de l'équité, n'empêche pas d'être lucide et de chercher à connaître la réalité et ses problèmes.

Par exemple, mettre au point des épreuves permettant de suivre le niveau d'apprentissage des élèves dans les diverses disciplines, selon les classes, les établissements, les types d'enseignement, ou bien participer à des enquêtes internationales comparant les systèmes scolaires et suivant l'évolution de leurs performances d'année en année, voilà des domaines d'application immédiats pour le modèle statistique de la généralisabilité. Ce dernier aide ainsi à prendre conscience du fonctionnement du système éducatif.



Mais d'une façon plus spécifique, cette théorie nous fournit surtout des points de repère pour connaître l'importance des fluctuations d'échantillonnage affectant les diverses mesures de l'apprentissage. Dans ce domaine, elle peut très légitimement localiser des problèmes, poser des questions, susciter des réflexions et suggérer des approches nouvelles.

La critique des examens à laquelle elle conduit n'est d'ailleurs pas uniquement négative. Elle a amené à se demander si l'approche psychométrique héritée de Galton ne faisait pas fausse route en comparant les élèves les uns aux autres, donc en suscitant la compétition entre eux et en élargissant leurs différences, avec les conséquences sociales que l'on connaît: marginalisation, etc.

Elle montre maintenant, comme on vient de le voir, les limites d'une évaluation par objectifs, telle qu'on l'envisageait dans les années 60, visant à mesurer les connaissances sur une échelle «absolue» de compétence.

Par opposition, la réflexion suscitée par cette critique nous conduit à revenir à la conception pédagogique proposée par Benjamin Bloom, d'une éducation individualisée, sur le modèle du tutorat, où l'élève n'est comparé qu'à lui-même, l'évaluation scolaire étant centrée alors sur ses difficultés et ses progrès. Le but de l'éducation n'est-il pas de développer au mieux chaque personnalité?

La théorie de la généralisabilité devrait pouvoir fournir les instruments statistiques pour aider à réaliser cet ambitieux programme.

Que proposer?

Pour illustrer les applications de la théorie de la généralisabilité possibles à l'avenir, nous avons choisi trois domaines particuliers: l'analyse de l'importance relative des sources d'influences, le contrôle des facteurs indésirables et l'individualisation de la mesure. Leur discussion sera nécessairement plus technique. Nous essayerons cependant d'éviter les formules.

Analyser l'importance relative des sources d'influences

Estimation des composantes de variance dans des plans mixtes

Cronbach a toujours insisté sur le fait que l'estimation des composantes de variance pour les facteurs principaux et leurs interactions (les influences positives ou négatives des conditions d'examen, qui rendent compte de la variabilité des résultats de l'individu observé) était un travail scientifique essentiel, sans lequel il était impossible d'améliorer les dispositifs de mesure. Les études de généralisabilité ont pour but d'obtenir de telles estimations.



Selon les plans expérimentaux choisis pour recueillir les mesures, certaines composantes peuvent être estimées indépendamment, ou sont au contraire confondues avec d'autres. Elles peuvent également être estimées de façon plus ou moins stable, selon le nombre de degrés de liberté du carré moyen correspondant. C'est évidemment le rôle du statisticien de planifier des études de généralisabilité aussi informatives que possible, pour permettre d'analyser de façon précise la variance des scores observés, comme le demande Cronbach.

Mais les statisticiens anglo-saxons s'en sont tenus jusqu'ici à des plans de mesure classiques, où une population d'élèves supposée infinie était confrontée à des examens comportant aussi une infinité de conditions d'observations, (dont certaines facettes pouvaient néanmoins être fixées ou aléatoires finies). Ils n'ont jamais traité de dispositifs où la population d'élèves était considérée comme fixée, c'est-à-dire où l'on s'intéressait aux fluctuations d'échantillonnage des conditions d'observation pour ces élèves uniquement. Le problème paraît pourtant légitimement se poser et les estimations de variance sont alors différentes.

Cette hésitation s'explique sans doute par la complexité que les plans mixtes introduisent dans les formules de généralisabilité si l'on veut, comme Robert Brennan par exemple, dériver les formules d'erreurs pour chaque plan de mesure séparément (Brennan, 1992).

Mais l'algorithme de calcul proposé par Cardinet et Tourneur (Cardinet & Tourneur, 1985) fournit des estimations utilisables, si du moins on accepte de considérer les coefficients de généralisabilité comme de simples coefficients d'importance d'effet, comme proposé ci-dessous.

Coefficients d'importance d'effet

La pratique est devenue courante de calculer un test de signification statistique chaque fois que l'on veut prouver l'existence d'une source de variation quelconque. Comme chaque fois qu'une démarche est appliquée mécaniquement, cela peut conduire à des abus. Les chercheurs ont eu en effet tendance à confondre signification statistique et grandeur de l'influence étudiée. Or une différence de moyenne peut avoir très peu de chances d'être l'effet du hasard, sans être nécessairement grande. Il suffit que la taille de l'échantillon soit très importante. D'où le développement en cette fin de siècle, pour contrebalancer et nuancer les tests de signification, de coefficients d'importance d'effet, qui ne sont pas affectés par la taille des échantillons utilisés.

Le principe de ces coefficients est simple. On calcule la composante de variance pour l'effet qu'on veut étudier (*Variance systématique*), et on la compare à la variance d'échantillonnage du dispositif utilisé, en calculant le rapport: *Variance systématique / (Variance systématique + Variance d'erreur)*. Un tel rapport peut varier de 0 à 1 et



donne ainsi une idée de l'importance de l'effet étudié relativement à celle des effets d'échantillonnage non contrôlés.

On peut calculer ce rapport non seulement pour chaque source de variance analysée séparément dans l'étude de généralisabilité, mais encore, de façon très libre, pour toutes les combinaisons de variances systématiques qui peuvent avoir un sens pour l'expérimentateur (Abdi, 1987).

On peut montrer que ce sont en fait des coefficients d'importance qui sont fournis par l'algorithme de Cardinet et Tourneur et les logiciels qui l'appliquent, plutôt que des coefficients de généralisabilité au sens où Robert Brennan, par exemple, les définit, en analysant la variance de l'erreur relative dans chaque dispositif (Brennan, 1992). On peut noter en particulier que ce dernier ne corrige pas les estimations de variance pour les facettes finies (*Correction de Whimbey*) alors que les coefficients d'importance (et les logiciels développés à partir de «Assurer la Mesure») le font régulièrement (pour le calcul de *Oméga carré*).

Cette réinterprétation du coefficient de généralisabilité comme un type particulier de coefficient d'importance constitue une nouvelle étape dans la réintégration de la psychométrie au sein de la théorie statistique générale. Elle facilite grandement la mesure de l'influence relative des différentes sources d'erreur, base elle-même de toute intervention visant à optimiser les dispositifs de mesure.

Contrôler les facteurs indésirables

Si l'on voit que l'échantillonnage d'une facette de l'univers des conditions d'observation introduit une grande variance d'erreur, la démarche statistique habituelle pour en réduire l'importance est de prendre davantage d'observations sur cette facette. La variance correspondante sera réduite en proportion.

Mais cette solution simple atteint vite ses limites, en raison du coût des prises d'observations. C'est pourquoi d'autres approches ont aussi été essayées.

Problèmes d'hétérogénéité: analyse de facettes

Il peut arriver que la variance d'erreur soit grande en raison d'une forte hétérogénéité des niveaux de la facette en question.

Si par exemple une question, dans un test de mathématiques, traite d'un sujet très éloigné du thème principal, on peut s'attendre à ce que les réussites à cette question ne soient pas liées à la réussite moyenne à l'ensemble du test. Du coup la différence sera plus grande pour cette question que pour les autres, entre le score effectif et le score qui peut être prévu sur la base des moyennes marginales (de la question et de



l'élève). Autrement dit, le composant de score lié à l'interaction sera plus élevé. Comme le phénomène se répétera pour tous les sujets, la part de la somme des carrés d'interaction apportée par cette question sera particulièrement grande.

Or cette part peut être calculée facilement. C'est ce que Cardinet et Tourneur ont appelé l'analyse de facettes. Les logiciels Etudgen et WinGT font apparaître ainsi la proportion de la variance d'erreur (relative ou absolue) qui est causée par chaque question.

Sur cette base, il est possible d'examiner ce qui peut causer l'hétérogénéité de la question et de redéfinir en conséquence l'univers des conditions d'observation auquel on veut généraliser (pour pouvoir éliminer cette source de variabilité non souhaitée). On peut ainsi déterminer un univers de référence plus étroit, mais beaucoup plus homogène.

Dans bien des cas, le responsable d'un examen aura intérêt à limiter un peu la portée de la certification, autrement dit le domaine couvert par le diplôme, en échange d'une plus grande sûreté dans son jugement.

Problèmes de pondération: échantillonnage matriciel

Dans d'autres cas, le contrôle des facteurs indésirables peut se faire en remplaçant un échantillonnage purement aléatoire par un échantillonnage où les facettes qui sont sources principales d'erreur sont examinées plus souvent que les autres. On en donnera simplement un exemple.

On a dit plus haut que grâce à la théorie de la généralisabilité, les plans expérimentaux avaient pu être introduits en psychométrie. L'échantillonnage matriciel est un type particulier de plan expérimental qui répartit la prise d'information sur plusieurs échantillons de sujets et plusieurs échantillons de questions.

Ceci est important dans les enquêtes portant sur l'ensemble d'un système scolaire. On veut couvrir le plus largement possible tous les chapitres du programme. On veut aussi différencier le mieux possible certains sous-ensembles d'élèves pour pouvoir effectuer des comparaisons. Dans ces conditions, choisir un seul échantillon de questions et un seul échantillon d'élèves serait très peu efficace. Il faut au contraire que certains élèves répondent à certaines questions, tandis que d'autres élèves traitent d'autres questions, de façon à multiplier les prises d'information sur ces deux plans, sans surcharger personne.

Si l'on sait que la variabilité des résultats se situe surtout entre les classes, par exemple, on pourra augmenter le nombre de classes échantillonnées en réduisant en contrepartie le nombre d'élèves interrogés dans chaque classe. De la même façon, si l'on veut réduire l'incertitude en ce qui concerne le niveau atteint dans certains chapitres du programme, on multipliera le nombre de questions qui s'y rapportent.



L'apport principal de la théorie de la généralisabilité dans des dispositifs de ce genre est de rappeler que l'erreur de mesure n'est à chercher ni dans les questions, ni dans les sujets, mais uniquement dans les fluctuations d'échantillonnage. Ainsi la même composante de variance, par exemple interquestions, peut faire partie de la variance d'erreur si l'on étudie les élèves et de la variance vraie si l'on étudie le programme.

En particulier, individualiser la mesure

Pourtant dans le passé, on avait voulu faire de la variance entre élèves la variance vraie, celle qui portait la fidélité et la validité de l'examen, celle qu'il fallait donc maximiser pour améliorer la mesure. De même, et par opposition, la pédagogie de maîtrise qui voulait amener tous les élèves à l'atteinte de l'objectif cherchait à maximiser la différence entre les questions non encore étudiées et les questions déjà étudiées, où la réussite devait être maximale.

Sans rejeter complètement ces deux modèles, on peut leur préférer un troisième, où l'on compare les élèves à eux-mêmes, soit dans le temps, en évaluant leur progrès, soit dans leurs performances relatives, en analysant leur profil de résultats.

Mesure du changement individuel

Au départ, on peut s'attendre à des résultats faibles, alors qu'après apprentissage, on devrait pouvoir observer un progrès marqué, sur les mêmes questions, ou pour des questions parallèles.

Divers dispositifs sont envisageables pour mettre en évidence ce progrès. Le plus simple est certainement celui qui répète la même épreuve, avant et après apprentissage. On compare alors les pré- et post-tests, aux questions appareillées. S'il est optimal du point de vue statistique, ce schéma risque pourtant de rebuter les élèves qui ne comprendront pas qu'on les interroge sur un sujet qu'ils n'ont pas appris. De plus, la mémorisation des questions, inévitable, peut avoir des effets perturbateurs sur la valeur de la mesure.

Une approche plus prometteuse est celle qui analyse au départ une production complexe, en la cotant sur un ensemble d'une dizaine d'objectifs, faisant tous l'objet d'un apprentissage en cours d'année. En fin d'année, une autre production complexe peut être cotée sur les mêmes objectifs. Le progrès moyen peut alors être calculé. La variance intermoments peut être comparée à la variance d'échantillonnage, représentée dans ce cas par l'interaction *Objectifs X Moments*. Un coefficient d'importance de l'apprentissage peut être estimé sur cette base, ceci élève par élève et sans aucune comparaison interindividuelle. La faiblesse de ce plan est cependant d'admettre que les objectifs choisis sont un échantillon des objectifs possibles et pourraient être remplacés par d'autres, ce qui n'est pas toujours réaliste.



Un autre dispositif envisageable pour la mesure du progrès serait de coter chacun des objectifs du plan précédent par rapport à un certain nombre de critères, nichés dans chaque objectif et choisis aléatoirement à chacun des deux moments (avant et après étude). On aurait ainsi une meilleure estimation de la variance d'erreur expérimentale, donnée par la variance entre critères à l'intérieur à la fois des objectifs et des moments. On pourrait comme précédemment calculer un coefficient d'importance de l'apprentissage (coefficient de généralisabilité du progrès) et si l'on veut, fixer des seuils de progrès individualisés. Pour des élèves aux performances très homogènes, un progrès faible pourrait en effet être déjà significatif et certifiable, alors que pour des sujets plus brouillons, il faudrait des progrès plus marqués pour qu'on puisse les prendre en compte.

Une dernière approche, moins efficace du point de vue statistique mais peut-être plus facile à mettre en pratique, serait d'avoir un pré-test et un post-test non appareillés, les items étant choisis aléatoirement aux deux moments.

L'essentiel dans tous ces cas est que l'on réfère l'élève uniquement à lui-même. Les exigences effectives sont différentes, puisqu'elles sont fonction de la variabilité de chaque sujet. Le souci d'équité prend donc une forme nouvelle, elle aussi individualisée.

Mesure d'une différence ou d'une interaction

Le principe d'une mesure individualisée peut prendre une autre forme, dans laquelle le sujet est comparé à lui-même, mais cette fois au même moment.

La psychométrie connaissait les mesures ipsatives, pour lesquelles le score était calculé en faisant la différence entre deux résultats individuels. Il s'agissait le plus souvent de soustraire la moyenne d'un profil, pour que tous les sujets aient la même valeur moyenne. Avec la théorie de la généralisabilité, il devient possible de connaître la marge d'erreur que comporte la comparaison de deux points d'un profil, et même de connaître la généralisabilité d'un effet d'interaction.

Pour comparer les résultats à deux des dimensions d'un profil, il faut connaître la variance d'échantillonnage pour chacun, puis additionner ces deux variances d'erreur. On pourra ensuite établir un intervalle de confiance pour la différence, à partir de cette variance combinée. Cette façon de faire suppose que chaque dimension est mesurée par un même nombre d'items, et que la variance interitems à l'intérieur des dimensions est prise comme variance d'échantillonnage.

Si c'est une interaction qui est étudiée, les deux facettes qu'elle combine font nécessairement partie de la face de différenciation. Cette interaction est à son tour en interaction avec au moins une autre facette (de généralisation). C'est cette dernière com-



posante d'interaction (de trois facteurs) qui fournit, après pondération appropriée, le terme d'erreur pour le coefficient de généralisabilité de l'effet d'interaction.

La rôle de la théorie de la généralisabilité

Extension de l'analyse de la variance

On a beaucoup discuté des relations entre la théorie de la généralisabilité et l'analyse de la variance. Puisque ces pages doivent faire un bilan, disons avec prudence que ces relations sont «bonnes»!

D'un certain point de vue, la théorie de la généralisabilité peut partir, comme le fait Robert Brennan, de la définition des composants de scores et se développer presque en vase clos jusqu'au calcul de l'erreur-type sur la mesure. Le fait qu'elle se réfère exclusivement au modèle d'ANOVA de Cornfield et Tukey tend aussi à l'isoler et à rendre difficile pour chaque modèle l'utilisation des résultats de la théorie «d'en face».

Mais d'un autre point de vue, selon Cardinet et Tourneur, l'étude de généralisabilité commence là où finit l'analyse de la variance. Partant de là, on peut considérer qu'une analyse de généralisabilité ajoute à l'ANOVA le souci de la mesure, qui implique le choix des facettes de différenciation et des facettes d'instrumentation. Au-delà encore, il faut citer le calcul de divers coefficients et indices dont l'usage semble spécifique à ce type d'analyse. Mais on peut dire que la théorie de la généralisabilité constitue surtout une extension originale de l'ANOVA, tout en reposant essentiellement sur la même base.

La possibilité évoquée plus haut de réinterpréter le coefficient de généralisabilité comme une forme particulière de coefficient d'importance contribue à rapprocher encore plus les deux modèles, en les unifiant par le haut cette fois, c'est-à-dire au niveau de leurs applications et de leurs conclusions.

Outil pour des recherches de toute nature

Une autre évolution importante à noter à l'heure du bilan, c'est que le modèle psychométrique dont partait Cronbach à l'origine est devenu un modèle quasi-universel. A la suite de l'extension proposée par les spécialistes européens (Cardinet, Tourneur & Allal, 1981-1982) l'idée de la symétrie du modèle a été généralement bien acceptée. Son application n'est donc plus limitée à la mesure des traits individuels. Ses conclusions peuvent être prises en compte dans l'estimation des marges d'erreur des enquêtes et des *surveys*, et même pour l'interprétation des résultats des recherches expérimentales.



Références bibliographiques

- Abdi, H. (1987). *Introduction au traitement statistique des données expérimentales*. Grenoble: Presses Universitaires de Grenoble.
- Brennan, R. L. (1992). *Elements of generalizability theory* (2nd ed.). Iowa City: ACT Publications.
- Cardinet, J. & Tourneur, Y. (1985). *Assurer la mesure*. Bern: Peter Lang.
- Cardinet, J., Tourneur, Y. & Allal, L. (1981-1982). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement*, 18 & 19 (4), 183-204 et 331-332.
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.



Zur Bedeutung international vergleichender Leistungsmessungen im Bildungswesen

Armin Gretler

Im letzten Jahrzehnt des 20. Jahrhunderts sind international vergleichende, empirische Untersuchungen zweifellos zu einem Schwerpunkt der Bildungsforschung geworden. Dies gilt insbesondere für Messungen schulischer Leistungen in verschiedenen Fächern und auf verschiedenen Altersstufen. Im folgenden Beitrag, der sich bemüht, auf einer konkreten Ebene zu bleiben, wird zuerst den Gründen für diese Entwicklung und damit der Geschichte dieser Untersuchungen nachgegangen. Anschliessend wird anhand einiger schweizerischer und eines amerikanischen Beispiels gezeigt, welche Folgerungen die Forschung aus den Ergebnissen dieser Untersuchungen für das jeweils eigene Bildungssystem ziehen kann. Daraus werden Möglichkeiten und Grenzen dieser Art von Bildungsforschung sichtbar. Eine kurze Betrachtung gilt der auf verschiedenen Ebenen angesiedelten Kritik international vergleichender Leistungsmessungen. Das Fazit ist klar: solche Untersuchungen sind ein unverzichtbares Instrument der Weiterentwicklung unserer Bildungssysteme; sie teilen diese Funktion aber mit anderen Forschungstypen. Abschliessend wird eine Vermutung darüber angestellt, welche bisher relativ vernachlässigten Forschungsaspekte in diesem Bereich in Zukunft eine grössere Rolle spielen werden.

Geschichte international vergleichender Leistungsmessungen

Ein Blick zurück legt nahe, in der bisherigen Entwicklung international vergleichender Leistungsmessungen zwei Phasen zu unterscheiden: eine etwa die Jahre 1960-1995 umfassende Pionierphase, die organisatorisch auf nicht-gouvernementaler Ebene angesiedelt und weitgehend von der *International Association for the Evaluation of Educational Achievement* (IEA) geprägt war, und eine etwa 1995 einsetzende Phase von Regularisierung und Institutionalisierung auf inter-gouvernementaler Ebene, in welcher der Stafettenstab von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) übernommen wurde. Über eine allfällige dritte, der Zukunft zugehörige Phase dieser Entwicklung, nämlich die Routinisierung internationaler Leistungsmessungen nach dem Ablauf des ersten OECD-Zyklus im Jahre 2010, kann im jetzigen Zeitpunkt nur spekuliert werden. An dieser Stelle muss aber

auch daran erinnert werden, dass vergleichende Erziehungswissenschaft eine hinter die Leistungsmessungen zurückreichende Tradition hat; sie war aber vorher nicht empirisch, sondern deskriptiv orientiert.

Die Pionierphase

Ende der fünfziger Jahre kamen im Hamburger Unesco-Institut etwa ein Dutzend Bildungsforscher zusammen, die sich mit Fragen befassten, welche innerhalb eines einzigen Bildungssystems nur mit grossangelegten, aber völlig unrealistischen vergleichenden Schulversuchen zu beantworten gewesen wären. So entstand die Idee, Systemmerkmale wirklich existierender Bildungssysteme und, wenn möglich, deren Wirkung untereinander zu vergleichen: die Welt wurde in diesem Sinne als natürliches «*Educational Laboratory*» aufgefasst (diese Grundidee wurde in der Folge immer wieder zitiert). In einer drei Jahre dauernden Machbarkeitsstudie (1959-1962), an der sich zehn europäische Länder, die USA und Israel beteiligten, wurde überprüft, ob Untersuchungen der geplanten Art methodologisch und datenverarbeitungsmässig überhaupt durchführbar waren. (Die Schweiz beteiligte sich unter der Verantwortung von Samuel Roller an dieser Studie. Abzuklären bleibt, warum unser Land während der ersten 25 Jahre ihrer Existenz der IEA dann doch fernblieb und erst seit 1986 in ihr vertreten ist; diese Funktion wird von Beginn an von der Schweizerischen Koordinationsstelle für Bildungsforschung / SKBF wahrgenommen). Als sich der erfolgreiche Verlauf der Machbarkeitsstudie abzuzeichnen begann, wurde die IEA 1961 offiziell gegründet (zur Geschichte der IEA siehe z. B.: De Landsheere, 1986, S. 231-237; Purves, 1989 und IEA, 1993, S. 57-74).

In ihrem sogenannten «*Mission Statement*» definiert sich die IEA wie folgt:

«The International Association for the Evaluation of Educational Achievement (IEA) is an independent international cooperative of research centres. It has taken as its mission the conduct of comparative studies focusing on educational policies and practices in order to enhance learning within and across systems of education. IEA has committed itself to a cycle of studies of learning in the basic school subjects and to additional studies of particular interest to its members.»

Such mission implies the following specific features which distinguish IEA studies:

- 1. By being international and cooperative they enable researchers and policy makers to enter into a dialogue with and learn from their colleagues around the world and they enable systems of education to view more clearly their unique cultural situation from an international comparative perspective.*
- 2. In focusing on educational policies and practices, they start with developing a conceptual framework that clarifies the issues, suggests appropriate methods of investigation, results in validated measures of educational outcomes and processes, and uses those analytic tools that can best elucidate key factors and issues» (IEA 1993, S. 1).*



Seit ihrer Gründung hat die IEA im wesentlichen folgende Untersuchungen durchgeführt (in Klammern wird jeweils die Zahl der beteiligten Länder beigefügt):

- *First International Study in Mathematics (12)*,
- *Six-Subjects-Survey [Reading (15), Literature (10), Science (19), Civic Education (10), English and French as Foreign Languages (18)]*,
- *Classroom Environment Study (10)*,
- *Second International Study in Mathematics (20) and Science (24)*,
- *Written Composition Study (14)*,
- *Preprimary Childcare Study (14)*,
- *Computers in Education Study (23)*,
- *Reading Literacy Study (31)*,
- *Third International Mathematics and Science Study / TIMSS* (rund 50; eine weitere Anzahl Länder kommt in einer zurzeit noch nicht abgeschlossenen, sogenannten TIMSS-Repeat-Study dazu),
- *Civic Education Study (rund 20)*.

TIMSS ist unter verschiedenen Aspekten der bisherige Kulminationspunkt international vergleichender Leistungsmessung; dies gilt in bezug auf die Zahl der teilnehmenden Länder, den methodologischen Aufwand, den hohen Organisationsgrad und die Raschheit der Datenauswertung, die Professionalität in der Verbreitung der Ergebnisse und das grosse Echo sowohl in der Fachwelt als auch bei den politischen Entscheidungsträgern und der breiten Öffentlichkeit.

Im Zusammenhang mit den obenerwähnten Untersuchungen hat die IEA mit fortlaufenden Verbesserungen und Verfeinerungen ein reichhaltiges methodologisches Instrumentarium und ein Organisationsmodell internationaler Zusammenarbeit entwickelt. Die Struktur eines typischen IEA-Projektes besteht aus:

- einem *International Steering Committee*;
- dem *International Coordinating Centre*;
- dem *International Project Committee* (mit den nationalen Projektkoordinatoren, dem internationalen Koordinator und den Mitgliedern des *International Steering Committee*);
- einem *National Project Coordinator*, einem *National Centre* sowie einem *National Committee* für jedes beteiligte Land.



Der IEA kam während der Pionierphase internationaler Leistungsmessungen die führende Rolle zu. Sie beschränkte sich nicht auf Leistungsmessung allein, sondern bezog auch das Curriculum und in zunehmendem Masse auch andere Kontextmerkmale in ihre Untersuchungen mit ein. Hingegen konnten gewisse organisatorische Schwächen der IEA nicht übersehen werden. Je mehr sich z. B. die politischen Entscheidungsträger für die Ergebnisse der Leistungsmessungen zu interessieren begannen, desto nachteiliger waren gewisse Verzögerungen in der Präsentation von IEA-Untersuchungsergebnissen (ihre Lernfähigkeit in dieser Hinsicht bewies die IEA dann allerdings mit TIMSS). Dass die IEA nicht alles leisten konnte, was auf dem Gebiet internationaler Vergleiche im Bildungswesen zu leisten wäre, kommt in einer Kritik des *U.S. National Research Council's* (nach einer vorausgehenden Würdigung der Pionierrolle der IEA) zum Ausdruck: *«IEA has not identified or attempted to deal with all the important issues that should be addressed by indicators, nor has it been sufficiently well financed or well organized to set up regularly scheduled international assessments in all important school subject matters»* (U.S. National Research Council, 1993, S. 6).

Gegen Ende der achtziger Jahre beginnt sich abzuzeichnen, was schliesslich zur zweiten Phase in der Entwicklung internationaler Leistungsmessungen führen wird. Dabei ist eine doppelte Bewegung zu unterscheiden. Einerseits wuchs der politische Druck im Hinblick auf die Errichtung eines umfassenden Indikatorensystems zur international vergleichenden Beurteilung von Bildungssystemen. Dabei war klar, dass diese Aufgabe am ehesten einer inter-gouvernementalen Organisation (konkret: der OECD) anzuvertrauen war. Andererseits wurden Leistungsmessungen dadurch auf ihre Funktion als Teilmenge eines solchen umfassenden Indikatorensystems zurückgeführt; andere Aspekte des internationalen Vergleichs von Bildungssystemen traten gleichberechtigt neben die in das Gesamtsystem zu integrierenden Leistungsmessungen.

Lange wurde erwartet, dass sich zwischen IEA und OECD eine enge Zusammenarbeit und Arbeitsteilung etablieren würde. Chester E. Finn, ehemaliger Sekretär im *U.S. Education Department* und eine der Haupttriebkkräfte für die Errichtung eines umfassenden Indikatorensystems im Rahmen der OECD, äusserte sich z. B. wie folgt zum Verhältnis zwischen IEA und OECD:

«Today the work of the IEA comprises the single most valuable asset in the world with respect to the cross-national appraisal and analysis of education outcomes... It seems to me that, instead of working in splendid isolation, a more stable and closer relationship between IEA and OECD would be most beneficial. IEA very likely will be the source of much of the cognitive outcomes data the OECD will want to gather, and it certainly would be a waste of resources if OECD would replicate what IEA is already doing. Organizationally, OECD represents governments, while IEA members represent mostly research organizations, universities, foundations, and some governments. OECD is a fairly stable enterprise, financed largely through fixed annual payments from member countries. IEA is rather more fragile, highly dependent on private



philanthropy and volunteer labor. There is, to my mind anyway, an obvious symbiosis of interests, capabilities, and strengths waiting to be developed» (Finn, 1989, S. 114-116).

Die Entwicklung verlief dann anders; die enge Zusammenarbeit zwischen IEA und OECD kam nicht zustande. Neben anderen Faktoren scheinen persönliche Konstellationen und Kommunikationsschwierigkeiten zwischen Hauptakteuren dabei eine Rolle gespielt zu haben. Das entsprechende historische Unterkapitelchen bleibt zu schreiben.

Die Regularisierungs- und Institutionalisierungsphase

Im Gefolge der Diskussionen der fünfziger und sechziger Jahre über die Möglichkeiten der Sozialwissenschaften, Beiträge zur rationalen Planung der modernen Industriegesellschaften und damit auch ihrer Bildungssysteme zu leisten, veröffentlichte die OECD 1973 ein Dokument mit dem Titel «*Un système d'indicateurs de l'enseignement visant à orienter les décisions des pouvoirs publics*» (OECD 1973). Das von grossen Hoffnungen begleitete Projekt musste aber aus verschiedenen Gründen, vor allem wegen mangelnden Interesses der politischen Entscheidungsträger, aufgegeben werden; die Zeit dafür war offensichtlich noch nicht reif (siehe dazu: Bottani & Tuijnman 1994, S. 25-26). Wie oben dargelegt, erwachte das Interesse an internationalen Bildungsindikatoren – diesmal unter bildungspolitischem Druck vor allem aus den USA – gegen Ende der achtziger Jahre neu; die OECD organisierte 1987 in Washington und 1988 in Poitiers vorbereitende Konferenzen, deren Zweck in der Ausarbeitung und Einführung eines internationalen Bildungsindikatorensystems bestand. Daraus entstand das OECD / CERi (Centre for Educational Research and Innovation) -Projekt für die Entwicklung internationaler Bildungssysteme-Indikatoren (INES: *International Indicators of Educational Systems*). Im Rahmen von INES veröffentlichte die OECD 1992 eine erste Serie von Bildungsindikatoren; seither wird die Reihe in kurzen Abständen periodisch fortgesetzt.

Die folgenden Angaben zum Inhalt der Indikatoren und zur Organisation des INES-Projektes stützen sich auf die vierte Ausgabe der Bildungsindikatoren (OECD 1996). Die insgesamt 43 Indikatoren werden in folgende sieben Kapitel gegliedert (in Klammern wird jeweils die Anzahl Indikatoren angegeben):

1. Demographisches, soziales und wirtschaftliches Umfeld der Bildung (4)
2. Bildungsausgaben sowie Human- und Finanzressourcen (8)
3. Bildungszugang, Bildungsbeteiligung und Bildungsfortschritte (5)
4. Schulumfeld und Bildungsabläufe in der Schule/Klasse (10)
5. Abschlussquoten der Bildungseinrichtungen (4)
6. Leistungen von Schülern sowie Lese- und Schreibfähigkeit von Erwachsenen (8)
7. Arbeitsmarktbezogene Bildungsergebnisse (4)

Aus dieser Kategorisierung wird deutlich, dass die Ergebnisse von Leistungsmessungen in einem von sieben Kapiteln oder in acht von 43 Indikatoren zusammengefasst werden. Sechs weitere Kapitel und 35 weitere Indikatoren betreffen andere Merkmale von Bildungssystemen; der Fortschritt gegenüber der früheren Beschränkung auf Leistungsmessung ist offensichtlich.

Die Vertretung der Schweiz im INES-Projekt wird durch das Bundesamt für Statistik (BFS) und einige weitere Experten wahrgenommen. Im Rahmen des nationalen INES-Nachfolgeprojektes «Bildungsindikatoren Schweiz» gibt das BFS regelmässig Veröffentlichungen heraus, die vor allem schweizerische Kennziffern, aber auch internationale Vergleichsindikatoren enthalten (BFS, 1993, 1995 und 1998).

Beruhen die OECD-Indikatoren im Bereich Bildungsergebnisse bisher auf Daten, die aus IEA – und einigen weiteren Untersuchungen stammten, so ist die OECD in der zweiten Hälfte der neunziger Jahre daran, im Bereich Leistungsmessungen eigene Datenerhebungen vorzubereiten. Dies geschieht in der Form von zwei Projekten. PISA (*Programme for International Student Assessment*), das erste dieser Projekte, ist auf die Jahre 2000-2010 angesetzt; ILSS (*International Life-Skills Study*), das zweite, wird die erwachsene Bevölkerung betreffen und kurz nach 2000 einsetzen.

Das Projekt PISA setzt den Typus der bisherigen Leistungsmessung fort (es untersucht die Bereiche *Reading Literacy*, *Mathematical Literacy* und *Scientific Literacy* bei den 15jährigen), setzt aber auch neue Akzente:

- Mehr Gewicht als auf Kenntnisse, die direkt mit dem Curriculum zusammenhängen, wird auf Fähigkeiten und Kenntnisse gelegt, von denen angenommen wird, sie seien für das künftige Erwachsenenleben wichtig (die Unterscheidung zwischen «*school*» *skills* und «*life*» *skills* ist für PISA wichtig, obwohl natürlich dazu zu sagen ist, dass «die Schule» auch nichts anderes will als Kinder und Jugendliche auf «das Leben» vorzubereiten). So kommen vermehrt sogenannte cross-curriculare Kompetenzen ins Blickfeld wie z. B. Kommunikations-, Problemlösungs- und Anpassungsfähigkeit, Flexibilität und die Fähigkeit, sich der modernen Informationstechnologien zu bedienen (hier ist allerdings beizufügen, dass schon TIMSS sich keineswegs auf «Schulwissen» konzentrierte, sondern das Alltagsleben und die Umgebung der Schüler/innen in die Fragestellungen miteinbezog).
- Die Datenerhebung erfolgt in einem Zyklus mit dreijährigen Abständen. Jedesmal werden alle drei Bereiche geprüft, jedesmal steht aber eine davon im Vordergrund: Lesen im Jahr 2000, Mathematik im Jahr 2003 und Naturwissenschaften im Jahr 2006. Es ergeben sich so neue Perspektiven für Zeitvergleiche und Trendanalysen.



- Da Indikatoren zwar Hinweise auf wichtige Fragen, aber in der Regel keine direkten Antworten für die Lösung von bildungspolitischen Problemen geben können, entwickelt PISA gleichzeitig einen politikorientierten Analyseplan.

Beim Start zu PISA werden 32 Länder teilnehmen. Das Projekt wird unter der Leitung der OECD von einem internationalen Konsortium unter der Führung des *Australian Council for Educational Research* (ACER) durchgeführt; dem Konsortium gehören ausserdem das niederländische *National Institute for Educational Measurement*, der *Service de pédagogie expérimentale* der Universität Liège sowie die amerikanische Firma *WESTAT* an (OECD, 1999).

Man kann Vermutungen darüber anstellen, wie die Arbeit auf dem Gebiet der international vergleichenden Leistungsmessungen nach dem Ablauf des ersten OECD/PISA-Vollzyklus im Jahre 2010 weitergehen wird, wobei – das zeigt die Geschichte – davon auszugehen ist, dass sich Bildungspolitik und Bildungsforschung in absehbarer Zeit auch wieder andern Schwerpunkten zuwenden werden. Im optimistischen Szenario verstetigen und routinisieren sich diese Leistungsmessungen im Rahmen eines umfassenden Indikatorensystems und werden damit zu einem Teil des Normalinstrumentariums, mit dem Bildungssysteme geführt werden. Mit der Entwicklung der Bildungsindikatoren verbindet sich dann die Vision, dass es einmal möglich sein werde, ein Bildungssystem aufgrund einiger Schlüsselindikatoren in gewissen Aspekten zu charakterisieren, ähnlich wie das heute schon für nationale Volkswirtschaften mit Indikatoren wie z. B. dem Pro-Kopf-Einkommen, der Teuerungsrate, der Arbeitslosenrate und ähnlichen, allgemein bekannten und anerkannten Kennziffern der Fall ist. Im pessimistischen Szenario gelingt diese Verstetigung und Routinisierung aus irgendwelchen Gründen nicht; Leistungsmessungen dieser Art blieben dann prekäre und unregelmässig stattfindende Unternehmen.

Beispiele für bildungspolitische Folgerungen, welche die Forschung aus international vergleichenden Leistungsmessungen zieht

Nachdem die bessere Kenntnis und das bessere Verstehen des eigenen Bildungssystems als Hauptgrund für die Teilnahme eines Landes an international vergleichenden Untersuchungen genannt wurden (U.S. National Research Council, S. 1) wird im folgenden an drei unterschiedlichen Beispielen – zwei schweizerischen und einem amerikanischen – gezeigt, welche Art von Aussagen Forscher/innen aufgrund von Ergebnissen internationaler Vergleiche zuhanden der Bildungspolitik machen können. Diesen Beispielen wird die Liste der Projekte des hier besprochenen Typs mit schweizerischer Beteiligung vorangestellt (das erste, COMPED, begann 1989, das letzte, ILSS, für das die Teilnahme der Schweiz offiziell noch zu bestätigen ist, wird kurz nach 2000 beginnen):



- *Computers in Education Study* (COMPED) / IEA
- *International Assessment of Educational Progress II* (IAEP II) / *Educational Testing Service* (ETS), Princeton, USA (ein privates Unternehmen) / Mathematik und Naturwissenschaften
- *Reading Literacy Study* / IEA
- *International Adult Literacy Survey* (IALS) / Statistics Canada und OECD
- *Third International Mathematics and Science Study* (TIMSS) / IEA
- *Survey of Mathematics and Science Opportunities* (SMSO) / Michigan State University (Spezialerhebung im Rahmen von TIMSS)
- *Civic Education Study* / IEA
- *Programme for International Student Assessment* (PISA) / OECD
- *International Life-Skills Study* (ILSS) / OECD

Ausgewählte schweizerische Ergebnisse aus internationalen Leistungsvergleichen

Das erste Beispiel bezieht sich auf den Versuch, für die Schweiz wichtige Ergebnisse aus fünf der oben aufgelisteten Projekte zuhanden von Bildungspolitik, -verwaltung und -praxis darzustellen. Einen entsprechenden Auftrag erteilte die Schweizerische Koordinationskonferenz Bildungsforschung (CORECHED) zwei in mehreren dieser Projekte führend beteiligten Forschern (siehe: Moser & Notter, in Vorbereitung). Die von ihnen ausgewählten Themen wurden durchwegs nach dem Raster «Problemstellung – Ergebnisse – (bildungspolitisches) Fazit» dargestellt. Die im folgenden referierten Beispiele beschränken sich jeweils auf eine Kurzumschreibung des von den Autoren gezogenen Fazits, d. h. der Überlegungen, die sich die Bildungspolitiker/innen aufgrund der Ergebnisse machen müssten/könnten:

- Zum Thema «Leistungen in Naturwissenschaften in der obligatorischen Schule» ergibt sich das Fazit, dass – sollte die Ausmerzungen der in der Schweiz vorhandenen Defizite ein prioritäres Ziel sein – die Verlängerung des naturwissenschaftlichen Unterrichtes und die stärkere Gewichtung der Fachterminologie in Betrachtung gezogen werden müssten.
- Aus den «Leistungen in Mathematik und Naturwissenschaften am Ende der Sekundarstufe II» lässt sich der bildungspolitisch wichtige Schluss ziehen, dass das duale Berufsbildungssystem eine ähnlich gute mathematisch-naturwissenschaftliche Grundbildung ermöglicht wie die allgemeinbildenden Vollzeitschulen und dass der Schweiz im Vergleich zu Ländern mit rein schulischer Ausbildung wie z. B. Schweden oder die Niederlande keine nennenswerten Nachteile erwachsen.



- Hingegen stellt sich aufgrund der ernüchternden Ergebnisse im Bereich «Kompetenzen der erwachsenen Bevölkerung im Lesen und Rechnen» die Frage, wie weit bzw. wie wenig nicht unbedeutende Teile dieser Bevölkerung auf eine kommende Informationsgesellschaft vorbereitet sind. Massnahmen sind vor allem in der Weiterbildung zu treffen, wo es gilt, nicht nur in die Bereiche der Spitzentechnologie, sondern vermehrt auch in jene der Grundkompetenzen zu investieren.
- Interessant und aussagekräftig ist auch das Fazit zum Thema «Schulleistung und Geschlecht»: Die im internationalen Vergleich in der Schweiz markant «schlechteren Leistungen der Mädchen in der Mathematik sind zu einem Teil auf Vorurteile über die Begabungen von Knaben und Mädchen und damit verbundenem Verhalten von Schülerinnen und Schülern, aber auch von Lehrerinnen und Lehrern im Unterricht zurückzuführen. Lehrpersonen sind nicht frei von Vorurteilen und haben einen Einfluss auf die Geschlechterdifferenzen im Interesse an der Mathematik und auf das Selbstvertrauen in die Mathematikleistungsfähigkeit... Dieser nachgewiesene Wirkungsmechanismus von stereotypen Vorurteilen hat insofern eine positive Seite, als die Verantwortung für Veränderungen zu einem Teil auch bei den Lehrerinnen und Lehrern liegt» (Moser & Notter, in Vorbereitung).
- Ein komplexes Fazit ergibt sich zum Thema «Schulleistungen und Klassengrösse», denn der Vergleich fällt international anders aus als im Innern eines nationalen Bildungssystems. Der internationale Vergleich zeigt, dass Spitzenleistungen in Mathematik und Naturwissenschaften auch in Ländern mit durchschnittlich sehr grossen Klassen (z. B. in Ostasien) erzielt werden können. Trotzdem haben die kleinen Klassen in der Schweiz (und in anderen Ländern) ihre Berechtigung, denn die Mathematikleistungen innerhalb der Schweiz sind in grossen Klassen (mit mehr als 24 Schüler/innen) signifikant schlechter als in kleinen Klassen (mit weniger als 16 Schüler/innen). Allerdings profitieren nicht alle, sondern nur die schwächeren Schüler/innen von den kleinen Klassen.
- Das Fazit zur Frage «Allgemeine oder einseitige Begabung der Schülerinnen und Schüler?» ist vor allem für bildungspolitische Entscheide betreffend die Struktur der Sekundarstufe I von Bedeutung, denn die Ergebnisse zeigen, dass die traditionelle Gliederung dieser Stufe in zwei, drei oder vier Züge (z. B. Real-, Sekundar- und Bezirksschule) zwar für etwa die Hälfte der Schüler/innen angemessen ist, dass sie aber den Bedürfnissen der anderen Hälfte nicht Rechnung trägt, da diese in einzelnen Fächern unter- oder überfordert wird. «Niveaueurse in einzelnen Fächern könnten darum den 50 Prozent einseitig Begabten besser gerecht werden, ohne die 50 Prozent allgemein Begabten zu benachteiligen» (Moser & Notter, in Vorbereitung).
- Aus der Untersuchung der «Schulleistung bei verschiedenen Unterrichtsformen» ergibt sich, dass Unterrichtsformen für den Lernprozess eine untergeordnete Rolle spielen... «Begriffe wie ‘Neue Unterrichtsformen’ erleichtern zwar die Diskussion über Unterricht, sie werden der Komplexität des Lehr-Lern-Prozesses

ses aber nicht gerecht und suggerieren, dass bereits durch die bloße Anwendung einer Unterrichtsform eine Verbesserung des Lernens stattfindet. Die Anwendung einer Unterrichtsform garantiert allein aber noch längst keinen wirksamen Unterricht, denn sowohl guter als auch schlechter Unterricht kann auf sehr verschiedene Weise gehalten werden. Mathematikunterricht ist aber vor allem dann besonders wirksam, wenn er von der Lehrperson didaktisch klar strukturiert wird und wenn die Schülerinnen und Schüler das Gefühl haben, sie würden im Unterricht individuell unterstützt. Diese Merkmale erfolgreichen Lernens sind nicht an eine bestimmte Unterrichtsform gebunden. Sie werden sowohl bei Lehrpersonen, die den Unterricht mehrheitlich durch die Schülerinnen und Schüler selbststeuernd organisieren, als auch bei denjenigen, die den Unterricht mehrheitlich direkt steuern, wahrgenommen» (Moser & Notter, in Vorbereitung).

- Als letztes Beispiel dient das Fazit zum Thema «Arbeitszufriedenheit von Lehrpersonen»: diese ist in der Schweiz im internationalen Vergleich hoch. Berücksichtigt man den vermutlichen Zusammenhang zwischen dieser Variablen und der Schulqualität, so kann bildungspolitisch der Schluss gezogen werden, dass die Investitionen in die Lehrer/innengehälter – eine der Determinanten der Arbeitszufriedenheit – sich lohnen.

Anhand dieser Beispiele aus dem demnächst erscheinenden CORECHED-Band von Urs Moser und Philipp Notter können zwei Dinge schön gezeigt werden. Zum ersten wird ersichtlich, welchen bildungspolitischen Stellenwert Schlussfolgerungen aus internationalen Leistungsvergleichen haben können. Beruhen sie auf einem grundsätzlich unbefriedigenden Sachverhalt (Leistungen in den Naturwissenschaften, Kompetenzen der erwachsenen Bevölkerung im Lesen und Rechnen, Schulleistungen und Geschlecht), so lassen sich – allerdings nicht direkt und vorbehaltlos – mögliche Massnahmen vorschlagen, die zu einer Verbesserung des unbefriedigenden Sachverhaltes führen können. Beruhen sie auf grundsätzlich befriedigenden Sachverhalten (Leistungen in Mathematik am Ende der Sekundarstufe II, Klassengrösse, Arbeitszufriedenheit von Lehrpersonen), so geben sie positive Rückmeldungen zu bereits getroffenen Massnahmen, die umso bedeutsamer sind, je mehr diese Massnahmen allenfalls noch Anlass zu bildungspolitischen Auseinandersetzungen geben (z. B. Klassengrösse und Lehrergehälter). Bei Sachverhalten, die nicht einfach den Kategorien «befriedigend» oder «unbefriedigend» zugeordnet werden können (allgemeine oder einseitige Begabungen, Schulleistungen bei unterschiedlichen Unterrichtsformen), steuern die internationalen Leistungsvergleiche rationale Argumente zu jeweils aktuellen bildungspolitischen Diskussionen bei. Dies gilt im übrigen durchgehend: Untersuchungen der hier besprochenen Art erhellen Tatsachen und tragen so zur empirischen Untermauerung von bildungspolitisch verwendeten Argumenten bei. Bildungspolitische Entscheide können rationaler als bisher gefällt werden. Nie aber können Schlussfolgerungen aus internationalen Leistungsvergleichen mechanisch und direkt in bildungspolitische Lösungen umgesetzt werden; sie müssen im kulturellen Kontext des jeweiligen Bildungssystems bewertet werden bevor



sie – zusammen mit anderen relevanten Elementen – in bildungspolitische Entscheidungen einfließen können.

Zum zweiten wird schon aus den wenigen hier ausgewählten Beispielen ersichtlich, dass bei Untersuchungen der beschriebenen Art um den Kern der Leistungsmessung herum eine Reihe anderer Daten erhoben werden, die weit über die von den Medien in der Regel privilegiert behandelten Länderranglisten hinausgehen. Es handelt sich vor allem um die sogenannten Kontextvariablen. An dieser Stelle kann auch erwähnt werden, dass international vergleichende Leistungsmessungen zusätzlich zum Standarderhebungsprogramm nationale Optionen anbieten. Zudem steht es den teilnehmenden Ländern selbstverständlich frei, das obligatorische Erhebungsprogramm um eigene Fragestellungen zu erweitern und zu bereichern; die schweizerische TIMSS-Erhebung, die deswegen als TIMSS-Plus bezeichnet wird, liefert dazu eine schöne Illustration.

Regionale, kantonale und lokale Auswertungen internationaler Leistungsvergleiche

Das zweite Beispiel für Folgerungen, die aus den Ergebnissen internationaler Leistungsvergleiche für das eigene Bildungssystem gezogen werden können, bezieht sich – stellvertretend für ähnliche Unterfangen in anderen Kantonen oder Regionen – auf den am *Cycle d'orientation* (CO) in Genf unternommenen Versuch, die Daten von TIMSS lokal zu analysieren und didaktisch auszuwerten (siehe Pini & Gabriel, 1998a und 1998b, und Bain & Barblan, 1998). TIMSS bestätigte und präziserte vorerst zwei nicht völlig unbekannte Befunde: die im CO sich verbreiternde Leistungsdistanz zwischen prägymnasialem und nicht-prägymnasialem Zug und die trotzdem vorhandenen leistungsmässigen Überschneidungen zwischen den Schüler/innen der verschiedenen Abteilungen.

Die Analysen im Bereich Mathematik kreisen angesichts dieses Befundes um den Begriff der Heterogenität und die inter-individuellen Leistungsunterschiede in einer Schule, die ursprünglich auf eine gewisse Homogenität innerhalb der verschiedenen Schülergruppierungen angelegt war. TIMSS erlaubt also festzustellen, dass zwischen einem implizit anerkannten Prinzip (Homogenität) und der gegenwärtigen Wirklichkeit (Heterogenität) des CO ein Widerspruch besteht. Die Frage jedoch, ob die Ursachen dieser Heterogenität eher struktureller oder pädagogischer Art sind und damit die Frage, wie sie reduziert bzw. besser bewältigt werden kann (in einer «*conception plus dynamique et plus flexible des parcours de formation*»? [Pini & Gabriel, 1998, S. 50]) muss ausserhalb von TIMSS beantwortet werden.

Im Bereich der Physik und Chemie stellten die Forscher in einem qualitativen Ansatz bei jedem TIMSS-Item oder jeder Itemgruppe die Fragen: «*Pourquoi et comment les élèves réussissent-ils ou échouent-ils? Quels moyens mobilisent-ils pour réussir? A quels obstacles ou difficultés s'achoppent-ils lorsqu'ils échouent?*» (Bain & Barblan,



1998, S. V). Die Antworten auf diese Fragen wurden vor allem aufgrund der Erfahrung von Lehrpersonen und Naturwissenschaftsdidaktiker/innen am CO formuliert. Zu den von international vergleichenden Untersuchungen produzierten Daten wird zwar kritisch angemerkt, sie seien oft zu allgemein und zu abstrakt und die erklärenden Variablen stünden häufig in einem nur mittelbaren Zusammenhang mit dem wirklichen Unterricht, doch wird einer Reihe von TIMSS-Daten durchaus attestiert, sie hätten sich auch im Rahmen dieser didaktischen und lokalen Auswertung als nützlich erwiesen. Wenn dann bei den beiden Autoren trotz dieses Brauchbarkeitsattests eine stark kritische Note betreffend die Grenzen international vergleichender Untersuchungen anklingt und die Frage der Effizienz dieser Art von Wirksamkeitsforschung (französisch eleganter: *l'efficience des recherches sur l'efficacité...*) gestellt wird, wird man den Eindruck nicht ganz los, sie hätten ein bisschen vergessen, was sie vorgängig selbst zu Recht festgestellt hatten, dass nämlich die lokaldidaktische Auswertungsmöglichkeit nicht das oberste Ziel internationaler Leistungsvergleiche ist (Bain & Barblan, 1998). Eine weitere spezifisch didaktische Auswertung der TIMSS-Daten legt Jacquet im Bereich der Mathematik vor (Jacquet, 1998), während Labudde und Pfluger aufgrund eines im Rahmen von TIMSS verwendeten Zusatzinstrumentes die Lern-Lehr-Kultur im Physikunterricht auf der Sekundarstufe II empirisch analysieren (Labudde & Pfluger, 1999).

Im übrigen haben verschiedene Kantone und Regionen TIMSS-Daten auf ihrer spezifischen Ebene ausgewertet, wobei sich jeweils Schlussfolgerungen für die Bereiche Curriculum und/oder Schulstruktur und -organisation und/oder Didaktik (siehe oben) ziehen lassen (siehe dazu: Huber & Ramseier, 1997; Kuratle, 1999; Moser & Oggenfuss, 1998; Pedrazzini-Pesce, 1998; Pedrazzini-Pesce & Vanetta, 1998). Ein Überblick über die Auswirkungen von TIMSS auf die Schweiz findet sich im übrigen in dem von der internationalen Projektleitung herauszugebenden «TIMSS-Impact-Book» (Ramseier & Gretler, in Vorbereitung).

Vermutungen über den Zusammenhang zwischen Leistungen, Systemmerkmalen und übergeordneten Werten in einer amerikanischen TIMSS-Analyse

Am folgenden Beispiel – einer vertieften Analyse der amerikanischen TIMSS-Daten und dem Versuch, daraus bildungspolitisch relevante Folgerungen zu ziehen – soll gezeigt werden, dass diese Daten zwar eine äusserst detaillierte, auf viele Merkmale eingehende Standortbestimmung der mathematischen und naturwissenschaftlichen Bildung in den USA erlauben, dass aber zwischen den Befunden und den allenfalls zu treffenden Massnahmen keine rigorose Beziehung, sondern eine Art Grauzone besteht, die mit Hypothesen und Vermutungen überbrückt werden muss. Insofern entspricht diese Situation der allgemeinen Einsicht, dass bildungspolitische Entscheidungen nicht direkt und unmittelbar aus empirischen Ergebnissen abgeleitet werden können. Von besonderem Interesse ist diese von William H. Schmidt von der Michigan State University geleitete Untersuchung aber, weil ein Zusammenhang



zwischen den teilweise alarmierenden Leistungen amerikanischer Schüler/innen und Systemmerkmalen angenommen wird, von denen wiederum vermutet wird, sie seien tief in den Werten der amerikanischen Gesellschaft verankert («... *the authors admit frankly to moving beyond the empirical data to speculate about relationships that underlie those data*» [Schmidt et alii, 1999, S. 181]). Die Ergebnisse der zahlreichen Einzelanalysen, die im Rahmen dieser Untersuchung durchgeführt wurden, werden in den folgenden sieben Merkmalen amerikanischer Mathematik- und Naturwissenschaftserziehung zusammengefasst:

- eine grosse Zersplitterung in der Umschreibung der Bildungsziele;
- zu weite, aber zu wenig in die Tiefe gehende Lehrpläne («*a mile wide and an inch deep*»);
- Lehrmittel (in der Regel aus privaten Verlagen), die alles einzuschliessen versuchen und damit weiter zur Fragmentierung des Unterrichts beitragen;
- wenig kohärente Unterrichtsgestaltung (jeweils nur wenige Lektionen zum gleichen Thema; mehr verschiedene Aktivitäten in einer Lektion als in andern Ländern);
- hoher, in der Regel in den Lehrplänen verankerter Stoffwiederholungsgrad über mehrere Klassen hinweg;
- Mangel an klar strukturierten Inhalten in den «*Middle Grades*» (siebtes und achtes Schuljahr), wo die Zersplitterung auf zahlreiche Themen ihren Höhepunkt erreicht; die Diagnose des Unterrichts in diesen Schuljahren ergibt auch, dass dieser seinem Charakter nach eher der Primar- als der Sekundarstufe angehört;
- Ungleicher Zugang verschiedener Schüler/innengruppen zu mathematischer und naturwissenschaftlicher Bildung (Schmidt et alii, 1999, S. 191-199).

Diese sieben, von den TIMSS-Daten gestützten Merkmale werden ihrerseits mit drei übergeordneten, teilweise sich allerdings überschneidenden strukturellen Charakteristika des amerikanischen Bildungssystems in Verbindung gebracht, nämlich erstens der Dezentralisierung der Entscheidungsbefugnisse, zweitens dem ungleichen Zugang verschiedener Schülergruppen zu verschiedenen Bildungsinhalten (hier die Überschneidung mit einem der oben erörterten Merkmale) und drittens mit der unklaren Stellung des siebten und achten Schuljahres (*Middle Grades*) im Bildungssystem (auch hier eine gewisse Überschneidung).

Hier soll kurz die Situation erörtert werden, die sich aus der Konfrontation von TIMSS-Daten mit der für die USA und andere föderalistische Länder charakteristischen Mehrebenen-Entscheidungsstruktur mit starker Betonung der dezentralen Ebene ergibt: «*The authors believe that American education is shaped by an accretion of choices... We believe that the accretion of those choices and decisions – and how they are shaped by our beliefs, values and preferences – have created current U.S. mathematics and science education... Because they are the result of many*

choices over time, our educational systems lack the internal coherence found in systems more deliberately designed» (Schmidt et alii, 1999, S. 185-186). Nach einer differenzierten Analyse der bildungspolitischen Entscheidungsstrukturen in den TIMSS-Ländern stellen die Autoren fest: *«These data suggest that not only was distributed decision-making a characteristic feature of U.S. mathematics and science, but it was almost uniquely so among TIMSS countries»* (Schmidt et alii, 1999, S. 189). Schweizerinnen und Schweizer werden sich sagen, dass der in dieser Hinsicht sehr ähnlich liegende Fall ihres Landes durch das «almost» abgedeckt sei; zudem werden sie feststellen, dass gerade die Schweiz mit sehr ähnlichen Entscheidungsstrukturen, aber guten Leistungen in Mathematik, ein Gegenbeispiel zu den USA darstellt. Den Autoren der hochdifferenzierten Studie fällt es offensichtlich nicht leicht, bzw. gelingt es begrifflicherweise nicht, eine klare Folgerung aus dieser Situation zu ziehen. Einerseits suggerieren sie – wenn auch sich absichernd durch den Rekurs auf die Begriffe Hypothese und Vermutung – einen kausalen Zusammenhang zwischen den alarmierenden TIMSS-Ergebnissen der USA und diesen Entscheidungsstrukturen. Andererseits betonen sie selbstverständlich gebührend, dass föderalistische Mehrebenen-Entscheidungsstrukturen wie jene der USA vielfältige Vor- und Nachteile haben. Macht man an dieser Stelle einen etwas vereinfachenden Sprung, so müsste man von den Autoren eigentlich die Formulierung folgender, zugegebenermassen noch nicht sehr realitätsnahen Frage erwarten: Sollen die USA mit dem Ziel der Verbesserung ihrer Leistungen in Mathematik und Naturwissenschaften ihre bildungspolitischen Entscheidungsstrukturen und damit auch die ihnen zugrunde liegenden Werte ändern oder – die Alternative dazu – werden diese Werte so hoch geschätzt, dass ihretwegen ein tiefes Niveau der in Frage stehenden Leistungen in Kauf genommen werden kann? Es liegt nahe, dass die Antwort auf diese Frage dann in der Suche nach einem dritten Weg bestünde (sind z. B. Harmonisierung, Koordination und kooperativer Föderalismus schweizerische Antworten auf eine implizit analoge Fragestellung?).

Frühere, nicht wie bei Schmidt prospektive, sondern retrospektive Versuche, die Auswirkungen international vergleichender Leistungsmessungen auf ein nationales Bildungssystem zu beschreiben, liegen z. B. für Schweden (Marklund, 1989) und Ungarn (Bathory, 1989) vor, während Husén auf bildungspolitische Auswirkungen solcher Untersuchungen allgemein eingeht (Husén, 1987).

Im übrigen ist es interessant, der umfassenden Untersuchung von Schmidt mit ihrer starken Akzentsetzung auf aggregierende Systemmerkmale eine vertiefende Analyse der naturwissenschaftlichen Leistungen in der Schweiz gegenüberzustellen (Ramseier, 1997). Mit grosser Stringenz wird dort erläutert, welche bildungspolitische Aussagekraft die beiden als wichtig erkannten Variablen «Unterrichtsdauer» und «inhaltliche Prioritäten» haben. Weit über diesen konkreten Untersuchungszweck hinaus gehen aber Ramseiers abschliessende allgemeingültige Überlegungen und Kommentare zur Logik internationaler Vergleiche. Sie betreffen sowohl den wissenschaftstheoretischen Hintergrund als auch Erkenntnisse über die Aussagekraft solcher Vergleiche. Realistischer Anspruch können nicht kausale Erklärungen, sondern «nur» empirisch erhärtete Interpretationen sein. Zudem ist ein Bildungssystem



immer auch ein historisch bedingter Einzelfall. So ist die Aussagekraft internationaler Vergleiche zwar immer beschränkt, und doch ist ihr Nutzen unverzichtbar, denn «letztlich liegt die Stärke internationaler Vergleiche nicht im Aufstellen von klaren Wenn-dann-Aussagen, die einfach in technische Handlungsanleitungen umgesetzt werden können. Ihr Nutzen liegt vielmehr darin, dass sie neue Perspektiven eröffnen. Neben der Standortbestimmung zeigen sie,

- wie Leistungen mit Merkmalen von Bildungssystemen zusammenhängen;
- dass Leistungen und Merkmale unserer Schule, die uns selbstverständlich sind, international im Guten wie im Schlechten bemerkenswert sein können;
- dass Bildungssysteme gut funktionieren, bei denen für uns kaum hinterfragbare Selbstverständlichkeiten fehlen, oder die Elemente realisiert haben, die für uns revolutionär wären» (Ramseier, 1997, S. 56).

Zur Kritik an internationalen Leistungsvergleichen

Die Aufmerksamkeit, die international vergleichende Leistungsmessungen gegenwärtig finden, ist so gross, dass auch vielfältige Kritik sich einstellen und entwickeln muss. Hier wird vorgeschlagen, diese Kritik auf drei Ebenen zu kategorisieren und durch je ein oder zwei Beispiele kurz zu illustrieren.

Für die Kritiker der ersten Ebene überwiegen die Gefahren, die unerwünschten Nebenwirkungen und andere negative Aspekte internationaler Leistungsvergleiche. Sie sind deshalb – selbst wenn allenfalls auch einige positive Aspekte zugestanden werden – grundsätzliche Gegner der Produktion und Verwendung solcher Indikatoren. Es ist anzunehmen, dass der Präsident der *Société Pédagogique Genevoise*, der einem kurzen Statement den Titel «Von der Diktatur der Noten zur Diktatur der Indikatoren?» gegeben hat, damit einverstanden wäre, dieser Ebene der Kritik zugerechnet zu werden (Pasquier, 1997). Er befürchtet, dass die Effizienz der Schule künftig an diesen Indikatoren gemessen werde und wirft den vergleichenden Studien vor, sie täten so, «als wären die Ziele der Schule überall dieselben». Diese von tiefem Misstrauen und der Angst vor der Bedrohung des eigenen freien Beitrags zur Qualität der Schule geprägte Defensivhaltung ist verständlich und zeigt, wie hoch der Dialogbedarf noch ist, um die Kritiker dieser Ebene – wenn möglich – davon zu überzeugen, dass internationale Vergleiche zu einem unverzichtbaren Instrument der Kenntnis und somit der Verbesserung unserer Bildungssysteme geworden sind (der vorliegende Artikel versteht sich selbst als ein kleiner Beitrag zu diesem Dialog).

Auf der zweiten Ebene wird die nützliche Funktion internationaler Leistungsvergleiche im Hinblick auf die Weiterentwicklung unserer Bildungssysteme zwar bejaht, aber gleichzeitig stark relativiert. Als Beispiel dieser Art von Kritik dienen hier die sieben Anmerkungen zu TIMSS aus einer Rede zur Verleihung der Ehrendoktor-



würde durch die Fakultät für Pädagogik der Universität Bielefeld. Diese in der deutschen TIMSS-Diskussion beheimateten Überlegungen versuchen, auch das in den Blick zu nehmen, was TIMSS nicht erhebt. «Sie richten sich vor allem gegen eine verkürzte Interpretation solcher Untersuchungen, gegen vorschnelle Schlüsse daraus, gegen die Verabsolutierung einer Messmethode und damit gegen einen rein quantitativen Effizienzbegriff» (Von der Groeben, 1999, S. 7). Knapp zusammengefasst betreffen diese Anmerkungen folgende Kritikpunkte:

- Wider das Ranglistendenken und wider die Vorstellung von Schule als Rennstrecke. Ungeprüfte und problematische Voraussetzungen von TIMSS seien: « 1) Lernen sei ein linearer Prozess der Akkumulation von Wissen, 2) Schule habe folglich möglichst vielen Kindern in möglichst kurzer Zeit möglichst viel davon zu vermitteln und 3) die Qualität von Schule bemesse sich allein oder vorrangig an der Erfüllung dieser Aufgabe» (Von der Groeben, 1999, S. 7).
- Lernwege seien keine Lehrplanwege: Lernprozesse seien individuell verschieden und könnten gegen vorgegebene Normen verlaufen. Qualität von Schule könne auch daran gemessen werden, wie sie dieser Unterschiedlichkeit gerecht zu werden vermag.
- Gute Leistungen im Lande X können auch unter Randbedingungen erzielt werden, die im Lande Y nicht als gute Schulkultur gelten (was in dem Masse nicht gegen TIMSS spricht, als die Randbedingungen in die Untersuchung miteinbezogen werden).
- An geisteswissenschaftlichen und künstlerischen Fächern, konkret an den Formen der Aneignung von Literatur, wird gezeigt, dass das Messbare nicht alles und oft gerade nicht das Wichtige sei.
- Für die gegenüber der Mathematik besseren Biologieleistungen der deutschen Schüler/innen bietet die Forschung «eine besonders persönlichkeitschützende Lernwelt» als Erklärung an. Die kritische Frage laute, warum diese nicht auch in der Mathematik vorhanden sei (das allerdings ist keine Kritik an international vergleichender Leistungsmessung, sondern am deutschen Mathematikunterricht).
- Der Sinn eines Vergleiches zwischen Ländern mit grundlegend verschiedenen Bedingungen (z. B. Entwicklungs- und Industrieland) und damit die Gültigkeit eines «Weltcurriculums» wird bezweifelt.
- Die grossen Unterschiede zwischen den Leistungen an Gymnasien und Hauptschulen hätten in Deutschland nicht zu einem Aufschrei der Öffentlichkeit geführt (dies ist wiederum keine Kritik an TIMSS, sondern an Ungleichheiten im deutschen Bildungssystem und an der Art ihrer Wahrnehmung in der Öffentlichkeit).



Diese Anmerkungen zu TIMSS können hier nicht weiter diskutiert werden; bemerkenswert aber ist, dass in zwei oder drei Punkten nicht TIMSS, sondern zwei gerade durch TIMSS ins Licht gehobene Befunde kritisiert werden und damit der Nutzen von TIMSS attestiert wird.

Auf der dritten Ebene der Kritik schliesslich bleibt der unverzichtbare Nutzen internationaler Vergleiche völlig unbestritten. Es geht hier um methodologische Auseinandersetzungen; kritisiert wird also nicht das Prinzip, sondern das «Wie» solcher Untersuchungen und auch dieses meist nur in methodologischen Einzelheiten. Klassisches Beispiel einer solchen Auseinandersetzung ist jene zwischen Goldstein und Masters, die sich vor allem darüber streiten, was, warum und wie gemessen werden soll (Goldstein, 1995). Aber auch alle Anstrengungen, die Möglichkeiten und Grenzen international vergleichender Leistungsmessungen zu präzisieren, sind dieser Ebene zuzurechnen (siehe z. B. Notter, 1999).

Vermutungen über künftig wichtiger werdende Forschungsaspekte im Bereich internationaler Vergleiche

An zahlreichen Stellen dieses Beitrages ist der Zusammenhang zwischen Leistungen, Bildungszielen und den diesen letzteren zugrunde liegenden Werten angesprochen worden, die Beziehung also zwischen Elementen, die drei verschiedenen Ebenen angehören. Hauptgegenstand internationaler Vergleiche aber sind zur Zeit horizontale Einebenen-Leistungsvergleiche zwischen verschiedenen Ländern. Zum Abschluss dieses Beitrags wird nachstehend die These entwickelt, zukünftige Forschung im Bereich internationaler Bildungssystemvergleiche habe sich vermehrt mit drei bisher relativ vernachlässigten Aspekten zu befassen, nämlich erstens mit der vertikalen Beziehung zwischen Elementen der drei Ebenen (Entwicklung einer Methodologie, Beginn mit nationalen Fallstudien), zweitens mit der Erhellung und Strukturierung der Wertebene und drittens mit horizontalen Ländervergleichen nicht nur auf der Leistungsebene (wie bisher), sondern vor allem auf der Bildungsziel- und der Wertebene.

Vorerst sollen einige Stellen in Erinnerung gerufen werden, in denen die Wertebene angesprochen wird. Um ganz einfach zu beginnen: Was heisst es, wenn Thomas J. Alexander in der Eröffnungsansprache zur zweiten INES-Generalversammlung sagt: «*Ces pays (gemeint sind die asiatischen Länder) nous confrontent à des valeurs culturelles et sociales différentes...*» (Alexander, 1994, S. 18)? Wie wird diese zweifellos richtige Feststellung in das Design international vergleichender Untersuchungen einbezogen? Von Werten wird auch in der *Collaborative Agenda for Improving International Comparative Studies in Education* des U.S. National Research Council's mehrfach gesprochen: Die Kenntnis der implizierten Werte ist zum Kontextverständnis eines Leistungsvergleichs wichtig; internationale Untersuchungen zeigen die bildungspolitischen Optionen auf, mit denen bestimmte Werte in die Praxis umgesetzt werden können, und schliesslich werden Wertfragen ganz allgemein als



Bestandteil von international vergleichenden Untersuchungen bezeichnet (U.S. National Research Council, 1993, S. 10, 11 und 20). Eine sehr grosse Rolle spielen Werte in der oben kurz besprochenen amerikanischen TIMSS-Studie; mit der Frage *«Is there a main story underlying all of these empirical stories about U.S. science and mathematics education?»* zielen deren Autoren auf zentrale Systemmerkmale und die ihnen zugrunde liegenden Werte (Schmidt et alii, 1999, S. 181). Sehr klar wird die Rolle von Werten und Zielen bei Ramseier: *«Wenn wir aus internationalen Vergleichen praktische Konsequenzen ziehen wollen, so müssen wir über die Logik der quantitativen empirischen Sozialforschung hinausgehen. Nicht nur müssen wir Werte und Ziele setzen bzw. von der Bildungspolitik setzen lassen...»* (Ramseier, 1997, S. 56). Schliesslich sei das Projekt PISA zitiert, von dem gesagt wird: *«Its ambition is to assess achievement in relation to the underlying objectives (as defined by society) of education systems, not in relation to the teaching and learning of a body of knowledge»* (OECD, 1999, S. 11).

Insgesamt ergibt sich gegenwärtig der Eindruck eines Missverhältnisses zwischen der unbestrittenen Bedeutung, die der Ziel- und Wertebene zuerkannt wird, und dem tatsächlichen (tiefen) Strukturierungs-, Forschungs- und Wissensstand auf diesen beiden Ebenen. Dieses Defizit führt denn auch dazu, dass die beiden Ebenen – und insbesondere die Ebene der Werte – noch kaum konkret in internationale Leistungsvergleiche und in das System internationaler Bildungsindikatoren integriert sind. Ist dies einmal der Fall, dann können Fragen wie die folgenden gestellt und vielleicht beantwortet werden:

- Wie kann eine Evaluation gleichzeitig auf einen universellen und auf einen nationalen Referenzrahmen bezogen werden?
- Über welche Elemente eines Bildungssystems verwirklichen sich die der entsprechenden Gesellschaft (oder Teilen von ihr) zugrunde liegenden Werte? Der Umgang mit dieser Frage wird wahrscheinlich dazu führen, dass die Beschränkung internationaler Vergleiche auf Mathematik, Naturwissenschaften und *Reading Literacy* durchbrochen werden muss.
- Tragen international vergleichende Untersuchungen im Bildungswesen zu einer universellen kulturellen Homogenisierung oder im Gegenteil zur Erhaltung kultureller Vielfalt bei, bzw. können sie gleichzeitig zur Globalisierung und zur Erhaltung kultureller Vielfalt beitragen? Internationale Vergleiche werden sich damit in die Diskussion einer der wichtigsten Fragen unserer Zeit eingliedern, die weit über den Bereich des Bildungswesens hinausgeht.



Bibliographische Angaben

- Alexander, T. J. (1994). *Allocution d'ouverture (prononcée lors de la deuxième Assemblée générale du projet INES tenue à Lugano-Cadro, Suisse, les 26 et 27 septembre 1991)*. In OCDE. Centre pour la Recherche et l'Innovation dans l'Enseignement (CERI), *Evaluer l'enseignement. De l'utilité des indicateurs internationaux* (S. 15-21). Paris: OCDE.
- Bain, D. & Barblan, F. (1998). *A propos d'une enquête internationale. Physique et chimie*. Genève: Développement et innovation pédagogique au Cycle d'orientation (DIPCO).
- Bathory, Z. (1989). *How Two Educational Systems Learned from Comparative Studies. The Hungarian Experience*. In A. C. Purves, A. C. (Hrsg.), *International Comparisons and Educational Reform* (S. 45-50). Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Bottani, N. & Tuijnman, A. (1994). *Les indicateurs internationaux de l'enseignement: cadre, élaboration, interprétation*. In OCDE. Centre pour la Recherche et l'Innovation dans l'Enseignement (CERI), *Evaluer l'enseignement. De l'utilité des indicateurs internationaux* (S. 25-40). Paris: OCDE.
- Bundesamt für Statistik (1993). *Bildungsindikatoren Schweiz*. Bern: BFS.
- Bundesamt für Statistik (1995). *Bildungsindikatoren Schweiz. Bildungssystem (e) Schweiz im Wandel*. Bern: BFS.
- Bundesamt für Statistik (1998). *Ausgewählte Bildungsindikatoren Schweiz. Standardisierte Kennziffern 1996/97*. Bern: BFS.
- Finn, C. E., Jr. (1989). *A World of Assessment. A Universe of Data*. In A. C. Purves (Hrsg.), *International Comparisons and Educational Reform* (S. 112-116). Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Goldstein, H. (1995). *Interpreting international comparisons of student achievement*. Paris: UNESCO Publishing, Educational studies and documents 63.
- Huber, M. & Ramseier, E. (1997). *Schulleistungen in den Kantonen Basel-Landschaft und Bern. Eine Auswertung der Third International Mathematics and Science Study*. Bern: Amt für Bildungsforschung, Erziehungsdirektion des Kantons Bern.
- Husén, T. (1987). Policy Impact of IEA Research. *Comparative Education Review*, 1, 31, 29-46.
- International Association for the Evaluation of Educational Achievement (IEA) (1993). *IEA Guidebook 1993-1994. Activities, Institutions and People*. The Hague: IEA.
- Jaquet, F. (1998). TIMSS, les apports pour la didactique des mathématiques. *Education et recherche*, 1, 55-69.
- Kuratle, R. (1999). *Was müssen Schülerinnen und Schüler am Ende der obligatorischen Schulzeit können? Bericht über das Hearing vom 30. November 1998*. Solothurn: Erziehungsdepartement des Kantons Solothurn, Abteilung Pädagogik.
- Labudde, P. & Pfluger, D. (1999). *Physikunterricht in der Sekundarstufe II. Eine empirische Analyse der Lern-Lehr-Kultur aus konstruktivistischer Sicht (vervielfältigt)*.
- Landsheere, G. de. (1986). *La recherche en éducation dans le monde*. Paris: Presses Universitaires de France.
- Marklund, I. (1989). *How Two Educational Systems Learned from Comparative Studies: The Swedish Experience*. In A. C. Purves (Hrsg.), *International Comparisons and Educational Reform* (S. 35-44). Alexandria, Virginia: Association for Supervision and Curriculum Development.

- Moser, U. & Notter, Ph. (in Vorbereitung). *Schulische Leistungen im internationalen Vergleich. Ausgewählte schweizerische Ergebnisse aus internationalen Leistungsvergleichen bei Kindern, Jugendlichen und Erwachsenen*. Bern: Schweizerische Erziehungsdirektorenkonferenz (EDK) / Schweizerische Koordinationskonferenz Bildungsforschung (CORECHED).
- Moser, U. & Oggenfuss, F. (1998). *Was leisten internationale Leistungsmessungen? Bericht über die regionale Auswertung der TIMSS-Studie*. Ebikon: Zentralschweizerischer Beratungsdienst für Schulfragen (ZBS).
- Notter, Ph. (1999). Schulen im Test. Möglichkeiten und Grenzen internationaler Leistungsvergleiche. *Zürcher Lehrerinnen- und Lehrerverband (ZLV) Magazin*, 5, 16-19.
- OCDE (1973). *Un système d'indicateurs de l'enseignement visant à orienter les décisions des pouvoirs publics*. Paris: OCDE (vervielfältigt).
- OCDE, Centre pour la Recherche et l'Innovation dans l'Enseignement (CERI) (1994). *Evaluer l'enseignement. De l'utilité des indicateurs internationaux*. Paris: OCDE.
- OECD, Centre pour la Recherche et l'Innovation dans l'Enseignement (CERI) (1996). *Bildung auf einen Blick. OECD-Indikatoren*. Paris: OECD.
- OECD (1999). *Measuring Student Knowledge and Skills. A New Framework for Assessment*. Paris: OECD.
- Pasquier, G. (1997). Von der Diktatur der Noten zur Diktatur der Indikatoren? *Dossier Schweizerische Lehrerinnen- und Lehrerzeitung (SLZ)*, 11, 59 (Übersetzung: Daniel V. Moser-Léhot).
- Pedrazzini-Pesce, F. (1998). *Sistemi scolastici a confronto. Risultati conseguiti dagli allievi della Svizzera italiana nell'ambito del terzo studio internazionale sulla matematica e sulle scienze naturali (TIMSS)*. Bellinzona: Ufficio studi e ricerche (USR).
- Pedrazzini-Pesce, F. & Vanetta, F. (1998). TIMSS: le ricadute sul piano regionale di uno studio internazionale. *Education et recherche*, 1, 83-94.
- Pini, G. & Gabriel, F. (1998a). *A propos d'une enquête internationale. Maths et hétérogénéité*. Genève: Développement et innovation pédagogique au Cycle d'orientation (DIPCO).
- Pini, G. & Gabriel, F. (1998b). Connaissances en mathématiques des élèves genevois du Cycle d'orientation. Résultats de l'étude internationale TIMSS. *Education et recherche*, 1, 70-82.
- Plomp, T. (1998). *The potential and challenges of international comparative studies of educational achievement*. Enschede: Universität Twente (vervielfältigt).
- Purves, A. C. (Hrsg.) (1989). *International Comparisons and Educational Reform*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Ramseier, E. (1997). *Naturwissenschaftliche Leistungen in der Schweiz. Vertiefende Analyse der nationalen Ergebnisse in TIMSS*. Bern: Amt für Bildungsforschung, Erziehungsdirektion des Kantons Bern.
- Ramseier, E. & Gretler, A. (in Vorbereitung). *Impact of TIMSS in Switzerland*. TIMSS Impact Book.
- Schmidt, W. H. & alii. (1999). *Facing the Consequences. Using TIMSS for a Closer Look at U.S. Mathematics and Science Education*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- U.S. National Research Council. Commission on Behavioral and Social Sciences and Education. Board on International Comparative Studies in Education. (1990). *A Framework and Principles for International Comparative Studies in Education*. Washington D.C.: National Academy Press.



U.S. National Research Council. Commission on Behavioral and Social Sciences and Education. Board on International Comparative Studies in Education. (1993). *A Collaborative Agenda for Improving International Comparative Studies in Education*. Washington D.C.: National Academy Press.

Von der Groeben, A. (1999). Leistung wahrnehmen, Leistung bewerten. Sieben Anmerkungen zu TIMSS – oder: Plädoyer für einen differenzierten Blick in der Leistungs- und Qualitätsdebatte. *Pädagogische Führung*, 1, 7-10.



Adresses des auteurs

Jean-Paul Bronckart

Université de Genève
Faculté de psychologie et
des sciences de l'éducation
Rte de Drize 9
CH – 1227 Carouge
E-mail: Jean-Paul.Bronckart@pse.unige.ch

Jean Cardinet

Faubourg du Lac 31
CH – 2000 Neuchâtel
E-mail: Jean.Cardinet@vtx.ch

Marcel Crahay

Université de Liège
Faculté de psychologie et
des sciences de l'éducation
Service de Pédagogie expérimentale
Bd du Rectorat 5
Bâtiment B32
B – 4000 Liège
E-mail: Marcel.Crahay@ulg.ac.be

André Giordan

Université de Genève
Faculté de psychologie et
des sciences de l'éducation
Laboratoire de didactique
et d'épistémologie des sciences
Rte de Drize 9
CH – 1227 Carouge
E-mail: Andre.Giordan@pse.unige.ch

Armin Gretler

Richtergasse 14
CH – 5742 Kölliken
E-mail: a.gretler@echo.ch

Aletta Grisay

Rue Pleyel 2
F – 75012 Paris
E-mail: agrisay@ibm.net

Philippe Perrenoud

Université de Genève
Faculté de psychologie et
des sciences de l'éducation
Bd du Pont-d'Arve 40
CH – 1205 Genève
E-mail: Philippe.Perrenoud@pse.unige.ch

Jacques Weiss

Institut de recherche
et de documentation pédagogique
Fg de l'Hôpital 43
CH – 2007 Neuchâtel
E-mail: Jacques.Weiss@irdp.unine.ch



Dernières publications du SRED

Dans la collection «Cahiers du SRED»

- N° 1: DUCRET J.-J. & JAGASIA N.
La ligne du temps
Enquête psychologique.
Mai 1998. 92 pages.
ISBN 2-940238-00-6.
- N° 2: LURIN J. & SOUSSIA.
La littérature à Genève
*Enquête sur les compétences des adultes
dans la vie quotidienne.*
Septembre 1998. 182 pages.
ISBN 2-940238-01-4.
- N° 3: Textes rassemblés par LURIN J.
& NIDEGGER C.
**Expertise et décisions dans les politiques
de l'enseignement**
Actes du colloque de Penthes, février 1998.
Janvier 1999. 180 pages.
ISBN 2-940238-02-2.
- N° 4: Sous la direction de HUTMACHER W.
Culture religieuse et école laïque
*Rapport du groupe de travail exploratoire
sur la culture judéo-chrétienne à l'école.*
Mars 1999. 164 pages.
ISBN 2-940238-03-0.
- N° 5: VUILLE M. & GROS D.
Violence ordinaire
Juin 1999. 228 pages.
ISBN 2-940238-04-9.

A paraître prochainement

- N° 7: DUCRET J.-J.
Jean Piaget, 1968-1979
*Une décennie de recherches sur les mécanismes
de construction cognitive.*
Premier trimestre 2000.
Environ 300 pages.

**Dans la collection «Documents de travail»**

FAVRE B., NIDEGGER C., OSIEK F.,
SAADA E. & coll.

Le changement: un long fleuve tranquille?

*Dossier exploratoire concernant la rénovation
de l'enseignement primaire genevois.*

Février 1999, 114 pages.

SOMMER A.

**Représentations de la réussite et de l'échec
scolaires.**

*Le point de vue des chercheurs...
et celui des enseignants.*

Juin 1999, 54 p.

SCHAPIRA A.-L.

**Que pensent les élèves de l'enseignement
de l'histoire ?**

Recherche exploratoire chez des pré-adolescents.

Juillet 1999, 85 p.

SAADA El Hadi

**Ecole primaire genevoise, bilinguisme
précoce et contexte interculturel.**

Juillet 1999, 28 p.

Dans la collection «Documents statistiques»

BAERTSCHI C.

Les coûts unitaires en éducation

Aspects méthodologiques.

Octobre 1998, 50 p.

SECTION DE STATISTIQUES

**Mémento statistique de l'enseignement
et de la formation à Genève.**

Edition 1999. Juin 1999, 33 p.

BARRO M. S.

**Indicateurs des finances et des ressources
de l'éducation à Genève.**

Une évaluation préliminaire. Document bilingue
français-anglais. Juillet 1999, 43 p.

WASSMER P.-A.

Ressources humaines et financières du DIP

Données générales et approche par l'activité.

Septembre 1999, 94 p.

SECTION DE STATISTIQUES

**Annuaire statistique de l'enseignement public
et privé à Genève, édition 1999**

Année scolaire 1998-1999.

Septembre 1999,

disponible sur CD-Rom, disquettes

ou sur le site Web du SRED.

© SRED 1999
Service de la recherche en éducation
12, quai du Rhône - 1205 Genève
Tél. (022) 327 57 11 - Fax (022) 327 57 18
<http://agora.unige.ch/sred>

Responsable de l'édition: Narain Jagasia
narain.jagasia@etat.ge.ch

Graphisme: Sophie Jaton
Impression: RoCHAT-Baumann SA / Imprimerie Nationale

Imprimé à Genève en novembre 1999
ISBN 2-940238-05-7

Voyage dans un espace multidimensionnel

Daniel Bain, licencié en lettres et en psychologie de l'Université de Genève, docteur en psychologie de l'Université de Lausanne, a été en 1963 le fondateur et le premier directeur du Centre de recherches psychopédagogiques (CRPP) du Cycle d'orientation de Genève. C'est dans ce cadre qu'il a développé l'essentiel de ses travaux, se consacrant dès 1969 entièrement à la recherche. De 1972 à 1997, il a fait bénéficier ses étudiants de l'Université de Lausanne de sa connaissance approfondie des processus d'orientation scolaire.

Daniel Bain a su, comme nul autre, anticiper les besoins et être à l'écoute des différents acteurs de l'institution scolaire. Ses collaborations avec les enseignants et les chercheurs, à chaque fois stimulées par son savoir théorique, sa rigueur méthodologique, son humour subtil, l'ont conduit à développer une grande diversité de thèmes.

Ce cahier, qui rassemble des textes écrits par différents chercheurs dans des domaines abordés par Daniel Bain tout au long d'une carrière exceptionnelle, lui est remis en témoignage d'amitié.