

**EVALUATION DES SYSTEMES SCOLAIRES PAR DES BATTERIES
D'INDICATEURS DU TYPE PISA :**

**VERS UNE REGRESSION DES PRATIQUES D'EVALUATION
DANS LES CLASSES ?**

Pierre Varcher

Genève - janvier 2002

Pierre Varcher
2, rue Beau-Site
1203 Genève

pvarcher@worldcom.ch

AVERTISSEMENT

Le texte qui suit a été écrit entre 1999 et 2000, soit avant la passation des tests de l'enquête PISA par les élèves. Bien que nous puissions faire l'hypothèse que cette nouvelle enquête menée par l'OCDE ne change pas radicalement les critiques formulées ci-après, le texte devrait néanmoins être revu et complété en intégrant les éléments les plus récents fournis par l'enquête PISA.

D'autre part, ce texte devait, à l'origine, être un des chapitres d'une brochure consacrée plus largement à l'évaluation scolaire, et notamment aux obstacles à l'évolution des pratiques d'évaluation¹. Son approche à contre-courant des politiques scolaires actuellement admises a rendu sa publication difficile dans un ouvrage édité par un organisme étatique. Aussi, les auteurs ont-ils jugé préférable de ne pas compromettre la diffusion de la brochure par le maintien de ce chapitre. Ils restent néanmoins convaincus que les modèles sur lesquels reposent les enquêtes internationales devraient faire l'objet d'un débat critique et constructif.

Pour situer brièvement le cadre général de la publication, il faut mentionner que les auteurs défendent l'idée d'une école génératrice de moins d'échecs et ayant davantage de sens pour les élèves. Ils estiment dès lors que les pratiques d'évaluation devraient changer et ne plus se contenter des traditionnelles épreuves papier-crayon qui conviennent surtout à vérifier une capacité de restitution par les élèves d'un savoir transmis par le maître. Pour être davantage en cohérence avec des modes d'apprentissage plus constructivistes, les pratiques d'évaluation doivent tendre à devenir plus interactives et formatrices. Refusant le discours qui tend à séparer l'évaluation formative de l'évaluation certificative, les auteurs aboutissent à la conclusion que seul un véritable travail de fond sur l'évaluation certificative et une prise en charge de celle-ci dans l'acte d'apprentissage peuvent véritablement faire évoluer les pratiques en classe.

La publication met en évidence les diverses sources de blocages qui s'opposent au développement d'une telle évaluation (partie I) et évoque quelques tentatives d'innovations qui sont néanmoins menées à différents niveaux (partie II). Une troisième partie s'intitule : " Penser la certification autrement pour l'inscrire dans un nouveau paradigme ". L'ouvrage réserve enfin une place importante à la description détaillée de pratiques de classe qui mettent en œuvre la démarche théorique proposée (partie IV). Le texte qui suit devait s'insérer dans la partie consacrée aux innovations car il permettait aux auteurs de poser un regard sur les initiatives institutionnelles en matière d'évaluation et de s'interroger sur leur impact possible au niveau des pratiques de classe.

¹ Davaud, C., Battus, R., Rimaz, J.-L. et Varcher, P. (2001). Approches théoriques et pratiques d'évaluation : l'épreuve de la certification. Genève : DIPCO (Développement et innovation au cycle d'orientation). Mai 2001.

Le développement par les pouvoirs politiques d'indicateurs de l'efficacité des systèmes d'enseignement: les compétences des élèves comme outil de mesure

Un peu partout en Europe et dans les pays industrialisés, " les décideurs des principaux pays membres de l'OCDE ont éprouvé vers la fin des années 1980 le besoin de discuter ensemble des instruments, tels les indicateurs, qui pourraient les aider à mieux suivre l'efficacité de leurs systèmes d'enseignement. L'OCDE a collaboré à deux conférences internationales destinées à clarifier les enjeux : la première, organisée en novembre 1987 à l'invitation du gouvernement américain, était centrée sur les indicateurs de performance, tandis que la deuxième, à l'initiative des autorités françaises, explorait en mars 1988 les interactions des indicateurs et de l'évaluation. " (OFES, 1992, p. 6)

La Suisse a participé à ce type de recherche depuis le début des années 90 en prenant part à l'enquête internationale sur la lecture (sous l'égide de l'IEA, International Association for Evaluation of Educational Achievement) et au programme TIMSS (Third International Mathematics and Sciences Study) qui a cherché à évaluer les connaissances des élèves en mathématiques et en sciences. La décision de continuer à s'engager dans cette voie a été confirmée par la suite : " La Suisse participera à l'évaluation internationale, mise sur pied par l'OCDE, visant à tester les compétences des élèves de 15 ans. Cette évaluation, dont l'appellation internationale est PISA (Program for International Student Assessment), a démarré officiellement cette année [1998]. (...) La première évaluation complète aura lieu au début de l'an 2000. " (PISA-CH, 1998). " La réalisation du projet en Suisse est cofinancée par la Confédération et les cantons. Tous les cantons ont décidé de participer au projet PISA. " (OFS, 1999). 7000 élèves nés en 1984 ont été sélectionnés aléatoirement sur le plan suisse pour le premier test de l'an 2000 et les cantons francophones ont tous choisi l'option qui leur était proposée d'augmenter le nombre d'élèves testés de manière à pouvoir procéder à leurs propres analyses et comparaisons (OFS, 1999).

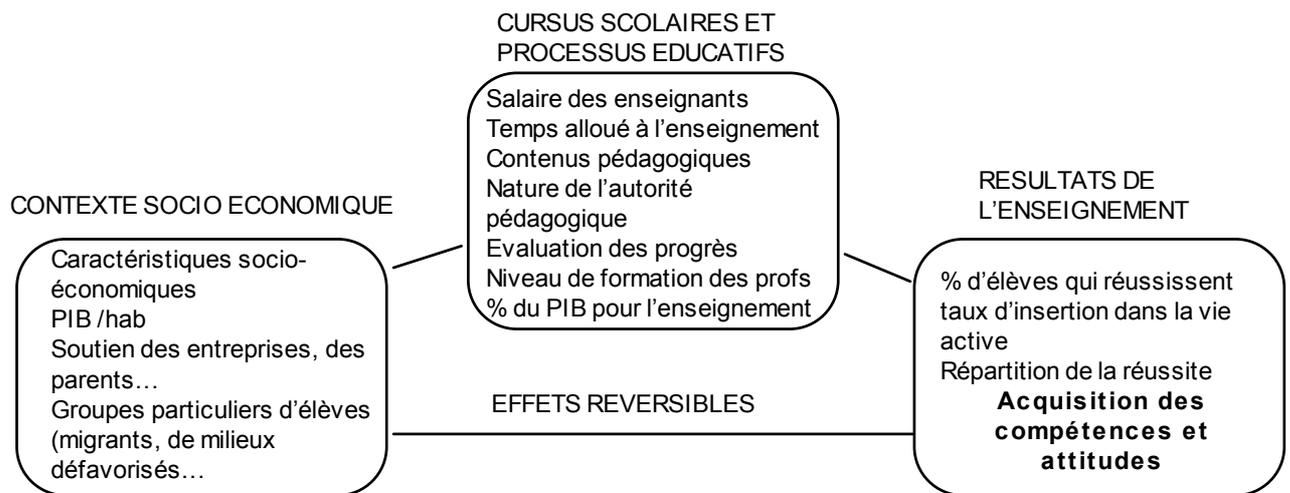
A Genève, la ligne de la Présidente du Département de l'instruction publique est claire à ce sujet. Elle désire, elle aussi, voir développer des indicateurs : " C'est bien sur l'évaluation du système scolaire que l'accent doit désormais être porté. (...) Elle a pour objectif de réguler le fonctionnement de l'institution et de vérifier la réalité des compétences, des maîtrises, des savoir faire qui sont visés à son terme, afin de pouvoir prendre des mesures correctrices touchant à la bonne adéquation des plans d'études, de développer, voire réajuster les dispositifs de formation continue de ses formateurs. " (Département de l'instruction publique, Genève, 1997/1, p. 3). " Quels que soient les motifs qui guident nos dépenses, nous avons des comptes à rendre aux habitants de ce canton. (...) Les résultats scolaires représentent certes un instrument de mesure, mais ils ne sauraient suffire. C'est la raison pour laquelle nous estimons prioritaire de doter notre canton d'instruments d'évaluation et d'indicateurs devant permettre à chacun, élèves, enseignants, parents, citoyens-contribuables, de vérifier les performances de notre système de formation. " (Brunschwig Graf, 1997). Cette intention a été mise en pratique puisque, au début de l'année 2001, le Service de la recherche en éducation a publié un ouvrage intitulé " Le système d'enseignement et de formation genevois. Ensemble d'indicateurs " (SRED, 2001).

Dans tous les pays, de nombreux indicateurs cherchent ainsi à mesurer le contexte démographique, social et économique alors que d'autres portent sur l'étude des cursus scolaires, sur la distribution des ressources matérielles, sur la nature et le nombre des diplômes délivrés. C'est un premier plan, celui des données statistiques.

Mais sur un autre plateau de la balance, il est question aussi de déterminer dans quelle mesure les compétences définies comme essentielles sont acquises par les élèves. C'est la question de la définition d'un indicateur que Scheerens (1995) nomme le "*student academic achievement in the basic skills*" et qu'il définit comme le critère central d'efficacité. Mais, comme l'admet l'OCDE, ce domaine "*demeure la lacune la plus sérieuse des indicateurs*"; les spécialistes déplorent "*l'absence de mesures régulières et fiables des acquis éducatifs des élèves des divers pays - en particulier de mesures de leurs compétences.*" C'est pourquoi "*durant ces dernières années, l'OCDE a pris l'initiative de contribuer directement à améliorer les mesures des résultats, en travaillant avec les pays Membres à l'organisation d'enquêtes comparatives internationales centrées sur les compétences qui sont nécessaires dans la vie d'aujourd'hui.*" (OCDE, 2000, p. 4).

On peut donc dire qu'" *il existe deux principales sources de statistiques sur l'éducation. La première provient des bases de données des organisations internationales, qui contiennent principalement des chiffres sur les ressources, les coûts et les flux; la seconde résulte de travaux (...) qui s'intéressent essentiellement aux acquis des élèves.*" (OFES, 1992, p. 13, n° 18).

La communauté scientifique des chercheurs doit dès lors se préoccuper de mettre en lien ces deux sources de statistiques, notamment pour essayer de montrer dans quelle mesure différents éléments peuvent avoir une influence sur les résultats des élèves. Mais "*la construction d'un ensemble d'indicateurs pour le système d'enseignement sous-tend une représentation des éléments de ce système (...). Or, il n'y a pas de représentation qui s'impose et qui soit unanimement acceptée.*" (SRED, 1999, p. 10). Voici un exemple de mise en lien de certains éléments du système éducatif :



D'après OFES, 1992, p. 44

Il n'est pas question ici de mettre en perspective critique la première de ces sources, celle qui se situe sur le plan des données statistiques, "objectivables". Nous nous intéresserons en revanche au développement et à l'institutionnalisation de la deuxième source d'information, celle qui porte essentiellement sur les acquis des élèves, notamment sous la forme du programme OCDE/PISA².

Dans un premier temps, on pourrait nous objecter que le développement des indicateurs n'a rien de commun avec notre problématique centrée sur l'évolution des pratiques d'évaluation dans les classes. En effet, il est clairement annoncé que la volonté de développer des indicateurs procède d'une intention de séparer la question de l'évaluation du système scolaire de celle de la certification des élèves : "Les travaux de Cardinet, notamment, montrent qu'on ne peut viser ces deux objectifs à la fois" (Département de l'instruction publique, Genève, 1997/2, p. 5); "d'autres modalités que la certification des élèves sont nécessaires pour évaluer l'efficacité d'un système de formation" (ib., p. 3). "Le but de PISA est d'effectuer une enquête sur les caractéristiques globales des élèves dans chaque pays, plutôt que d'examiner les performances d'individus particuliers." (OCDE, 2000). Cette volonté de séparer les deux fonctions et de se limiter à celle d'évaluation de l'efficacité du système scolaire se traduit d'ailleurs dans le projet PISA par un échantillonnage des élèves et des items (les élèves ne reçoivent pas le même cahier de tests) ce qui rend difficile une évaluation individuelle. Les élèves n'auraient aucun souci à se faire puisque ces tests ne sont pas destinés à mesurer, donc à certifier, leurs acquis individuels. On devrait dès lors saluer ces initiatives actuelles et se réjouir de la clarification qu'elles sont censées apporter en levant

² Notons qu'une telle évaluation des acquis des élèves a déjà été réalisée à Genève, avant la mise en place de PISA, avec la démarche OPEC (Observation des performances et estimation des compétences) qui a concerné l'ensemble de l'école primaire genevoise et qui avait pour but de montrer comment évoluent les compétences et les acquis des élèves. A cet effet, des tests ont été organisés pour des échantillons d'élèves des degrés 1, 3 et 6.

certaines des blocages institutionnels que nous dénonçons – les épreuves des maîtres dans leur classe sont encore trop souvent la “ marque ” extérieure que le travail a été fait, que les plans d'études sont respectés ; les maîtres pourraient alors se concentrer sur la formation de leurs élèves et se lancer dans des innovations en matière d'évaluation puisque leurs épreuves ne serviraient plus d'outil de régulation du système.

Il convient pourtant d'être plus circonspect et de tempérer une conception aussi “ aseptisée ” de cette recherche d'indicateurs. L'intention de bien distinguer les fonctions de l'évaluation est peut-être illusoire et elle risque de nous mener en réalité sur une route de mieux en mieux balisée effectuant, sans en avoir l'air, un demi-tour vers une école qui, à Genève, n'aurait plus grand chose à voir avec les finalités de l'article 4 de la loi sur l'instruction publique (LIP). En effet, les promoteurs du programme OCDE/PISA n'ont pas l'intention de limiter leurs ambitions à des constats sur les systèmes scolaires ; ils veulent “ inciter les établissements scolaires et les systèmes éducatifs à se préoccuper des défis de l'époque moderne ” (OCDE, 1999/1, p. 14). Or, il n'y a pas une manière unique, neutre, objective, de définir les défis d'une société ni de déterminer “ les connaissances et compétences fonctionnelles qui permettent de participer activement à la vie de celle-ci ” (OCDE, 1999/1, p. 14). En imposant leur volonté de comparaison à large échelle comme si toutes choses étaient égales par ailleurs – mêmes enjeux de la formation élémentaire, même contexte culturel et socio-économique, mêmes valeurs – les promoteurs de ce programme occultent le débat de fond sur l'école et les intentions des différents acteurs ; jamais, par exemple, n'est rappelé l'objectif fondamental de l'OCDE qui est de promouvoir des politiques visant notamment “ à réaliser la plus forte expansion de l'économie et de l'emploi et une progression du niveau de vie dans les pays Membres ” (extrait de l'art. 1 de la Convention), c'est-à-dire dans les pays déjà les plus riches de la planète. Ces programmes d'évaluation des compétences posent donc des questions essentielles et c'est pourquoi il faut que dès maintenant les différents acteurs de l'école puissent se prononcer et être au clair sur les réels avantages et risques de ces initiatives.

Sur le plan plus précis qui nous occupe ici, celui de l'évaluation des acquis des élèves dans le cadre scolaire, ces programmes imposent en outre leurs conceptions sans que celles-ci ne soient vraiment explicitées et puissent être mises en discussion.

C'est pourquoi, dans le cadre de notre problématique, les deux points suivants sont à examiner attentivement :

1. Quelles sont les finalités précises d'une recherche et d'une mise au point d'indicateurs de pilotage du système scolaire et plus précisément du programme PISA/OCDE ?
2. Quel(s) type(s) d'indicateur sera ou seront utilisé(s) pour “ mesurer ” les acquis des élèves et quelle attitude adopteront les promoteurs face à l'ampleur et à la complexité de cette tâche ?

1. Les finalités précises d'une recherche et d'une mise au point d'indicateurs

Les indicateurs visent essentiellement à mesurer le système éducatif et reposent fondamentalement sur le paradigme de la mesure. Or, même des non spécialistes de l'école voient bien les dangers des démarches de ce type et en dénoncent l'absurdité : *" L'école est l'institution qui fait prévaloir le sens véritable du mot 'qualité'. (...) On ne quantifie pas la qualité, on témoigne qu'elle échappe à toute quantification; qu'elle est précisément et à jamais le contraire même de la quantité. Une 'norme de qualité', dans le monde de l'esprit ? Autant dire : une froideur d'amour, une prison de liberté. Renonçons, pendant qu'il en est encore temps, à ce sinistre contresens."* (Barilier, 1997).

Néanmoins, que ce soit à l'OCDE, en Suisse ou à Genève, on ne remet pas en question l'intention de *" mettre au point un ensemble cohérent d'indicateurs de l'enseignement et un système de points de référence (...) pour les besoins de la planification et de l'élaboration des politiques d'enseignement "* (OFES, 1992, p. 23, n° 4 et 5) et la cohérence est recherchée au niveau de l'ensemble des pays industrialisés, puisque *" les résultats [doivent être] publiés sous forme d'indicateurs des performances comparables sur le plan international "* (OFS, 1999). Le problème est que les éléments permettant de fonder la cohérence recherchée par les concepteurs restent implicites : *" Les théories sur l'efficacité des écoles ne sont presque jamais explicitées avant que la recherche de terrain ne démarre."* (Bosker, 1994). Comment pouvons-nous alors connaître les finalités précises de ce genre de recherche ? N'y a-t-il pas risque qu'un système éducatif, comme celui de Genève, se fasse imposer de " l'extérieur " en quelque sorte des objectifs qui sont contraires ou fort différents de ceux que le législateur lui a fixés ?

La question est d'autant plus légitime que les experts de l'OCDE relèvent eux-mêmes que *" l'analyse de la littérature sur le développement des indicateurs frappe dans ce qu'elle révèle de similitude de thèmes et d'expressions – efficacité, compétitivité économique, qualité, rendre compte – qui se retrouvent de pays à pays par-delà les différences de contextes nationaux comme si avaient été soudain découvertes des vérités d'application universelles."* (OFES, 1992, p. 12). Il n'est pas étonnant de retrouver de tels termes dans les discours portant sur les indicateurs, puisque c'est de manière quasi ontologique que ceux-ci sont liés à la recherche d'efficacité : les auteurs anglo-saxons parlent de l'élaboration d'un dispositif de monitoring de l'enseignement et relèvent qu'il convient de démontrer scientifiquement qu'un tel monitoring ajoute de la valeur à la productivité éducationnelle (Pelgrum & alii, 1995). Scheerens est encore plus explicite puisque, selon lui, la documentation consacrée à l'efficacité de l'entreprise inspire un cadre général et est un modèle implicite de la plupart des recherches sur l'efficacité de l'école (Scheerens, 1995, p. 18). A l'échelle de l'établissement et de l'ensemble d'un système régional ou national, la perspective dans laquelle s'inscrit le modèle de l'" *effectiveness* " de l'école est celle du *" rational goal model "* dont les critères centraux sont l'efficience et la productivité (ib., p. 22).

A l'échelle plus large de l'ensemble des pays de l'OCDE, rappelons que le postulat de base est de confirmer le leadership économique et politique des pays riches. Or, si l'école doit contribuer à faire en sorte que le pays " gagne " la

“ guerre économique ”, on peut légitimement penser qu'elle sera appelée à former des futurs cadres ambitieux, gagners, des “ winners ”, et qu'on est dès lors bien loin des idéaux de l'article 4 de la LIP genevoise par exemple, bien loin de l'idée de l'homme et du citoyen qu'elle cherche à promouvoir. Prenons garde toutefois à ne pas caricaturer la réalité et à ne pas tirer des conclusions immédiates trop simplificatrices : comme le relèvent de nombreux documents d'information de l'OCDE, il existe encore aujourd'hui une certaine confusion dans les pays membres en ce qui concerne les critères permettant de définir l'efficacité de l'enseignement, et les débats font apparaître des divergences de vue, certains s'intéressant davantage à la compétitivité économique et d'autres plus généralement au bien-être et à l'épanouissement personnel. Le débat politique est donc encore ouvert, des opinions différentes s'affrontent, même si les pays de l'OCDE sont engagés dans la même compétition internationale et s'ils estiment que leurs chances de gagner dépendent de leur capacité d'innover dans le domaine des sciences et de la technologie. La pression est dès lors très vive : de facto, l'acquisition des connaissances, le progrès intellectuel, le développement des capacités mentales, l'accès généralisé à la culture scientifique sont devenus des objectifs prioritaires par rapport à ceux qui se rapportent à l'épanouissement et au développement harmonieux de la personne, au plaisir de découvrir et d'explorer le monde, par exemple.

Comme le dit Lowe, “ les études montrent que les responsables de l'enseignement se doivent de fixer et d'atteindre des niveaux élevés de résultats dans chacune des matières d'un programme régulièrement mis à jour, et ce pour quatre raisons principales :

- a) s'assurer que la population active nationale est réellement qualifiée et capable de s'adapter à de constants ajustements structurels de l'économie et du marché du travail et à une concurrence internationale toujours plus intense dans l'économie globale;
 - b) améliorer la qualité et l'efficacité de l'offre d'enseignement et de formation, et répartir les ressources disponibles de façon plus efficace malgré la dureté des contraintes budgétaires ;
- (...) ” (Lowe, 1995).

Jusqu'à présent, ces questions ont peu interféré avec celles des savoirs à acquérir à l'école car chaque pays est resté maître de ses programmes, de ses plans d'études. Même des enquêtes internationales comme TIMSS ont respecté ce principe : elles “ se sont concentrées sur les compétences directement liées aux programmes d'enseignement et n'ont en outre porté que sur les parties des programmes qui sont en principe communes à tous les pays participants ” (OCDE, 1999/1, p. 13). Mais les promoteurs du programme OCDE/PISA annoncent clairement une évolution et relèvent explicitement que leur “ apport le plus important et le plus ambitieux ” porte sur “ les connaissances et les compétences évaluées : celles-ci ne sont pas avant tout définies en tant que dénominateur commun des programmes scolaires nationaux, mais en tant qu'aptitudes jugées essentielles dans la vie future. (...) Le programme OCDE/PISA examine le degré de préparation des jeunes gens à la vie adulte. ” (OCDE, 1999/1, p. 13-14). La brochure de présentation sur Internet, dont le sous-titre est évocateur : “ Pilotage des connaissances et savoir-faire au nouveau millénaire ” (OCDE, 2000), affiche sans ambiguïté l'ambition des concepteurs du programme : “ Grâce au programme PISA, une évaluation des compétences pour la vie sera menée pour la première fois auprès de jeunes d'âge scolaire. ”.

Il est précisé que " *ce sont les gouvernements qui ont pris l'initiative de mettre en œuvre une enquête conçue de manière à répondre à leurs besoins* " et que l' " *ambition est d'évaluer le niveau de formation des élèves par rapport aux objectifs sous-jacents des systèmes éducatifs* " (OCDE, 1999/1, p. 13-14). Et c'est entre parenthèses que l'on rajoute : " *tels qu'ils sont définis par la société* ". A-t-on déjà vu une société définir ce qui est sous-jacent ?

Les programmes internationaux se fondent-ils sur l'hypothèse qu'il n'y a pas de différences importantes entre les objectifs éducatifs des pays participants, ou bien font-ils l'impasse sur les nivellements inévitables que susciteront les indispensables compromis ?

Est-ce qu'il va de soi que les gouvernements, et donc les pays, ont les mêmes besoins en termes de formation, les mêmes critères d'efficacité, les mêmes objectifs pour le système éducatif ? Un futur citoyen aura-t-il besoin des mêmes aptitudes pour réussir sa vie, quel que soit le contexte économique, géographique, politique, historique dans lequel il se trouve ?

Quand on sait que la réalisation du projet sur les plans pratique et technique a été confiée à un consortium formé de trois institutions et d'une entreprise américaine spécialisée dans ce genre de tests (OFS 1999/1), on peut légitimement craindre des dérives. Certes, l'OCDE précise que son secrétariat assume la direction générale du projet et que chaque pays participant dispose d'une voix dans l'organe qui élabore le concept; certes, chaque pays organise la réalisation du projet sur son territoire. Il n'empêche qu'un petit groupe d'experts a pour tâche de définir les aptitudes jugées essentielles dans la vie future pour un habitant d'un pays industrialisé, de définir ce qu'on entend par compétences et de définir les critères d'une existence réussie, dans le cadre relativement restreint de colloques regroupant une vingtaine de personnes, comme ceux du " DeSeCo working group " (OCDE, 1999/2).

Nous touchons là un point extrêmement sensible qu'il conviendrait d'étudier de près, mais ce n'est pas notre objet. Relevons simplement que les bases sur lesquelles s'appuient les experts pourraient être discutées et que les documents officiels cachent le problème en survalorisant les vertus de la représentation et de la délégation de pouvoir.

Sans chercher à poursuivre ici cette analyse, il nous importe de mettre en évidence l'amplification de la coupure entre les milieux qui élaborent ces référentiels de compétences et les enseignants. Nous avons relevé par ailleurs³ qu'un des principaux obstacles à l'évolution des pratiques d'évaluation réside dans le fait que les finalités et les objectifs sont souvent perdus dans les mémoires. Plus les enseignants seront confrontés, dans un débat public et ouvert, à la nécessité de dire au nom de quoi ils agissent, plus leur action sera en

³ Davaud, C., Battus, R., Rimaz, J.-L. et Varcher, P. (2001). Approches théoriques et pratiques d'évaluation : l'épreuve de la certification. Genève : Développement et innovation au cycle d'orientation. Mai 2001.

cohérence avec les buts poursuivis. Or, c'est exactement l'inverse qui se passe dans les enquêtes internationales : en technocratisant le débat sur les connaissances, compétences ou aptitudes fondamentales, au lieu de le politiser, on rend plus difficile la tâche de ceux qui, sur le terrain, essaient de faire évoluer de manière fondamentale les pratiques d'évaluation des enseignants. Le seul moyen qui restera aux autorités scolaires sera d'exiger une application mécanique de nouveaux instruments d'évaluation, sous la menace de restrictions budgétaires en cas de mauvais résultats des élèves ... Mais il est vrai que cela n'est pas en contradiction avec l'approche la plus automatique du " *rational goal model* " dont il est question ci-dessus.

Par ailleurs, les experts de l'OCDE ont été amenés à effectuer des choix et il a été décidé que les enquêtes internationales ne porteraient que sur certains aspects du savoir :

" Dans la mesure où les systèmes d'indicateurs sont nés de la nécessité de contrôler la capacité des écoles à satisfaire les objectifs de la société, ils sont davantage susceptibles de mesurer les aspects de l'enseignement jugés comme ayant un rapport fondamental avec les objectifs économiques, politiques et sociaux – par exemple les résultats des élèves en mathématiques, en sciences, en lecture et en instruction civique – que ceux liés à d'autres domaines du développement cognitif ou de l'épanouissement personnel – comme l'amour du savoir, le développement des aptitudes artistiques et esthétiques ou la capacité de trouver en soi des ressources.

Mesurer l'accomplissement des objectifs de la société exige nécessairement que l'on limite en partie les objectifs des programmes scolaires. Pour évaluer l'acquisition de cette 'quantité déterminée de connaissances positives' qu'un directeur d'école du 19^{ème} siècle a qualifié d'indispensable (Tyack, 1974, 49), il faut que quelqu'un définisse avec précision ce qu'elle est et, partant, également ce qu'elle n'est pas. " (OFES, 1992, p. 78).

Le choix de limiter les enquêtes à certains champs de la connaissance n'est pas motivé par des questions liées aux procédures d'évaluation ou à la nature des objectifs. Il n'est pas non plus intrinsèque aux modalités d'évaluation. Prime ici une volonté utilitariste au niveau de la définition du projet éducatif : " *La place faite aux mathématiques et aux sciences dans les programmes scolaires nationaux reflète en partie la volonté de préparer les jeunes à jouer un rôle moteur dans les activités de recherche-développement qui sont le fondement des économies modernes (...)* " (CERI, 1996). Et, quand l'Office fédéral de la statistique présente PISA au grand public au moyen d'un dépliant, il ne justifie plus le choix des trois " domaines " retenus, comme si ceux-ci allaient de soi. Pour quelle raison les sciences humaines sont-elles écartées de telles recherches ? " *Ces sujets n'intéressent pas suffisamment les organismes concernés pour qu'ils acceptent de dégager les fonds nécessaires.* " (N. Bottani, directeur du Service de la recherche en éducation, SRED, Genève, in Conseil de direction du cycle d'orientation, 1999, p. 22).

De plus, même si les intentions de départ étaient très ambitieuses quant au choix des domaines, les réductions successives liées aux exigences de faisabilité créent un décalage immense avec le produit final. Ainsi, le domaine " compréhension de l'écrit " est en soi une sérieuse limitation du domaine de la langue. Il

comprend des capacités qui, certes, doivent permettre de prendre une part active dans la société (OCDE, 1999/1, p. 15), mais il écarte les compétences de l'ordre de l'oral, souvent encore plus importantes.

A l'intérieur des systèmes scolaires, en se limitant à quelques champs de savoirs, les enquêtes internationales contribuent à focaliser l'intérêt des élèves, des parents, des autorités scolaires et des politiques sur les acquis dans ces deux ou trois disciplines, ce qui peut conduire, à terme, à déséquilibrer la formation des élèves. Les autres apports sont progressivement minimisés et peuvent être réduits au rôle d'options, ou disparaître complètement de la grille horaire des élèves. Ce choix des instances internationales va aussi, dans l'esprit des enseignants, renforcer l'idée que les questions de barèmes, de docimologie, de certification, ne concernent que les branches fondamentales du système scolaire, que seules quelques branches sont dignes d'être enseignées et évaluées, alors que, trop souvent déjà, les enseignants des disciplines qualifiées de "secondaires" ont tendance à ne pas s'impliquer dans les réflexions sur l'évaluation. PISA présente deux caractéristiques qui peuvent amplifier ces phénomènes : d'une part la périodicité des tests qui comporteront à chaque fois des questions concernant les trois domaines retenus, d'autre part, l'affirmation que ces tests évaluent des aptitudes jugées essentielles dans la vie future, ce qui renforce indéniablement le statut des disciplines qui servent de support.

Reste à savoir s'il est possible que de tels tests puissent influencer les contenus d'enseignement, voire les fondements des projets éducatifs eux-mêmes. Les participants aux assemblées générales du CERI (Centre de recherche et l'innovation dans l'enseignement) de l'OCDE se sont interrogés sur l'impact de ces tests et ont pu laisser croire qu'ils sont plutôt inoffensifs : *"L'une des conclusions que l'on peut tirer est que les indicateurs de l'enseignement risquent d'être largement adoptés, mais peu utilisés au niveau de la prise de décision. Les partisans des indicateurs considèrent ceux-ci comme un moyen d'améliorer la décision politique en améliorant les connaissances des décideurs ; mais les chercheurs qui s'intéressent à l'usage qui est fait des sciences sociales savent bien que les indicateurs ont rarement un impact immédiat sur les décisions ou sur les connaissances. Dans le meilleur des cas, la recherche contribue à l'élaboration de la décision politique par le biais de processus larges et diffus 'd'éclairage', c'est-à-dire par une modification du climat de l'opinion (...). S'il en est ainsi, les systèmes d'indicateurs seront susceptibles d'exercer une influence beaucoup plus grande sur l'opinion que sur la décision."* (OFES, 1992, p. 69). Mais un autre rapport présenté à cette même assemblée générale est beaucoup plus lucide, ou peut-être plus sincère : *"Ce que l'on décidera de mesurer dans le contenu des programmes et des processus, peut avoir de sérieuses répercussions sur ce qui est enseigné et appris dans les écoles."* (OFES, 1992, p. 81).

Par quel mécanisme, ces enquêtes pourraient-elles avoir de l'influence sur les contenus scolaires et sur les pratiques d'évaluation des enseignants ? Notre observation du système scolaire nous permet de mettre en évidence deux possibilités de rétroaction : l'une par des mécanismes purement internes et l'autre par une pression externe.

Sur le plan interne au système scolaire, en servant de modèle de référence, les tests officiels pourront instiller leur conception de l'école et du savoir, leurs valeurs, ce qui modifiera petit à petit les finalités du projet éducatif, ou en tout cas celles des disciplines concernées directement par les tests. En effet, ces tests disent, implicitement, " ce qu'il faut réussir " et " ce que réussir veut dire ". Il est facile d'imaginer quels effets les résultats des élèves produisent quand les maîtres concernés voient que leur classe est située bien en dessous de la moyenne cantonale ou nationale sur l'organigramme : la tentation du bachotage est grande, les élèves peuvent être drillés sur des questions semblables à celles des tests. L'effet est déjà bien connu au cycle d'orientation avec les épreuves communes : comme il convient qu'il n'y ait pas trop d'écart entre leurs résultats et ceux des épreuves faites habituellement en classe, les enseignants sont amenés à opérer des ajustements afin que les résultats ne soient pas trop contradictoires. Plus pernicieux encore, ces tests risquent de placer la barre très haut : en cherchant non à certifier les acquis individuels des élèves mais bien à mesurer le niveau global des acquis d'une population d'élèves, ils peuvent conduire à une hausse progressive des exigences des maîtres afin que leurs élèves se situent plus favorablement sur l'échelle des compétences. Par exemple, le test d'entrée OPEC en première année primaire mesurait les capacités des élèves en lecture alors que celles-ci ne sont pas censées être maîtrisées; il a permis certes de dénombrer les élèves sachant lire à l'entrée à l'école, mais il peut faire passer pour des enfants aux capacités peu développées ceux qui n'ont pas appris à lire avant d'entrer à l'école. Il peut aussi inciter les enseignantes d'école maternelle à insister davantage sur l'apprentissage formel de la lecture.

Ces enquêtes servent aussi de modèles de référence en termes de cotation des items, de détermination des barèmes, de présentation des bilans d'évaluation. La légitimité, l'officialité des tests internationaux, nationaux ou cantonaux poussera peu à peu tous les acteurs du système scolaire à évaluer les élèves de la même manière. Elles renforcent l'idée que la quantification des acquis est possible et qu'elle peut être objective ; elles perpétuent les pratiques d'évaluation des élèves au moyen de tests écrits standardisés, construits indépendamment d'une démarche d'apprentissage. Or, nous avons vu dans l'ouvrage déjà cité combien la quantification fait obstacle à une évolution des pratiques d'évaluation scolaire vers une évaluation plus qualitative, globale et formatrice. Sur ce plan, l'intention de séparer la certification des acquis des élèves de l'évaluation du fonctionnement de l'école a de fortes chances de n'être dès lors qu'un vœu pieux.

Une autre pression, extérieure au système scolaire, peut aussi, suite aux enquêtes internationales, influencer les contenus et les pratiques d'évaluation : les réactions de la presse aux résultats de l'enquête TIMSS sont éloquents et montrent le genre de pressions qui peuvent s'exercer sur les écoles et les systèmes scolaires les moins bien notés. " *France, mère des cancrs ?* " titre la revue " *La Recherche* " dans son éditorial de mai 1997. Face aux résultats des élèves genevois qu'elle qualifie de " *pas particulièrement brillants* ", la Conseillère d'Etat chargée du Département de l'instruction publique de Genève réagit promptement : " *Nous avons le devoir d'analyser ces résultats, d'en tirer les enseignements et, le cas échéant, de prendre les mesures adéquates.* " (Brunschwig Graf,

1997). Bien que les sociologues aient affirmé que la prise en compte de ces risques de dérive devrait permettre qu'ils ne se réalisent pas, les récentes exploitations des résultats des tests TIMSS en mathématiques et en sciences démontrent le contraire ; bien que l'OCDE ait mis en garde experts et décideurs sur le fait que *" le désir de certains groupes de pression de disposer de plus d'informations peut se traduire par la mise en place de formes alternatives du recueil de données – par exemple par la publication de classements dans la presse – qui ne répondent pas nécessairement aux critères de fiabilité et validité "* (OFES, 1992, p. 12, n° 13), le communiqué de presse rédigé sur papier à entête du Fonds national suisse de la recherche scientifique pour communiquer les résultats de TIMSS porte comme titre : *" La palme revient aux écolières et écoliers d'Extrême-Orient "* (FNRS, 1996). Le texte insiste surtout sur les classements des élèves suisses comparés aux élèves d'autres nations ; il met en exergue par des sous-titres le fait que *" la jeunesse suisse est dans la moyenne en sciences naturelles, dans le peloton de tête en mathématiques "* et offre quatre pleines pages de classements sous forme de graphiques. Quant à la présentation officielle de ces résultats dans des publications comme *" Regards sur l'éducation. Les indicateurs de l'OCDE "*, elle se passe de tout commentaire : les tableaux et graphiques permettent d'un seul coup d'œil de savoir si la moyenne obtenue dans un pays est statistiquement inférieure, égale ou supérieure à celle d'un autre pays (CERI, 1996, pp. 200 et 201). On ne peut plus parler de risque de dérive, mais quasiment de dérive organisée ... Le travail des journalistes pour trouver un titre alléchant est presque fait et l'Hebdo n'a pas eu besoin de trop réfléchir pour intituler son dossier du 21 août 1997 : *" L'école suisse est-elle assez performante ? "*. Deux ans plus tard, TIMSS était toujours médiatique ; la revue *" Pour la Science "* (N° 265, novembre 1999) reprenait les résultats avec un argument massue pour les rendre crédibles : 500.000 testés dans 50 pays, cela permet d'évaluer objectivement leur niveau. TIMSS donne aussi lieu à d'autres exploitations : deux chercheurs américains ont, par exemple, comparé systématiquement les résultats obtenus aux tests à la dépense per capita pour l'éducation dans chaque état des USA, et globalement entre les Etats-Unis et 12 autres pays (Goldschmidt, P. & Eyer mann, T.S., cités dans *" Sciences humaines "* 101, janvier 2000, pp. 9-10).

La boucle est bouclée : les tests sont développés pour rendre les systèmes scolaires plus efficaces, c'est-à-dire moins dévoreurs de deniers publics, les résultats des tests permettent de lancer le débat dans la presse et de créer, comme on l'a vu plus haut, *" une modification du climat de l'opinion "*. *" Selon la théorie du 'choix public', (...) créer une pression en favorisant un mécanisme de marché peut contribuer à rendre les écoles plus efficaces : financement lié aux outputs, (...), publicité des résultats des écoles, etc. "* (Bosker, 1994). Evidemment, cela nous éloigne d'une éventuelle voie de sortie à notre problème d'évaluation des acquis des élèves dans une perspective plus globale, qualitative, formatrice, et ne reniant pas la fonction certificative.

2. La recherche d'instruments pour mesurer les acquis des élèves : une tâche complexe !

Dans un premier temps, ce sont les propositions de l'IEA et de l'IEAP, organismes qui ont mis sur pied les tests de mesure des acquis linguistiques ou scientifiques (TIMSS) qui ont servi de référence. Pourtant, cette base a semblé peu sûre aux experts même de l'OCDE : *" Il s'agit souvent de recherches ponctuelles, non répétées ou dans des conditions telles - épreuves changées, niveaux différents, échantillons non comparables - qu'il est quasiment impossible de construire des séries temporelles ou d'analyser des tendances. Les techniques d'échantillonnage utilisées et les méthodes de passation des tests ont également fait l'objet de critiques de la part de certains commentateurs ou chercheurs. Enfin, et peut-être pour ces raisons, ces études, bien que très coûteuses, sont souvent victimes de difficultés budgétaires qui ne poussent pas exactement à la rigueur. Aussi les résultats doivent-ils en être examinés avec grande attention et les données disponibles par ce biais sont-elles encore moins standardisées et incontestables que les autres. Il est certes impossible d'évaluer avec précision ce qui se passe dans les salles de classe et nulle mesure des acquis des élèves ne sera jamais parfaite. Il n'en reste pas moins que la standardisation des acquis des élèves est, dans la perspective d'une évaluation globale, d'une telle importance que c'est là un des problèmes les plus urgents à résoudre au plan international. "* (OFES, 1992, p. 14).

Depuis 1992, des progrès méthodologiques ont certes été enregistrés et, à l'occasion de la publication des résultats de TIMSS, la presse a pu relever que *" mille et une précautions semblent avoir été prises pour minimiser les biais méthodologiques "* (revue *" La Recherche "*, 298, mai 1997, p. 5). Pourtant, même des chercheurs à la pointe du mouvement comme Hutmacher admettent que ces tests sont encore *" un laboratoire méthodologique et conceptuel "* (Cycle d'orientation, 1997). Des sommes impressionnantes ont néanmoins été dégagées pour mettre sur pied le programme PISA dont les concepteurs annoncent clairement qu'il sera en rupture avec les enquêtes précédentes auxquelles ils adressent d'ailleurs de nombreuses critiques (OCDE, 1999/1, notamment pp. 13 et 18), sans que ces objections de fond ne remettent pour autant en cause la présentation complètement réifiée des résultats de TIMSS, ni l'utilisation de ceux-ci comme base pour des travaux universitaires de comparaison avec les coûts des systèmes éducatifs. Dès lors deux questions peuvent être posées :

- comment les auteurs s'y prennent-ils pour mesurer les acquis des élèves ?
- ces enquêtes peuvent-elles amener des éléments qui contribueraient à améliorer les pratiques d'évaluation des maîtres ou sont-elles de nouveaux obstacles dressés sur la route de l'innovation ?

Examinons ces deux questions à propos de trois des caractéristiques principales de ces tests :

a) les enquêtes internationales posent, sans leur donner une réponse satisfaisante, les redoutables problèmes de la référence épistémologique des savoirs évalués et du modèle d'apprentissage

En ce qui concerne le choix des items, et pour les enquêtes précédant PISA, les textes officiels se veulent rassurants et rappellent qu'il a été procédé à *"une analyse approfondie des manuels et guides utilisés dans les pays"* et au recueil *"de conseils de spécialistes de ces matières"* (CERI, 1996, p. 194). De fait, comme le relèvent Bain et Barblan qui ont analysé en détail l'enquête TIMSS de 1995 sur les acquis des élèves en physique et en chimie, le cadre conceptuel qui a servi à l'élaboration des tests et des questionnaires est certes le résultat d'analyses approfondies de ce qui est enseigné dans divers pays participants, *"mais aussi la résultante de négociations et de compromis divers. La conception finale, y compris et surtout son opérationnalisation sous formes d'items, est nécessairement de type 'multitudiniste'. Pour des raisons évidentes, les questions posées ne peuvent refléter fidèlement les présupposés épistémologiques et didactiques de chaque système de formation impliqué dans l'enquête."* (Bain & Barblan, 1998, p. 113).

Avec PISA, certains aspects du problème subsistent : la forme et le contenu de l'enquête seront aussi le résultat de négociations et de compromis. Mais deux changements notables sont à relever :

- sur le plan formel, la procédure est beaucoup plus explicitée, les acteurs, les groupes d'experts, sont identifiés ;
- sur le fond, les références sont également explicitées, les trois domaines évalués sont développés. Le problème, c'est qu'ils sont présentés de façon neutre alors que le modèle de référence ayant permis de les définir n'est ni neutre ni explicite.

Par exemple, il est explicitement mentionné comment le programme OCDE/PISA définit la culture scientifique : *"La culture scientifique est le fait de pouvoir utiliser des connaissances scientifiques, d'identifier les questions et de tirer les conclusions fondées sur des faits, en vue de comprendre le monde naturel et de prendre des décisions à son propos, ainsi que de comprendre les changements qui y sont apportés par l'activité humaine. (...) Les savoir-faire concernant les démarches scientifiques ne sont réellement des savoir-faire scientifiques que lorsqu'ils sont appliqués à des matières scientifiques."* (OCDE, 1999/1, p. 68). Il n'est pas de notre propos de discuter cette prise de position. Relevons simplement qu'elle relève d'une réflexion épistémologique qui distingue très nettement les matières scientifiques des matières non scientifiques - s'appuyant en cela sur le concept de science exacte - et qui repose sur l'idée qu'on peut étudier le monde naturel comme un objet distinct des hommes et de leurs activités. Sur ces deux points en particulier, cette conception est en complet désaccord avec la pensée complexe, notamment avec le paradigme constructiviste réaliste fondé sur un systémisme dialogique dans lequel l'intentionnalité des acteurs est un concept central et *"la réintroduction du connaissant dans toute connaissance"* (Morin, 1999, p. 109) est un principe fondamental. PISA prend donc bien position sur le plan épistémologique. Plus pernicieux est le fait que le choix et l'explicitation de ce choix n'apparaîtront plus lors de l'exploitation et de la présentation des résultats de l'épreuve : à ce moment, on ne montrera que les questions, les réponses attendues et les résultats

des élèves. Déjà, avec TIMMS, les présentations des enquêtes donnent l'impression qu'elles " *constituent un ensemble neutre que des pédagogies diverses pourraient utiliser pour tester leur efficacité* " (Bain & Barblan, 1998, p. 114). Or, comme le relèvent ces deux auteurs, " *ce serait faire preuve d'une certaine naïveté docimologique et épistémologique : tout instrument d'évaluation est la traduction d'une certaine conception de la connaissance et de ses modes de transmission.* " (ib., p. 114). Nous avons dénoncé – toujours dans le même ouvrage – les obstacles d'ordre épistémologique (on n'enseigne pas la même chose, donc on n'évalue pas les mêmes choses, si on ne partage pas les mêmes références en termes de savoirs) et ceux liés à la matrice disciplinaire (si les conceptions de la discipline sont différentes, elles induisent des pratiques d'enseignement et d'évaluation différentes). En occultant ces enjeux, PISA ne nous aidera donc pas à clarifier et à améliorer les procédures d'évaluation des enseignants.

De plus, la volonté des concepteurs de l'enquête OCDE/PISA de ne pas se limiter à l'évaluation des curriculums pose un autre problème épistémologique; nous avons déjà relevé que c'est bien l'évaluation des compétences et aptitudes dont les jeunes auront besoin dans leur vie adulte qui est visée. Or, les chercheurs et spécialistes ont encore beaucoup de peine à appréhender ce concept de compétence qui, en tout cas, nous éloigne de l'épistémologie et nous rapproche du sens commun. La conséquence, et cela a déjà été démontré dans une moindre mesure avec TIMMS, est que les notions testées – et c'est encore plus net pour les savoir-faire – n'ont pas trop de fondements clairs sur le plan épistémologique et sont en fait plutôt en relation avec la vie quotidienne, avec l'appréhension commune : " *L'enquête évalue ce que nous appellerons – faute d'un terme plus adéquat – une certaine 'culture scientifique'. En tant que telle, elle n'est pas inintéressante, mais il nous apparaît alors abusif d'invoquer ses résultats pour juger de l'efficacité de notre enseignement.* " (Bain & Barblan, 1998, p. X).

Or, un des gros obstacles à tout changement de paradigme de l'évaluation réside dans le fait que les enseignants continuent à utiliser des moyens traditionnels qui ne sont pas forcément adaptés aux finalités du projet éducatif. Il est alors nécessaire de les engager à revisiter les fondements épistémologiques de leur discipline et le modèle d'apprentissage auquel le curriculum de leur discipline se réfère pour qu'ils amorcent, à partir de là, une réflexion sur ce qu'il est essentiel de faire apprendre aux élèves et comment le leur faire apprendre. C'est de cette manière que non seulement les procédures d'évaluation seront en cohérence avec les finalités de l'enseignement-apprentissage mais aussi qu'elles auront une chance de sortir du carcan des épreuves sommatives de type papier-crayon. Or, les enquêtes internationales qui vont se poser en modèles vont exactement en sens inverse : les références épistémologiques restent peu fondées, la question du modèle d'apprentissage est peu abordée et les items de l'épreuve ne correspondent pas forcément au curriculum... Quel exemple ! Il est indéniable que cela ne va pas remettre en question des enseignants qui pratiquent une évaluation à référence normative, généralement sur la base d'épreuves regroupant des questions qui testent " un peu de tout ce qui a été vu en classe " pendant la période concernée.

b) les enquêtes internationales continuent à utiliser des moyens traditionnels d'évaluation

Visant une évaluation quantifiée d'un produit mesurable, les auteurs des premières enquêtes internationales comme TIMMS se sont limités à des moyens d'évaluation traditionnels. Ils ont ainsi même donné aux questions à choix multiple une importance que celles-ci n'ont pas forcément au niveau du secondaire I. Pourtant, ces questions à choix multiple ont de sérieuses limites : *" la plus souvent mentionnée est le fait que l'on n'atteint ainsi qu'un niveau de reconnaissance (reconnaissance de la réponse juste ou des réponses fausses) plutôt que d'évocation (construction de la réponse juste). "* (Bain & Barblan, 1998, p. X). Sur le plan du savoir, il est souvent souligné que ce type de questions conduit à un morcellement en micro-connaissances dénaturant le savoir disciplinaire. Conscients de ce type de critiques, les auteurs de PISA ont introduit des questions ouvertes : *" Dans l'ensemble, les items utilisés dans le programme OCDE/PISA seront très différents de ceux utilisés (...) dans les études (...) telles que TIMMS "* (OCDE, 1999/1, p. 18). *" Une innovation importante "* consistera à évaluer des *" compétences d'ordre supérieur "* *" souvent par le biais de questions ouvertes "* (ib., p. 17). En faisant cela, ils s'appuient sur l'hypothèse forte selon laquelle on peut mesurer, et quantifier, la résolution de problème ou la capacité d'analyse au moyen de tests papier/crayon standardisables et standardisés.

Est-ce possible ? Notre position est de dire qu'il s'agit d'une mission impossible. Ces questions ouvertes interviennent hors de tout contexte d'apprentissage ; elles présentent des difficultés autant pour les élèves, qui doivent arriver à expliciter leur pensée dans un cadre pas toujours adéquat, que pour le correcteur qui doit tenter de décoder ce qui est écrit. Il n'empêche qu'au bout du compte, la réponse sera évaluée et jugée correcte ou incorrecte ... Pour l'instant, les concepteurs de PISA reconnaissent que *" la place exacte à donner à ce genre d'items dépendra de l'intérêt qu'ils présenteront lors des essais de terrain ainsi que de la possibilité de mettre au point des grilles de correction fiables "* (ib., p. 18). Ils annoncent que ces items à réponse ouverte tiendront *" une place relativement modeste dans le premier cycle d'enquête "* où l'on se contentera donc surtout d'items de type fermé (ib., pp. 18 et 17). Toutefois les tentatives se poursuivent et un groupe d'experts se réunit régulièrement pour essayer de faire entrer des problèmes ouverts dans les étapes ultérieures de l'enquête (OCDE, 1999/2). Jusqu'à présent, les exemples d'items fournis sont peu convaincants (OCDE, 1999/1, p. 76). Pour tenter de mettre l'élève en situation, la consigne devient très longue : il est en effet quasi impossible de contextualiser un problème quand on ne le fait pas vivre par les élèves. De toute façon, l'élève doit *" entrer "* dans une situation problème *" parachutée "* qui reste arbitraire et n'a pas forcément de sens pour lui. Dans un article qui traite des compétences à partir de l'évaluation, Bain (2000) démontre à partir d'exemples d'items qui prétendent évaluer la compétence expérimentale ou la démarche heuristique que *" la nécessité de simplifier et de standardiser les conditions d'observation, de correction, de notation et d'appréciation pour assurer la comparabilité des résultats "* explique qu'en particulier *" la capacité à résoudre un problème est (...) déconnectée de ses dimensions motivationnelles au sens large (opinions,*

valeurs, intérêts, attitudes, habitudes) ” et que “ la démarche heuristique est plus ou moins préprogrammée à travers la consigne, et non pas à inventer ” [par l'élève] (pp. 135 et 134). Quelle différence avec une situation vraiment construite en classe où les questions émergent à partir des représentations des élèves et où l'évaluation s'appuie sur un cheminement vécu ! De plus, il faut encore souligner que ces items donnent un poids déterminant à la composante verbale, y compris dans l'évaluation des compétences dans les domaines mathématiques et scientifiques.

En insistant finalement sur des items fermés (ceux cherchant à mesurer la maîtrise de concepts scientifiques peuvent représenter jusqu'à 50% des points d'une enquête) (OCDE, 1999, p. 78) et en présentant des questions ouvertes peu convaincantes, ces enquêtes risquent bien de conforter les enseignants dans des modèles d'items traditionnels et d'être par là un élément de blocage dans toute tentative d'innovation des pratiques d'évaluation, même si l'on a pu montrer que certains items de TIMMS peuvent servir d'exemple pour des activités d'évaluation novatrices en mathématiques lorsque les pratiques habituelles restent confinées aux algorithmes (Correvon & Seligmann, 1997).

c) les enquêtes internationales sont fondées sur le paradigme de la mesure

Le but de ces enquêtes étant de fournir des éléments statistiquement “ fiables ”, il est clair qu'elles sont bâties sur l'idée qu'il est possible de quantifier les acquis des élèves et d'établir des échelles de performance. Aussi l'ambition de PISA est-elle de mettre au point une “ méthode d'évaluation des élèves (...) susceptible de fournir une mesure robuste des savoir-faire pertinents ” (OCDE, 2000, p. 4) et de fonder “ le matériel d'évaluation [...] sur des méthodes de mesure rigoureuses. ” (OCDE, 1999/1, p. 21). Nous ne nous arrêterons ici que sur un des éléments qui contribue à faire croire à la toute puissance de l'outil de mesure : la présentation des résultats avec des classements précis, des graphiques basés sur des données qui semblent irréfutables, alors qu'elles dépendent non seulement du modèle utilisé pour leur traitement mais également, en amont, du choix des items et des modalités de codification. Cette triple dépendance n'est jamais explicitée ou rappelée, elle est même soigneusement camouflée lors de la présentation des résultats : “ Les résultats de l'évaluation OCDE/PISA seront présentés sous forme de niveaux de performances atteints par les élèves sur les différentes échelles de compétence dans chaque domaine. Les échelles obtenues par la calibration des items du test permettront de décrire les résultats en terme des compétences que maîtrisent les élèves se situant aux divers niveaux de l'échelle. En d'autres termes, il sera possible de dire ce que les élèves, à un niveau donné de l'échelle, savent et sont capables de faire par rapport à ceux qui se trouvent à des niveaux inférieurs. ” (OCDE, 1999/2, p. 18). L'emploi du futur garantit la certitude, les résultats seront vite réifiés, même si les concepteurs eux-mêmes connaissent les limites de l'opération : “ On ne parviendra pas immédiatement à réaliser des évaluations parfaites ” (ib., p. 20). Leur volontarisme et leur foi dans les essais de terrain les rendent convaincus que la chose sera possible : “ Les travaux de recherche disponibles à l'heure actuelle, ainsi que les expériences des enquêtes précédentes suggèrent qu'il sera possible d'établir une échelle de la culture scientifique ” (ib., p. 79), alors qu'il n'y a pas si longtemps, au sein même de l'OCDE, les experts écrivaient : “ Il est impossible d'évaluer avec précision ce qui se

... passe dans les salles de classe et nulle mesure des acquis des élèves ne sera jamais parfaite. " (OFES, 1992, p. 14, déjà citée ci-dessus).

En ne relativisant pas assez les résultats que ces enquêtes permettront de dégager, les experts de l'OCDE contribuent donc à renforcer l'idéologie de la mesure dont on a vu combien elle influence les pratiques des enseignants en classe, notamment les pratiques d'évaluation, combien elle limite toute tentative d'évolution de ces pratiques.

En conclusion

A notre avis, la voie des enquêtes internationales mène à une impasse ou, pire, à une régression des pratiques d'évaluation dans les écoles. Ce n'est pas par un traitement global, par la passation d'épreuves communes à toute une cohorte d'élèves qu'on résoudra les questions de régulation entre enseignants, établissements, régions ou pays, pas plus qu'on ne garantira une objectivité parfaite.

Nous avons essayé de montrer en quoi ces enquêtes à large échelle peuvent avoir une influence négative sur l'évolution des pratiques d'évaluation. Conscients de ces problèmes, plusieurs chercheurs et didacticiens ne voient une issue que dans une séparation drastique entre le travail d'évaluation en classe et tout ce qui relève de la certification. Ainsi, l'équipe du Laboratoire de la Méthodologie de la Géographie (LMG) de l'Université de Liège, en pointe de la didactique de cette discipline en Belgique francophone, affirme-t-elle : *" Nous considérons comme incompatible avec nos valeurs d'effectuer une évaluation certificative en tant que pédagogue. 'L'entraîneur ne peut pas être l'arbitre'. " (LMG, 2000).* Tout en partageant les mêmes valeurs que nos collègues du LMG, nous parvenons à la conclusion inverse : seul un véritable travail de fond sur l'évaluation certificative et une prise en charge de celle-ci dans l'acte d'apprentissage peuvent faire véritablement évoluer les pratiques en classe vers une évaluation globale, qualitative et formatrice qui soit cohérente avec le modèle de référence du savoir et de l'apprentissage. Or, les enquêtes internationales ne nous conduisent pas dans cette voie.

Quant à ceux qui pensent que ces enquêtes sont une idée à creuser, ils se rendent compte que les instruments sont encore loin d'être à la hauteur de leurs ambitions : *" Le concept de 'qualité de l'enseignement' sera difficile à transformer en un indicateur utilisable; la recherche et le développement ont encore un long chemin à faire, mais les 'sponsors' attendront-ils ? " (Creemers, 1995, p. X).* C'est donc l'impatience des décideurs pour rendre l'école plus efficiente et pour trouver des moyens de justifier des coupes budgétaires qui va bloquer toute tentative de faire évoluer les pratiques d'évaluation. C'est pourquoi nous pensons avec Bain et Barblan qu'*" il apparaît raisonnable d'évaluer dès maintenant l'efficacité des recherches sur l'efficacité de la formation. " (Bain & Barblan, 1998, p. 116).*

BIBLIOGRAPHIE GENERALE

Bain, D. (2000). De l'évaluation aux compétences : mise en perspective de pratiques émergentes. In J. Dolz & E. Ollagnier (éds.). *L'énigme de la compétence en éducation. Raisons éducatives 2, 1999/1-2* (pp. 129-145). Bruxelles : De Boeck Université.

Bain, D. & Barblan, F. (1998). *A propos d'une enquête internationale. Physique et chimie*. Genève : DIPCO (Développement et innovation au cycle d'orientation).

Barilier, E. (1997). ISO 9000 dans les écoles. *Hebdo 1^{er} mai 1997*, 107.

Bosker, R.J. (1994). Où en est la recherche anglo-saxonne ? In Crahay, M. (éd), *Evaluation et analyse des établissements de formation. Problématique et méthodologie* (pp. 9-26). Bruxelles : De Boeck-Wesmael.

Brunschwig Graf, M. (1997). A quoi reconnaît-on une bonne école ? *L'Ecole 9, 2*. Genève : Département de l'Instruction Publique.

CERI, Centre pour la Recherche et l'innovation dans l'enseignement (1996). *Regards sur l'éducation. Les indicateurs de l'OCDE*. Paris : OCDE.

Conseil de direction du cycle d'orientation (1999). Procès-verbal du 21.01.1999, n° 1036. Genève : Cycle d'orientation (document interne).

Cornu, L., Pompugnac, J.-C. & Roman, J. (1990). *Le barbare et l'écolier. La fin des utopies scolaires*. Paris : Calmann-Lévy/Essai société.

Correvon, A. & Seligmann, J. (1997). Quelles mathématiques pour les élèves du cycle d'orientation ? Analyse du contenu des épreuves et interrogations pédagogiques. In F. Gabriel (dir.), *A propos d'une enquête internationale, les maths* (pp. 51-122). Genève : Centre de recherches psychopédagogiques.

Creemers, B. (1995). Process Indicators on School Functioning and the generalisability of School Factor Models Across Countries. In OCDE, Centre pour la recherche et l'innovation dans l'enseignement, *Mesurer la qualité des établissements scolaires* (pp. 103-119). Paris : OCDE (Indicateurs des systèmes d'enseignement).

Cycle d'orientation (1997). *Commission de liaison, Ecole primaire - Cycle d'orientation, Procès-verbal de la séance du 15 mai 1997, accepté le 19 août 1997*. Genève : Direction générale, Service de l'enseignement (document interne).

Davaud, C., Battus, R., Rimaz, J.-L. & Varcher, P. (2001). *Approches théoriques et pratiques d'évaluation : l'épreuve de la certification*. Genève : DIPCO (Développement et innovation pédagogique au cycle d'orientation).

Département de l'instruction publique de Genève (1997/1). In Martine Brunshawig Graf, M. *Perspectives pour le degré secondaire 1 ; rapport du département de l'instruction publique du canton de Genève*. Genève : DIP.

Département de l'instruction publique de Genève (1997/2). *Certification du secondaire 1*. Rapport du département de l'instruction publique du canton de Genève, adressé le 27 février 1997 à Moritz Arnet par M. Brunshawig Graf.

FNRS, Fonds national suisse de la recherche scientifique (1996). *Comparaison internationale des prestations scolaires en mathématiques et en sciences naturelles. La palme revient aux écolières et écoliers d'Extrême-Orient*. Berne : FNRS. Communiqué de presse du 19 novembre 1996. 12 p.

Goldschmidt, P. & Eyermaun, T.S. (1999). International educational performance of the United States : Is there a problem that money can fix ? *Comparative Education*, mars 1999. Cités dans *Sciences humaines 101*, janvier 2000. pp. 9-10.

LMG, Laboratoire de la Méthodologie de la Géographie de l'Université de Liège (2000). *Quel type d'évaluation ?* In Recherche sur les compétences terminales en géographie. Site Internet.

www.ulg.ac.be/lmg/competences/chantier/evaluation/evalu_type.html, page consultée le 15 janvier 2000.

Lowe, J. (1995). Vue d'ensemble. In OCDE, *Les normes de résultats dans l'enseignement. A la recherche de la qualité* (pp. 7-27). Paris : OCDE.

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

OCDE, Organisation de coopération et de développement économiques (1999/1). *Mesurer les connaissances et compétences des élèves. Un nouveau cadre d'évaluation*. Paris : OCDE.

OCDE, Organisation de coopération et de développement économiques (1999/2). *PISA Newsletter*, 3, octobre-novembre 1999. Site Internet OECD/PISA Online. www.pisa.oecd.org/docs/news.htm, page consultée le 15 janvier 2000.

OCDE, Organisation de coopération et de développement économiques (2000). *OCDE Programme International pour le Suivi des Acquis des Elèves*. Site Internet OECD/PISA Online.

www.pisa.oecd.org/docs/broch.htm, page consultée le 15 janvier 2000.

OFES, Office fédéral de l'éducation et de la science (1992). *Indicateurs internationaux de l'enseignement. OCDE-CERI (INES). Choix de documents de l'assemblée générale de Lugano-Cadro. 16-18 septembre 1991*. Numéro spécial de Coopération Internationale en Education S3/92. Berne : OFES.

OFS, Office fédéral de la statistique (1999). *Evaluation des compétences chez les jeunes de 15 ans*. Neuchâtel : OFS. Dépliant. 12 p.

Pelgrum, W.J., Voogt, J. & Plomp, T. (1995). Curriculum indicators in International Comparative Research. In OCDE, Centre pour la recherche et l'innovation dans l'enseignement, *Mesurer la qualité des établissements scolaires* (pp. 81-102). Paris : OCDE (Indicateurs des systèmes d'enseignement).

PISA-CH (1998). *Présentation générale du projet*. Office fédéral de la Statistique. Pisa.9802f, 16.4.1998.

Scheerens, J. (1995). The Selection and Definition of International Indicators on Teachers. In OCDE, Centre pour la recherche et l'innovation dans l'enseignement, *Mesurer la qualité des établissements scolaires* (pp. 59-79). Paris : OCDE (Indicateurs des systèmes d'enseignement).

SRED, Service de la Recherche en Education (2001). *Le système d'enseignement et de formation genevois. Ensemble d'indicateurs*. Genève . SRED. Edition 2001.