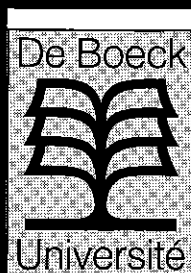


446.733 B

Le pilotage des systèmes d'éducation

Gilbert de Landsheere

PÉDAGOGIES EN DÉVELOPPEMENT



PROBLÉMATIQUES
ET RECHERCHES

PÉDAGOGIES EN DÉVELOPPEMENT

Direction

De Ketele Jean-Marie, Belgique

Comité scientifique international

Ben-Omar Boubker, Maroc; Chevrolet Daniel, France;

Girard André, Québec; Tapia Claude, France;

Tahiri Claudine, Côte d'Ivoire et Roosen Antoine, Belgique.

I
C
O

Le pilotage des systèmes d'éducation

Gilbert de Landsheere



© De Boeck-Wesmael, s.a., 1994.
Rue des Minimes 39 – 1000 Bruxelles

Toute reproduction d'un extrait quelconque de ce livre, par quelque procédé que ce soit,
et notamment par photocopie ou microfilm, est strictement interdite.

Printed in Belgium

D. 1994/0074/33

ISSN 0777-5245
ISBN 2-8041-1876-2

Avant-pr

Poussés par la crise culturelle dont la nouvelle révolution industrielle n'est que l'un des aspects, les pays industrialisés installent, les uns après les autres, le pilotage de leur système d'enseignement.

Ces innovations interpellent le monde de l'éducation par les questions de fond qu'elles soulèvent et par les problèmes techniques de grande complexité qu'elles posent.

Le présent ouvrage tente de faire le tour de la question et de dresser un tableau des principales réalisations - macroscopiques et microscopiques -, dans l'espoir d'y trouver leçons et inspiration. Il voudrait, en particulier, inciter les responsables de la collecte et de la publication d'indicateurs, à concentrer l'essentiel de leurs efforts sur l'évaluation des produits de l'éducation. Ce que les élèves apprennent n'est-il pas le plus important ? Chaque enseignant dans sa classe est aussi concerné que les décideurs, responsables au sommet, et tous les autres membres de la communauté éducative.

En raison de l'un des principaux objectifs poursuivis - inciter à la prise de décision -, le traitement du sujet prend, à plusieurs reprises, la forme d'un dossier, c'est-à-dire d'un ensemble cadré de pièces et de commentaires destinés à faire l'état général d'une question et à fournir, directement ou indirectement, des pistes d'action.

Le choix des quelques exemples, puisés dans un ensemble de travaux manipulant des centaines de millions de données, est arbitraire. Il n'a que valeur illustrative. La bibliographie orientera le lecteur qui souhaite des précisions et des approfondissements.

Enfin, il n'est pas rare que les démarches, en apparence simples, qui sont évoquées exigent, en réalité, le recours à des théories et à des techniques complexes. Certaines d'entre elles font d'ailleurs encore l'objet de controverses entre spécialistes. Dans un ouvrage introductif comme celui-ci, ces débats n'avaient pas leur place. Pourtant, les ignorer totalement eût été dangereux. C'est pourquoi un certain nombre de ces théories et de ces techniques, qui sont à l'avant-plan de l'actualité, sont esquissées en annexe.

Ce travail a été grandement facilité par une collaboration de plus de trente ans aux recherches de l'Association Internationale pour l'Eva-

luation du Rendement Scolaire (IEA) et par une participation aux activités du groupe d'experts qui, avec Norberto Bottani, ont préparé techniquement le lancement du projet "Indicateurs de l'enseignement" de l'Organisation de Coopération et de Développement Economiques (OCDE). Le chapitre consacré aux indicateurs s'appuie largement sur la masse de documents réunis sur ce sujet par le Centre pour la Recherche et l'Innovation dans l'Enseignement (CERI) de cette Organisation.

J'ai aussi bénéficié des encouragements et des remarques constructives de plusieurs collègues réunis par l'Institut Jules Destrée, en particulier de Gérard Fourez, qui ne m'a pas ménagé ses précieuses suggestions, et de M. Crahay, J.-P. Pourtois, H. Desmet et J.-E. Charlier. Qu'ils en soient chaleureusement remerciés ! Ma reconnaissance va aussi aux nombreux collègues et amis qui ont mis des documents, parfois inédits, à ma disposition et m'ont autorisé à les utiliser.

d'éd
duin
min,
Il do
qué
mém

quel
Il fau
niqu
tout
pend

Introdu

"Autrefois frileuse, gardant jalousement ses secrets et ses zones d'ombre, avare de résultats et se méfiant de chiffres, mue par une sainte horreur des <palmarès>, hérissée à l'idée que l'on puisse comparer, dans une région, un département ou une ville, les performances scolaires des établissements, l'éducation nationale a subi ces dernières années, une petite révolution. Sous la poussée des usagers de l'école, que le mythe de l'égalité de tous les élèves dans le creuset scolaire n'abuse plus guère, pressée aussi par les élus embarqués bon gré mal gré sur le navire mais rechignant à payer sans voir, elle a appris progressivement à jouer cartes sur table." (C. Garin, *Le Monde*, 14 janvier 1993).

"Seuls des outils complexes sont capables de dominer la complexité de la réalité" (R. Saint-Geours, *Eloge de la complexité*, *Economica*, 1987).

Etymologiquement, l'idée de pilotage, de guidage est inhérente à celle d'éducation - ce mot ne vient-il pas de *ducere*. En effet, éduquer, c'est *conduire* vers un état souhaitable. Le guide doit non seulement connaître le chemin, mais aussi veiller à ce que l'on ne s'en écarte pas au risque de se perdre. Il doit, en outre, respecter la liberté du voyageur, car, par cette liberté, l'éduqué est sujet au même titre que l'éducateur, particularité qui fonde la notion même d'éducation au sens noble du terme.

Le respect de la liberté a pourtant ses limites, car la culture dans laquelle un enfant doit peu à peu s'insérer n'est pas inscrite dans ses gènes. Il faut donc qu'il s'initie au langage, aux règles, aux connaissances, aux techniques et, plus généralement, à tout ce qui fait la civilisation qui est la sienne, tout en développant son esprit critique, sa capacité de diverger, son indépendance.

De façon plus directe, le pilotage peut se définir comme la prise de décision au niveau macroscopique (c'est-à-dire du système ou d'un sous-système éducatif) ou microscopique (c'est-à-dire d'un établissement, voire d'une classe), sur la base de constats aussi objectifs que possible, relatifs à l'état, au fonctionnement ou aux produits des systèmes. Ces constats ne parlent pas d'eux-mêmes : ils sont l'aboutissement d'une construction de sens à partir d'informations qualitatives et quantitatives constituant des *indicateurs*. Des décisions peuvent ainsi être prises en meilleure connaissance des tenants et des aboutissants des choix.

La comparaison avec la technologie médicale est, en partie, éclairante. Le *monitoring* s'opère à partir de signaux, de "clignotants", que des dispositifs électroniques déclenchent lorsqu'un indicateur (par exemple, le rythme cardiaque) atteint ou dépasse un seuil prédéfini ou lorsque le dispositif tombe en panne à cause d'une anomalie de fonctionnement ou d'alimentation. Toutefois, une telle comparaison a ses limites en raison de maints aspects rationnellement insaisissables du phénomène éducatif.

L'instauration d'un pilotage doit toujours être précédée d'une réflexion approfondie sur le projet éducatif, aussi bien dans le monde politique que dans la société civile. C'est de l'aboutissement, toujours provisoire, de cette réflexion et de la clarification des enjeux politiques que doivent dépendre les lignes de force du pilotage.

Cette clarification concerne en premier lieu les buts et les objectifs, dont le choix dépend, en dernier ressort, de jugements de valeur. Or, en matière d'enseignement, les intérêts divergent et les valeurs aussi. Elles divergent non seulement au niveau des pouvoirs organisateurs, mais aussi de ceux qui réfléchissent à la fonction de l'école dans la société, et des consommateurs directs que sont les élèves. Seule une négociation politique permet de s'accorder ou de trouver des compromis à propos des buts de l'éducation.

Sans qu'ils puissent faire la décision, il appartient aux spécialistes d'aider à définir les enjeux et d'imaginer des solutions à proposer aux acteurs sociaux. Comme le souligne G. Fourez (communication personnelle), "le spécialiste ne peut fonctionner adéquatement comme un expert neutre qui présenterait des contraintes absolues et des solutions uniques, mais il a sa place comme <représentant> des contraintes et comme créateur de solutions nouvelles, dans un débat sociopolitique."

L'idée du pilotage n'est pas neuve. Quand le président des Etats-Unis créa le premier Ministère de l'Education, en 1867, il fut précisé que celui-ci devrait publier chaque année "des statistiques et des observations informant sur l'état et les progrès de l'éducation". On le sait aujourd'hui, un long chemin restait à parcourir, non seulement aux Etats-Unis (qui publient mainte-

nant chaque année une étude statistique relative à 60 indicateurs, intitulée *L'état de l'éducation*), mais aussi dans les autres pays.

Jusqu'à un passé récent, les statistiques relatives à l'éducation étaient essentiellement conçues pour la planification et la gestion. A partir de données démographiques et d'options générales comme l'instauration de la scolarité obligatoire jusqu'à un âge donné, il s'agissait surtout de déterminer le niveau des ressources budgétaires, matérielles et humaines nécessaires (essentiellement des données d'*input*); la productivité des systèmes se mesurait principalement en termes de diplômés (dont on a peu à peu examiné l'origine socio-économique).

Les évolutions récentes de la science, de la technologie, de l'économie et les conquêtes sociales telles que la revendication de l'égalité des chances, le droit à la réussite des études, couplé au droit à la différence, ont notamment suscité de plus en plus d'interrogations sur **la qualité de l'éducation**. Une bonne partie de l'intérêt se déplace donc des variables d'*input* vers les variables de *traitement* ou *processus*, et d'*output*. Le pilotage de la qualité, à chaque moment du *processus* et dans les aspects essentiels des *produits*, est le principe clé de la gestion de la qualité générale.

L'instauration du pilotage, dans un nombre de plus en plus élevé de pays, a une explication culturelle, budgétaire, économique, politique et sociale.

Une innovation importante ne se produit jamais par hasard dans le système éducatif. En effet, celui-ci évolue en fonction du contexte culturel et des forces qui dominent la société. Ainsi, l'essor soudain du pilotage du système éducatif aux Etats-Unis s'explique, d'une part, par le choc psychologique que provoqua le lancement du Spoutnik par l'URSS (la science américaine n'était donc pas la première!) et par l'inquiétude grandissante que suscite l'insuffisance grave des rendements scolaires et, d'autre part, par un progrès des droits civiques garantissant, en principe, une éducation de qualité à tous, pauvres comme riches, noirs comme blancs.

Plus généralement, les sociétés avancées de notre époque se caractérisent par deux phénomènes majeurs : l'intellectualisation toujours accrue de tous les secteurs d'activité et, dans les cas favorables, une adhésion croissante aux valeurs fondamentales de la démocratie.

L'évocation de l'explosion de la science et de la technologie relève maintenant du lieu commun. Plus personne ne doute que le savoir et le savoir-faire de haut niveau (le facteur *k* - pour *knowledge* - cher aux économistes contemporains) sont les clés de l'avenir économique et social.

“Il est clair que le territoire et le capital existent toujours, seulement leur valorisation dépend des processus qui se déroulent dans un autre espace : celui de la connaissance. (...) Dans toutes les sphères de l'économie, et pas seulement dans les services, la gestion intelligente d'informations et de savoirs en trans-

formation permanente est devenue la condition de la survie. (...) Même si nous n'en avons pas l'habitude, même si cela est difficile, l'effort pour repenser le savoir est devenu indispensable. Désormais, le savoir est passé du côté de l'infrastructure. C'est lui qui fonde la prospérité économique et le rang des nations, c'est lui qui départage les entreprises en compétition, et c'est pour cela que nous ne pouvons plus éviter de penser la connaissance" (R.D. Collin, 1993, pp. 1-2).

La situation n'est pas sans danger : "La croissance quasi cancérogène des technologies menace l'existence même de la société, en même temps qu'elle crée de prodigieuses opportunités pour l'humanité" (Gonod, 1990, p.2). Ainsi s'explique l'importance à accorder à l'évaluation de l'impact social de ces technologies, et le pilotage tel qu'on le conçoit aujourd'hui en est une.

Les raisons budgétaires qui poussent au pilotage sont bien connues : on est aujourd'hui loin de la richesse des *Golden Sixties* qui a permis tant d'initiatives généreuses, notamment en matière d'égalité des chances. L'expansion économique s'est ralentie en Occident. Pour conserver leur place sur le marché mondial, les nations les plus avancées s'efforcent de promouvoir l'excellence et d'augmenter la qualité générale de leur enseignement, tout en limitant les dépenses au maximum.

Enfin, sur le plan de la politique générale, la lutte pour le respect des droits de l'homme, notamment le droit à une éducation de qualité, porte ses fruits. L'un des aspects majeurs de ce droit concerne le développement optimal de chacun dans le respect de sa différence.

Bref, deux préoccupations qu'il importe de concilier dominant :

1. L'éducation permet-elle à chacun de réaliser son projet personnel ?
2. L'éducation arme-t-elle bien les nations et les individus appelés à faire face à la nouvelle révolution industrielle et culturelle qui se produit à l'échelle universelle ?

Comme le souligne J.-E. Charlier (communication personnelle), on ne peut éviter une tension idéologique forte entre ceux qui estiment que l'école est une institution chargée de fournir à la société ce dont elle a besoin, s'il le faut en faisant violence à ses usagers, et ceux qui considèrent qu'elle doit répondre aux besoins des élèves. Postuler que les deux vont de pair, se combinent naturellement de façon harmonieuse, relève de l'utopie.

La problématique de la recherche de la qualité doit être mise en perspective. En effet, comme le souligne l'OCDE dans son ouvrage *Une éducation et une formation de qualité pour tous* (1992), l'accusation selon laquelle le "niveau" de l'éducation baisse ou n'est plus adéquat pourrait laisser entendre que les remèdes doivent être recherchés dans les objectifs poursuivis dans le passé et dans la façon d'enseigner d'alors. On omettrait ainsi de

tenir compte des changements économiques, sociaux, culturels et pédagogiques qui viennent d'être évoqués.

Le pilotage commence par la détermination, aussi objective que possible, de l'état de l'éducation et il importe qu'un accord puisse s'établir à propos de ce constat. Il est, en effet, la condition initiale d'une réflexion collective, d'un débat fructueux et constructif. "Tant que les interlocuteurs divergent sur le constat, ils peuvent difficilement converger dans la recherche de solutions" (W. Hutmacher, 1993, p.147). Or on sait combien est aiguë la controverse à propos du "niveau" des "produits" de l'école et de leur évolution.

Lorsque l'on dispose des informations voulues, que l'on a fait "l'état des lieux", il reste à tirer les conclusions et à décider des actions. Pour assurer le respect du projet éducatif de chacun, il importe d'observer, de façon continue, le comportement global du système éducatif (approche macroscopique) et d'établir, en même temps, dans quelle mesure il offre les occasions d'apprendre souhaitées par chacun, et donc de répondre aux demandes personnelles (approche microscopique). Les deux approches sont aussi importantes l'une que l'autre. Le système éducatif fonctionne bien dans la mesure où il réussit à servir simultanément la collectivité et chacun de ses membres en particulier.

Ce qui frappe, aujourd'hui encore, lorsqu'on réfléchit au pilotage du système scolaire, c'est finalement le peu d'informations précises et sûres que l'on possède sur son fonctionnement et sur ce qu'il produit en soi et par comparaison avec d'autres systèmes.

Certes, il n'est pratiquement pas possible d'expliquer de façon complète la genèse, les causes des performances scolaires, tant les facteurs déterminants sont multiples et, d'ailleurs, en bonne partie extérieurs à l'école. Cette complexité se retrouve dans tout ce qui touche l'homme. Ce n'est cependant pas une raison suffisante pour renoncer à réfléchir et à agir.

Le passé

Dans le passé, les responsables de l'enseignement n'ont assurément pas été indifférents au succès de leur entreprise. En particulier, les élèves ont, depuis longtemps, subi des examens et leurs maîtres sont soumis à l'inspection.

Les limitations, voire les injustices de ces contrôles, sont depuis longtemps dénoncées. Néanmoins, dans le cadre social du passé, les choses ne sont pas trop mal déroulées, en ce sens que, s'il y eut beaucoup de dégâts, de talents gaspillés et d'erreurs d'appréciation, les besoins de la société ont pu être satisfaits, de façon souvent inéquitable, il est vrai.

Piloter un système éducatif, c'est plus qu'accumuler des indicateurs. Le pilotage doit nécessairement comporter trois composantes : la collecte régu-

lière d'informations, l'évaluation de ces informations et leur traduction en actions institutionnelles.

Un pilotage qui informe de façon continue et permet de prendre les options éducatives les plus favorables est indispensable. A la limite, il peut être considéré comme une condition *sine qua non* de la survie de notre société d'aujourd'hui et, surtout, de celle de demain.

Un acte démocratique

Déterminer, aussi objectivement que possible, dans quelle mesure sont atteints, dans l'enseignement, les objectifs que la communauté éducative lui a assignés et informer tous les intéressés des résultats de cette observation est une obligation. En effet, en démocratie, tout citoyen a le droit d'être informé sur tout ce qui le concerne : ceux à qui il a confié un pouvoir ou une mission doivent lui rendre des comptes précis et intelligibles. En principe, ce droit s'étend au choix même des instruments et des modalités de l'évaluation.

Il ne suffit donc plus d'annoncer que Pierre ou Paul est premier ou dernier de sa classe (*standard social*), il faut faire savoir s'il progresse effectivement dans son projet personnel (*standard individuel*), et dans quelle mesure il maîtrise ou non les connaissances, les savoir-faire et les savoir-être indispensables à son développement et à son insertion dans la vie sociale (*standards critiques*).

"L'évaluation par indicateurs suppose que l'on fasse de l'utilisateur un membre à part entière du système éducatif, qu'il y soit pleinement intégré. On voit d'ailleurs mal comment les indicateurs de résultats ou de performances pourraient être définis sans l'avis des usagers: élèves, parents, employeurs, institués eux-mêmes comme sujets et non plus comme objets des systèmes d'éducation. Il est dans la logique du mouvement des indicateurs de conduire à un renforcement du pouvoir des usagers dans les systèmes éducatifs" (F. Ewald, 1991, p. 13). L'évaluation sociétale des technologies est appelée à jouer, ici aussi, un rôle majeur. Nous y reviendrons.

Une nécessité pédagogique

De façon plus ou moins consciente, tout enseignant réinterprète les plans d'études officiels, d'une part, en fonction de sa personnalité et de sa qualification professionnelle et, d'autre part, en fonction des caractéristiques de ses élèves. Il dispose, en général, de peu d'éléments qui lui permettent de savoir si ses élèves atteignent réellement les connaissances minimales nécessaires. De surcroît, on ne l'aide pas assez à identifier les difficultés d'apprentissage de ses élèves; pourtant, sans diagnostic correct, il n'est point de médiation adéquate.

ment

ac-

les
peut
so-

son
lui
tion
in-
une
e, ce
ion.

der-
tive-
sure
ndis-
tan-

r un
gré.
nan-
em-
des
s de
uca-
ap-

les
sa
ques
it de
ces-
ren-
re-

Nature du pilo

□ Types de pilotage

En matière d'enseignement, on distingue trois types de pilotage (C.E. Richards, 1988; V. de Landsheere, 1992, p.475) :

• **Le pilotage administratif ou monitoring de conformité**

Il veille au respect des dispositions réglementaires (*compliance monitoring*). Il s'agit d'établir si les écoles atteignent un niveau d'exigence prédéterminé. Dans ce cas, l'unité d'analyse est souvent l'établissement ou le canton scolaire. Les questions posées sont du type :

- Les enseignants possèdent-ils les titres requis ?
- Les curriculums mis en œuvre correspondent-ils au projet pédagogique ?
- Les équipements (bibliothèques, salles d'ordinateurs,...) sont-ils satisfaisants ?
- Les transports scolaires sont-ils bien assurés ?
- Etc.

Ce type de pilotage présente deux caractéristiques majeures :

- Il porte principalement sur l'input.
- Il a pour but principal d'assurer un niveau de service adéquat. Les critères de conformité sont définis par des lois, des règlements ou des instructions administratives. On évalue généralement sur place, notamment à l'aide de checklists.

• **Le pilotage formatif ou pilotage diagnostique**

Il est axé sur les individus. Par exemple, l'administration périodique de tests centrés sur les objectifs, couvrant l'essentiel d'un programme d'études, permet d'identifier les problèmes d'apprentissage qui peuvent se poser, dans un environnement scolaire donné, et d'en chercher la solution.

Ce type de pilotage se fait principalement à l'aide d'épreuves critérielles, épreuves pour lesquelles la définition des domaines principaux couverts par le curriculum est de première importance (Voir à ce sujet V. de Landsheere, 1988, p. 59 sq.).

Le déroulement de ce type de pilotage est, en gros, le suivant :

1. Les élèves subissent un prétest pour déterminer leur niveau de connaissances avant l'enseignement.
2. Les enseignants sont invités à concentrer leurs efforts sur les objectifs les moins bien atteints par leurs élèves.

3. Les enseignants proposent des tâches à réaliser, tant à l'école qu'à domicile, pour aider à atteindre ces objectifs. Dès qu'ils atteignent un niveau de maîtrise suffisant, les élèves en sont informés.
4. Les élèves sont soumis à un post-test.
5. En fonction des résultats obtenus, le professeur élabore éventuellement un plan de remédiation.

Les principes fondamentaux de la pédagogie de la maîtrise, - qui s'applique aussi bien à un enseignement collectif qu'à un enseignement individualisé ou semi-individualisé -, sont ainsi respectés.

• **Le pilotage du rendement scolaire**

Jusqu'à ces derniers temps, il a été généralement réalisé à l'aide de tests normatifs informant de façon comparative sur ce que "produisent" les écoles et les classes d'une région, d'une province ou d'un pays. On assiste actuellement à de nombreuses tentatives d'évaluations critérielles.

Le pilotage du rendement peut être conçu dans deux perspectives différentes. Selon nous, il doit essentiellement fournir des informations comparatives destinées, d'une part, aux pouvoirs organisateurs soucieux de connaître le niveau général et le degré d'homogénéité du produit du système, et d'en informer la communauté éducative et, d'autre part, d'informer *confidentiellement* les enseignants de la situation de leurs classes par rapport aux autres. Dans cet esprit général, les publications de résultats ne permettent jamais d'identifier les écoles.

Là où ces informations sont rendues publiques, notamment aux Etats-Unis (d'autres pays commencent à les imiter), on entend appliquer la **loi du marché** au système scolaire.

Sur le marché, la concurrence joue à plein : les consommateurs utilisent un maximum d'informations pour évaluer la qualité et le prix des produits qui leur sont offerts. Similairement, les parents, bien informés sur les qualités respectives des écoles, peuvent opérer leur choix en conséquence. Et si l'école fréquentée par leurs enfants se situe défavorablement par rapport aux autres établissements, les parents vont éventuellement exercer une pression pour que cette situation s'améliore. Parfois, un transfert vers un établissement plus performant sera opéré. A l'opposé, il arrive aussi qu'une école moins exigeante soit préférée pour les élèves en difficulté...

Arrivé à ce point, il importe de souligner combien une comparaison simpliste entre établissements peut être injuste. Lequel a le plus de mérite ? Celui qui reçoit des élèves en difficulté et réussit à les faire progresser significativement, ou celui qui ne reçoit ou ne garde que les élèves les plus forts et les classe aux premières places dans les palmarès ?

En fait, deux questions fondamentalement différentes se posent.

D'une part, une nation ou une autre entité politique importante peuvent souhaiter connaître ce que leur système éducatif produit, à la fois pour adapter, au besoin, leur politique et pour évaluer **le capital humain** dont elles disposent. Dans ce cas, ce sont des indicateurs globaux qui sont surtout souhaités (**macropilotage**).

D'autre part, l'interrogation peut porter sur la qualité d'écoles considérées individuellement. Dans ce cas, c'est la valeur ajoutée dans chacune d'elles qui offre le plus d'intérêt (**micropilotage**).

□ **Modèle technologique - Modèle humaniste**

De façon schématique, on oppose deux types de gestion des systèmes d'enseignement : le modèle technologique (ou modèle à tendance socio-économique) et le modèle humaniste (ou modèle à tendance psychologisante).

Dans une approche purement technocratique de la gestion de l'enseignement, on fait, en quelque sorte, l'économie de la condition humaine : un produit éducatif étant souhaité, on s'efforce de réunir les matières premières aux meilleures conditions (recrutement des personnes et acquisition des moyens) pour engager un processus ou traitement permettant d'obtenir le résultat désiré. Pratiquement, on cherche, de façon plus ou moins empirique, ce qui "marche le mieux", sans trop se préoccuper des conséquences individuelles et sociales.

A l'opposé, l'approche humaniste prend, elle, pour critère premier le respect de la personne humaine, qui ne peut jamais être considérée comme un simple outil, comme un acteur asservi. On entend toujours respecter sa liberté et tenir compte des projets personnels, pour autant qu'ils ne portent pas atteinte à la liberté et à la dignité des autres.

En fait, la réalité n'est pas aussi tranchée : le modèle technologique n'exclut pas un certain débat démocratique, tandis que le modèle humaniste n'est pas incompatible avec l'utilisation des technologies. Mais les technologies, utilisées de part et d'autre, ne sont pas toujours les mêmes.

• **Sur le modèle technologique**

Le modèle technologique *entrée, traitement, sortie, rétroaction* est aujourd'hui bien connu. Il vise à optimiser la productivité. Le produit de la meilleure qualité au moindre coût est la préoccupation majeure.

Le pilotage ainsi conçu est parfois rapproché du behaviorisme skinnérien, en particulier de l'ingénierie *comportementale*. D'évidence, le gestionnaire d'un système qui concerne des centaines de milliers de personnes doit nécessairement utiliser de façon rationnelle des outils adaptés au macroscopique. Cependant, le modèle technologique n'implique pas nécessairement

une approche mécanique du phénomène humain par excellence qu'est l'éducation. Il n'exclut pas les nuances et les réserves nécessaires lorsqu'on l'applique, et peut d'ailleurs tenir compte d'une marge d'indétermination.

Schématiquement, le modèle technologique peut se décrire de la façon suivante.

Après une analyse des besoins où le facteur économique occupe généralement une grande place, des objectifs sont assignés à travers un programme plus ou moins impératif. Les moyens les plus efficaces et les moins coûteux pour le mettre en oeuvre sont recherchés. Le *rendement*, c'est-à-dire le niveau des performances réalisées par les élèves, est évalué au terme de chaque étape des études. (Il ne faut, toutefois, pas perdre de vue qu'une analyse des besoins n'est jamais neutre : elle est influencée par le choix de société et par les outils d'analyse et les cadres théoriques utilisés.)

Une telle gestion est généralement réalisée par un pouvoir central qui décide des programmes et de la façon de les mettre en oeuvre. A ceux qui occupent le bas de la hiérarchie, il reste à exécuter le travail (*modèle centre-périphérie*).

Le recourt aux technologies peut avoir des effets sociaux indésirés et parfois irréversibles. Par exemple, le danger de décentration du curriculum en fonction des points de focalisation du pilotage ne doit pas être sous-estimé. La problématique des effets de reflux des examens, bien étudiés par Madaus et Kellaghan, nous servira de première illustration.

Madaus et Kellaghan (1992, p.143) distinguent deux grandes catégories d'évaluation. Dans la première, ils classent les évaluations réalisées presque quotidiennement en cours d'enseignement (simples contrôles, non menaçants, de savoirs ou d'habiletés; évaluations formatives,...), dont l'enjeu n'est pas considérable. Dans la seconde catégorie, ils rangent les évaluations dont les résultats peuvent avoir des conséquences graves pour l'individu (accès aux études supérieures, obtention d'un diplôme de fin d'études,...). Dans ce second cas, on observe généralement un important *effet de reflux négatif* : tant les contenus enseignés que la didactique tendent à être plus déterminés par le souci de faire réussir les épreuves que par la volonté d'assurer une éducation de la meilleure qualité possible.

Les mêmes auteurs signalent (p.144) que, dès 1868, une commission d'enquête anglaise conclut : "L'examen détermine le curriculum et non l'inverse. Il décourage l'expérimentation pédagogique, limite le choix des matières enseignées, entrave la liberté méthodologique et ne sert pas l'adhésion aux meilleures valeurs dans la classe."

En effet, l'enseignement qui, par exemple, vise prioritairement la réussite à un examen axé sur les connaissances factuelles risque de se faire aux dépens du développement social, esthétique, affectif et physique des élèves. Cette influence peut s'exercer non seulement sur l'année scolaire au terme

de laquelle l'examen aura lieu, mais aussi sur les précédentes (effort individuel d'un enseignant pour défendre sa réputation ou effort collectif pour défendre l'image de l'établissement). De leur côté, les élèves, quand ils en ont le choix, opteront souvent plus en fonction des exigences d'un programme que de leurs intérêts réels.

En outre, des épreuves nationales standardisées peuvent gêner l'adoption de curriculums locaux, adaptés aux besoins d'environnements particuliers.

A l'opposé, il est vrai, l'effet de reflux peut aussi être positif en incitant les enseignants et les élèves à accorder plus d'importance à des contenus ou à des processus importants, qu'ils ont éventuellement tendance à négliger.

Un autre aspect doit aussi faire l'objet de beaucoup d'attention : celui de la maîtrise démocratique des outils de pilotage. Le problème est épineux, car, dans la situation actuelle, la plupart des enseignants ne sont pas familiarisés de façon suffisante avec les méthodes et les techniques - de plus en plus sophistiquées - de l'évaluation.

• Sur le modèle humaniste

Sans ignorer l'inévitable enchaînement : *objectif - input - traitement - output - feedback*, l'approche humaniste déplace les priorités et fait une beaucoup plus large place au qualitatif, à l'éthique.

Pour le psychologue humaniste, l'expérience telle qu'elle est vécue par chaque personne et le sens que celle-ci lui donne sont les principaux objets d'étude. Le souci constant est la dignité et la valeur de chaque individu; il doit lui être permis de développer le potentiel qui est le sien. Une place importante est accordée à ce que l'on pourrait appeler une saine déviance, c'est-à-dire à l'expression de talents propres, et à la faculté de suivre librement le chemin que l'on choisit pour réaliser l'actualisation de soi. Etre soi, c'est, en effet, pouvoir diverger. Dans ce contexte, les sentiments et les valeurs ont toujours droit au respect et l'individu reste libre de construire son avenir. La liberté, l'amour, le don de soi, l'adhésion à des valeurs, l'autodiscipline, la loyauté, la créativité, l'espoir passent à l'avant-plan.

"La finalité de la maîtrise sociale de la technologie", écrit Gonod (1990, p.10), "c'est sa régulation en fonction des besoins de l'humanité, si l'on se place sur le plan global des sociétés et des collectivités, et, en définitive, des individus."

Le modèle périphérie-centre est associé à l'option humaniste. Il tend à décentraliser la gestion du système éducatif et à donner un pouvoir d'initiative maximum à la "base".

Cette tendance à la décentralisation est nette, non seulement vers les régions, les communautés éducatives locales, mais aussi vers les établisse-

ments encouragés à formuler un projet propre, ce qui implique aussi bien une certaine indépendance dans la gestion, l'utilisation des ressources, la conception du curriculum, que dans la formation ou le perfectionnement du personnel. "L'autonomie est sans doute l'expression la plus adaptée de la modernité, la clé de voûte de toute entreprise de rénovation.(...) Elle n'existera réellement que si elle apparaît dans la gestion quotidienne des établissements" (Dupont, 1988, p.150).

Actuellement, on observe deux tendances contradictoires. On accorde plus d'autonomie aux établissements, mais, en contrepartie, on renforce les dispositifs d'évaluation de façon à amener les établissements à rendre explicitement des comptes sur la qualité de leur travail. L'autopilotage est encouragé, mais le macropilotage se développe en même temps pour établir dans quelle mesure l'ensemble des établissements atteignent les objectifs jugés essentiels et donc prioritaires par le pouvoir organisateur ou la communauté éducative globale.

- **Attitudes technocratiques -
Attitudes démocratiques**

G. Fourez aime souligner que, dans le débat qui nous occupe, la dimension sociétale est, en dernière analyse, la plus décisive. Car, bien que presque toujours soucieux, - à des degrés divers, il est vrai -, des valeurs sociales et humaines, tous les pilotages font appel à des technologies. Certains accusent cependant un caractère technocratique très marqué en ce sens que les savoirs scientifiques et les techniques sont censés directement fournir à la société des solutions à ses problèmes : le débat politique, caractéristique des démocraties, est ainsi escamoté. "D'autres pilotages suivent, pour emprunter les catégories de J. Habermas, une logique décisionnelle, laissant aux décideurs politiques le choix des objectifs et prétendant se situer simplement au niveau des moyens. Cette manière d'aborder la question laisse déjà plus de place aux débats et aux négociations politiques, au moins par rapport aux finalités, - même si l'approche reste technocratique par rapport aux moyens. Enfin, le modèle pragmatique de Habermas souligne la nécessité d'un va-et-vient continu entre les analyses scientifiques et techniques, d'une part, et la détermination des buts, d'autre part. Cette attitude rend une place au débat politique, y compris à propos des contributions des spécialistes. Elle est essentielle au développement de la démocratie dans une société technologique." (G. Fourez, Communication personnelle).

Ces différentes attitudes ne sont pas basées sur une utilisation ou sur le mépris des technologies. Des tenants du débat démocratique, à tous les niveaux, peuvent accepter pleinement tout ce que les technosciences leur offrent. Et des technocrates peuvent être très soucieux des conséquences humaines et sociales des décisions. La différence réside essentiellement dans la

place qui sera réservée aux experts. Dans une approche technocratique, leur pratique est considérée comme socialement neutre, tandis que, dans le modèle pragmatique, la parole des spécialistes a sa place dans un débat de société.

En raison même de sa composante technologique, le pilotage de l'enseignement doit, comme les autres technologies, faire l'objet d'une *évaluation sociétale*. Elle porte sur la nature, l'importance, les incidences, les avantages relatifs d'un progrès technique, ainsi que sur les conséquences à court et à long termes pour l'économie, la politique et, plus généralement, pour la société globale.

"L'évaluation des choix technologiques", écrit Vinck (1988, p. 7), "implique de tenir compte des répercussions et des risques pour la société". Par exemple, à côté des grands services qu'elles rendent, les automobiles polluent et font perdre l'habitude de la marche à pied, bien utile pour la santé. "Mais", continue Vinck, "il est un autre risque, plus diffus : celui d'imposer progressivement une forme d'organisation sociale intolérable pour certains groupes." (Sur l'évaluation sociétale des choix technologiques, voir l'annexe).

□ Macropilotage - Micropilotage

• Le macropilotage

Tout pouvoir organisateur d'enseignement assigne des objectifs au système dont il prend la responsabilité et la charge. En démocratie, ce choix résulte d'une concertation entre toutes les parties intéressées, concertation qui s'appuie ou débouche sur une adhésion aux valeurs qui fondent les décisions. Les objectifs sont relatifs aux savoirs, aux savoir-faire et aux savoir-être.

Que l'on s'efforce d'établir dans quelle mesure les objectifs sont atteints et de prendre, à tous les niveaux, les mesures adéquates pour qu'il en soit ainsi, est non seulement un acte de raison, mais aussi un devoir civique : la communauté a droit qu'on lui rende des comptes à propos d'actions dont la réussite conditionne largement l'avenir des jeunes et de la société globale. En cas de carences ou d'insuffisances, les causes doivent en être recherchées et les remèdes doivent être appliqués.

Cette démarche générale fait l'objet du pilotage macroscopique. Il tend à se focaliser sur l'*input* et l'*output*, et est, pour cette raison, souvent qualifié de *technocratique*. On veut ainsi dire que la conception technique des problèmes éducatifs risque de se faire au détriment des conséquences humaines, sociales.

Pareille accusation peut être injustifiée. Par exemple, n'est-il pas capital d'établir quel pourcentage d'une population scolaire a acquis une maîtrise suffisante de l'habileté de lecture, sachant notamment que le niveau moyen de lisibilité des textes que l'on rencontre dans la presse quotidienne ou dans les entreprises est de x ou y . S'ils n'atteignent pas ce niveau d'habileté, les jeunes ont peu de chance de trouver un emploi, les entreprises manqueront de main-d'oeuvre qualifiée et les citoyens ne seront pas armés à suffisance pour jouer pleinement leur rôle dans la société.

La problématique est ainsi formulée de façon technique. Ce n'est pas une raison suffisante pour la condamner. Quelle que puisse être parfois la complexité de certaines démarches évaluatives, elles font aussi partie intégrante de la mission des éducateurs. Il importe, toutefois, que ceux-ci aient toujours leur mot à dire sur la conception et l'exécution des opérations, et sur l'interprétation des résultats.

• **Le micropilotage**

Niché dans le système éducatif global, chaque lieu d'éducation possède sa spécificité, sa personnalité, son environnement. Niché dans l'ensemble du corps enseignant, chaque maître est une personne unique, de même que l'élève est unique parmi ses pairs. On retrouve ainsi la célèbre phrase de Pop-ham : "La qualité de l'apprentissage que l'on pressent dans une situation d'enseignement donnée est fonction d'une méthode *particulière* employée par un professeur *particulier* pour des élèves *particuliers* poursuivant un but *particulier*".

Dans cette situation unique, l'élève est - par sa liberté - sujet au même titre que l'éducateur et, par sa propre cohérence, le maître comme ses élèves, articule constamment ses principes théoriques avec la façon dont il les vit dans le concret de son existence (Soëtard, 1983, p. 127).

Partant de là, V. de Landsheere (1992, p.3) écrit : "Si l'on admet que le principe premier de l'éducation est le respect de la personne, ni les lois psychologiques du développement de l'intelligence et de la personnalité, ni celles de la sociologie, ni la logique interne des branches du savoir ne résistent dans leur généralité dès qu'on franchit le seuil de la classe. Elles aident tout au plus l'éducateur à formuler sa propre loi *hic et nunc*". Reconnaître cette réalité implique-t-il la condamnation sans appel du pilotage macroscopique ? Assurément non.

• **Une complémentarité**

Le pilotage microscopique, qui s'opère à l'échelle d'une classe, voire d'un seul élève, n'est pas incompatible avec un projet éducatif global et, moins encore, avec un projet d'établissement. Il s'agit, en réalité, d'un em-

boitage de plus en plus fin. S'il n'est pas possible, c'est qu'il existe une défec-
tosité, une incohérence à au moins l'un des échelons du projet éducatif.

La méthode d'évaluation va évidemment différer selon les niveaux, car
à mesure que l'on affine l'observation, elle devient de plus en plus qualitative.
Les deux propositions suivantes sont, en effet, très éloignées l'une de l'autre :

- Selon les critères adoptés par l'Association nationale pour l'enseigne-
ment de la lecture, seulement 32 % des élèves de 14 ans peuvent être
considérés comme de bons lecteurs.
- A 14 ans, Paul n'est capable de lire que de courts textes, porteurs d'in-
formations purement factuelles, composés de phrases simples.

Il est clair que, pour remplir sa mission, l'enseignant a besoin d'infor-
mations dont la seconde proposition offre un exemple. Mais il est tout aussi
clair que c'est à des évaluations illustrées par la première proposition qu'un
ministre de l'éducation va recourir pour prendre une décision relative au pro-
gramme d'enseignement ou à la formation des maîtres.

La communauté des approches réside dans la volonté d'objectivation,
de rationalité et d'efficacité de l'oeuvre éducative, à quelque échelle que ce
soit, ce qui n'implique ni une quelconque mécanisation d'un processus, ni le
refus de la subjectivité, de l'intuition, de la construction individuelle du sens
que chacun confère ou croit conférer à ses actions. La préoccupation domi-
nante reste toujours le meilleur service possible à ceux qui s'éduquent.

Tout rapport de pilotage doit être un outil permettant d'éclairer adé-
quatement le débat pédagogique et le débat politique en matière d'action sco-
laire. "Si l'on n'articule pas ces deux dimensions, je craindrais qu'on aille à
l'échec ou, au moins à la non-pertinence : le débat politique qui ne se con-
fronte pas aux apports des pédagogues (comme à ceux des économistes et
d'autres) reste dans les nuages; tout comme le débat pédagogique, s'il ne s'ar-
ticule pas avec les manières dont une société négocie ses priorités et l'articu-
lation de ses institutions" (G. Fourez, communication personnelle).

Et, de même que la recherche en soi ne garantit pas le développe-
ment, de même il importe que les services que le pilotage peut rendre soient
correctement perçus. Croire que cette perception se produira du jour au len-
demain est une naïveté. Croire que les décideurs utiliseront la totalité des
données, même très sûres, mises à leur disposition, en est une autre.

□ Sur la valeur ajoutée

Une évaluation de *la valeur ajoutée* consiste à interpréter les résultats
scolaires (au sens large) en fonction de ce que les élèves savaient ou étaient
à l'entrée dans l'école ou dans la classe. On peut aussi tenir compte des dif-
férences d'aptitudes ainsi que de la qualité du soutien apporté par les familles.

C'est plus facile à dire qu'à faire. Il est notamment impossible de démontrer que, pour une discipline donnée, passer d'un score de réussite de 50 à 60 % équivaut à passer de 70 à 80 % ou de 85 à 95 %...

La solution adoptée en Californie est la suivante : toutes les écoles de l'Etat subissent un même test et les résultats d'une école particulière sont comparés aux résultats obtenus par l'ensemble de 10 % des écoles dont l'indice socio-économique est le plus proche de l'indice de l'école en question. Pareil indice n'explique évidemment pas tout; il constitue néanmoins un point de repère important.

Une innovation récente de la Direction de l'Evaluation et de la Prospective (DEP) du Ministère français de l'Education nationale s'inscrit dans la même perspective : les taux d'accès régionaux au niveau du baccalauréat, en 1991, ont été mis en relation avec "un taux d'accès attendu, compte tenu de la structure sociale de chaque région" (DEP, *L'état de l'école*, octobre 1992, p.23).

Sans sous-estimer l'intérêt des mises en relation qui viennent d'être évoquées, on ne peut toutefois dissimuler combien la réalité de "l'école efficace" est difficilement saisissable, même si l'on restreint le champ de l'efficacité à celui des apprentissages scolaires traditionnels.

Liensol et Meuret (1987), cités par A. Grisay (1988, p.5), ont étudié, dans 210 établissements secondaires des académies de Grenoble, Metz-Nancy et Toulouse, les résultats obtenus au baccalauréat par les élèves au cours de trois années consécutives. L'influence de l'origine socio-économique des élèves et de leurs retards scolaires antérieurs est manifeste.

Cependant, dans un même établissement, le taux de réussite est loin d'être stable. "Les performances d'un lycée sur trois peuvent être considérées comme radicalement différentes d'une année à l'autre. Il en va de même lorsque l'on mesure les performances avec le taux de réussite au baccalauréat." Il importe donc de définir de façon plus rigoureuse le concept d'efficacité ou de rendement des écoles.

Le bilan qu'A. Grisay établit des recherches quantitatives qui s'inscrivent dans la ligne du Rapport Coleman (recherches sur la relation entre l'*input* et l'*output*) et, aussi, des recherches plus qualitatives relevant du *modèle écologique*, est révélateur de la difficulté de l'entreprise et de la pauvreté des connaissances sûres dont on dispose actuellement sur le sujet.

L'exaspérant des résultats des recherches quantitatives *input-output* sur les rendements scolaires, conclut A. Grisay (p.17), "vient de ce que les différences de rendement entre écoles (indépendantes du recrutement, et donc attribuables, en principe, à la qualité de l'environnement scolaire) *existent* sans qu'on puisse en rendre compte (ou si peu) par des différences objectives de <traitements> suivis par les élèves. Car les observations sont probantes : *il existe des écoles plus performantes que d'autres.*"

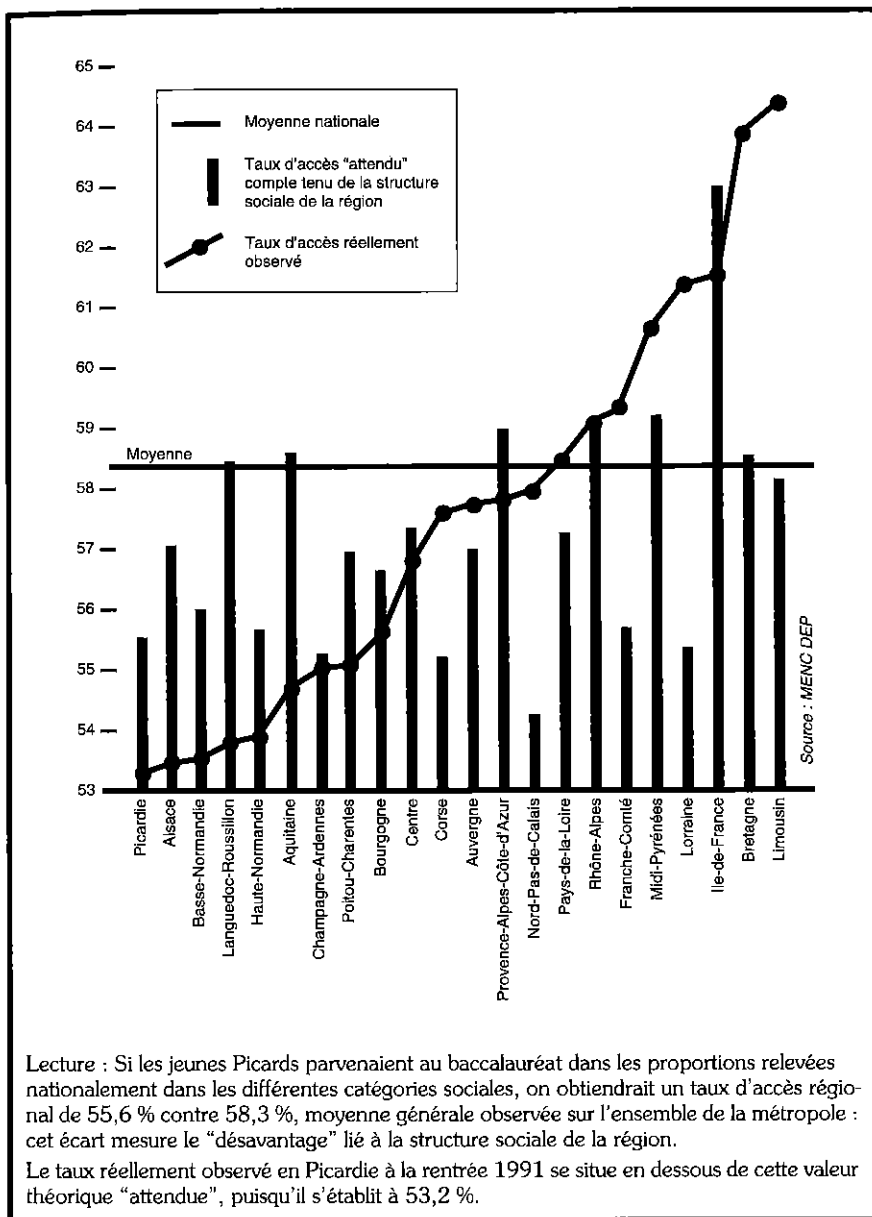


Figure 1 : Taux d'accès régionaux au niveau du baccalauréat (1991)
Effet de la structure sociale

Les conclusions générales de ce premier type de recherches sont (p.36) :

- La "bonne école" ne se reconnaît pas au premier coup d'oeil. Il faut, en particulier, se garder de confondre "bonnes écoles" et écoles "bien fréquentées". Malgré leurs résultats élevés, ces dernières ne sont pas nécessairement les plus performantes.
- Le bon rendement d'une école est principalement lié à la manière dont le temps d'apprentissage est exploité.
- "Il n'est pas évident qu'il existe un <profil> d'école performante qui soit partout et toujours valable. La réalité est plus complexe : il existe probablement des configurations de caractéristiques scolaires qui donnent de bons résultats avec *certaines* populations scolaires, et pour *certaines* matières, et de moins bons avec d'autres élèves ou pour d'autres domaines du savoir..."

Le bilan des recherches du type écologique n'est pas, non plus, évident.

La démarche générale consiste principalement en une observation ethnométhodologique d'établissements réputés soit performants, soit non performants : on tente de repérer, par une observation approfondie et systématique, les processus (sociaux, organisationnels, pédagogiques,...) qui peuvent expliquer l'efficacité ou la non-efficacité.

Des facteurs tels que la qualité de la discipline, les attentes positives vis-à-vis des élèves, la qualité du leadership exercé par le chef d'établissement comptent nettement parmi les plus importants. Mais les conclusions restent très prudentes : "Les résultats actuels ne nous disent pas encore ce que sont les écoles performantes. Ils nous fournissent seulement un certain nombre d'indications sur les facteurs associés à une élévation du taux de réussite à des tests standardisés" (p.57). C'est donc la validité même des critères qui peut-être mise en cause, car non seulement les effets cognitifs doivent être mieux explorés, mais l'aspect affectif doit aussi retenir beaucoup plus l'attention qu'on ne le fait actuellement.

La conclusion générale rejoint celle qui a été tirée à propos des recherches *input-output* : il faut explorer de manière plus approfondie les raisons qui rendent *certaines* styles d'écoles, dans certains contextes, plus performants que d'autres.

L'immense complexité de l'évaluation des effets de l'école se confirme donc une fois de plus. Devant pareille difficulté, ou bien on baisse les bras et l'on renonce à tout effort d'objectivation, ou bien on fait preuve du "courage de l'imperfection" : dans les sciences de l'homme, la complexité et la diversité sont telles qu'on ne travaille jamais que par grandes approximations. Le tout est d'en être conscient en permanence et de tirer le maximum de profit et d'enseignements du peu que l'on connaît.

□ Cohérence, qualité, équité

On peut s'efforcer de contrôler la productivité en savoirs, savoir-faire et savoir-être d'un système éducatif, soit indépendamment des objectifs du programme (on dresse le bilan de tous les acquis identifiables), soit en liaison avec les objectifs principaux que l'enseignement s'est explicitement donné pour mission d'atteindre.

C'est dans cette seconde hypothèse que le pilotage s'indique d'abord. On veut savoir si l'on avance bien dans la direction choisie et, à cette fin, on prélève des indicateurs qui deviendront éventuellement des clignotants si des problèmes particuliers se posent. La répétition périodique des évaluations permet de distinguer des évolutions et des tendances.

Le pilotage devrait établir dans quelle mesure le système éducatif répond à **trois exigences : la cohérence, la qualité et l'équité.**

La cohérence existe :

- si les *objectifs* poursuivis s'inscrivent effectivement, soit dans la ligne du projet éducatif global, démocratiquement formulé et adopté (macropilotage), soit dans la ligne de projets particuliers, formulés par des groupes distincts, voire par des individus (micropilotage);
- si le *curriculum implanté* - c'est-à-dire, ce qui est effectivement enseigné et appris - correspond aux objectifs fixés;
- si la nature et les modalités des *évaluations* sont fonction de ces mêmes objectifs.

La qualité peut être évaluée de façon absolue (L'éducation permet-elle un développement optimum de l'individu ?) ou relative (Nos élèves sont-ils aussi bien formés que les autres ?).

Le critère d'*équité* exige notamment que les ressources disponibles soient également accessibles à tous les élèves, ce qui ne signifie nullement qu'ils doivent tous en tirer parti de la même manière et en même quantité (*discrimination positive*). Il convient, en particulier, d'établir si la qualité et le niveau des acquis ne sont pas liés davantage à l'établissement fréquenté qu'aux capacités réelles de ceux qui le fréquentent. En démocratie, le système éducatif a pour idéal l'égalité de réussite, étant entendu que les modalités de celle-ci peuvent varier selon les aptitudes, les besoins et les intérêts de chacun.

Par exemple, l'un des aspects de l'équité à assurer concerne l'égalité des sexes. Les recherches évaluatives à ce propos portent notamment sur les aspects suivants :

- Tendances générales des performances des filles et comparaison avec celles des garçons.

- Performances comparatives des filles et des garçons en mathématiques et en sciences.
- Différences dans le choix des études.
- Orientation professionnelle des filles.
- Pourcentage des filles et des garçons qui font des études post-secondaires.
- Différences d'aspirations.

Dans l'*Etat de l'école*, (2, 1992, p.39), la Direction de l'Évaluation et de la Prospective, du Ministère français de l'éducation, donne à ce sujet les informations suivantes:

	Ensemble	10 % les plus faibles	10 % les plus forts
Français	60,6	24,4	85,7
Mathématiques	70,1	35,9	89,7

Lecture : en moyenne, les élèves ont répondu correctement à 60,6 % des questions de français.

Figure 2 a : Pourcentage de réussite aux questions de français et de mathématiques à l'entrée en 6e (septembre 1991)

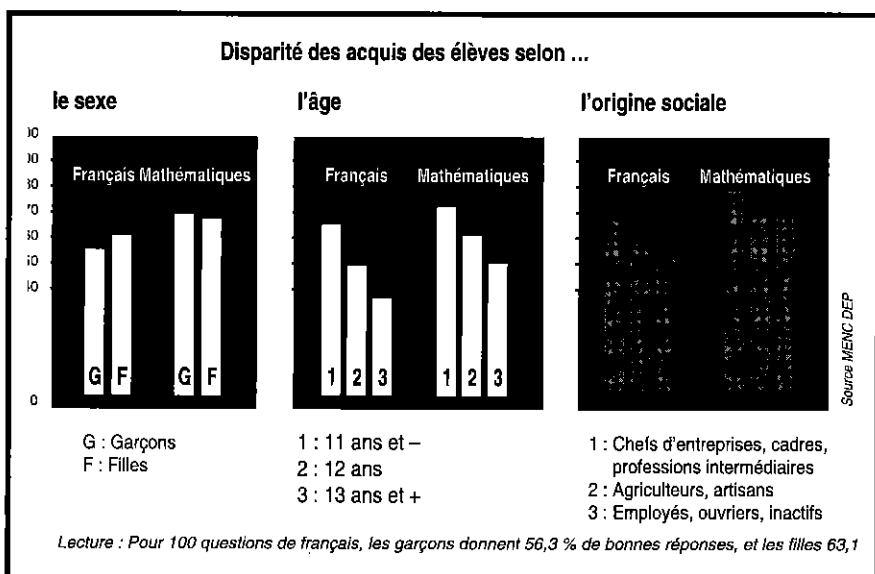


Figure 2 b : Disparité des acquis des élèves selon le sexe, l'âge et l'origine sociale

Bien des questions appellent donc des réponses claires et fondées. Qu'apprennent réellement les élèves ? Le système éducatif est-il équitable ? Les programmes répondent-ils aux exigences du monde d'aujourd'hui et de demain ? Les méthodes d'enseignement sont-elles adéquates ? Ces questions sont évoquées dans le tableau suivant.

• Récapitulation des questions

	Cohérence	Qualité absolue	Qualité relative	Équité
OBJECTIFS DES PROGRAMMES	L'enseignement répond-il aux besoins des élèves ?	Les programmes sont-ils assez, mais pas trop exigeants ?	Les élèves sont-ils aussi bien formés que ceux d'autres pays ?	Les programmes sont-ils conçus pour assurer l'égalité des chances, les discriminations positives ?
PRATIQUES RÉELLES DANS LES CLASSES	Correspondent-elles aux objectifs poursuivis ?	Le curriculum implanté est-il valide ?	Quelles pratiques s'avèrent plus efficaces que d'autres ? Quelles leçons tirer de ce qui se passe en classe ?	L'enseignement n'est-il pas biaisé en fonction des origines sociales ou ethniques, du sexe, de l'implantation géographique ?
ACQUIS DES ÉLÈVES	Le profil de compétences obtenu est-il le profil souhaité ?	Les acquis sont-ils suffisants pour assurer l'épanouissement personnel, pour exercer la citoyenneté responsable ?	Comment se situe : - l'élève moyen - l'élève du groupe supérieur, par rapport à ceux d'autres pays ?	Observe-t-on d'importantes différences entre les catégories ci-dessus ? Quels sont les taux d'échecs ?
EVOLUTION	En perçoit-on une ?	Le niveau du rendement baisse-t-il ?	Le système évolue-t-il mieux ou moins bien que d'autres ?	Le système évolue-t-il vers plus de démocratie ?

Selon D.H. Hargreaves (1988), le pilotage ne peut réussir que s'il se conforme à quatre conditions :

1. Être crédible non seulement pour le monde enseignant, mais aussi pour les autres membres de la communauté éducative.
2. Respecter l'anonymat quand cela s'indique. C'est pourquoi ceux qui sont chargés des opérations doivent jouir d'un certain degré d'indépendance vis-à-vis des autorités.

3. Combiner deux fonctions : celle du pilotage et celle du développement. Il faut, en effet, qu'on ne se contente pas de constats, mais, qu'à partir de là, on améliore réellement le système.
4. Apporter encouragement et soutien aux enseignants et aux chefs d'établissement.

□ Définir les buts de l'éducation et s'accorder à leur propos

Toutes les considérations qui précèdent reposent sur une condition première : avoir clairement défini les buts et les objectifs, et s'être accordé à leur propos.

Or, par exemple, les agences auxquelles les opérations de pilotage ont été confiées aux Etats-Unis ont, en vain, tenté de mettre d'accord les enseignants, les spécialistes des disciplines, les représentants des parents et les partenaires sociaux, sur les buts de l'éducation. On a parfois accusé les responsables du pilotage de vouloir produire un consensus qui n'existe pas.

De telles divergences existent plus ou moins partout. Elles ne portent pas seulement sur les fins, mais aussi sur les buts généraux des matières enseignées : les mathématiques seront-elles étudiées pour résoudre des problèmes de la vie quotidienne ou pour elles-mêmes ? Quelle est l'importance relative des apprentissages proposés à l'école ? Tient-on pour accord un minimum de désaccord, se demande V. de Landsheere (1988, p.231).

Ce phénomène ne doit pas être considéré comme anormal, comme une faiblesse. Tous les êtres humains sont différents. Hormis les besoins primaires liés à la survie, ils n'adoptent pas les mêmes valeurs, n'éprouvent pas les mêmes besoins, ce qui provoque une grande variation des objectifs poursuivis à un moment donné.

Cette constatation a valeur de mise en garde, mais ne doit cependant pas empêcher l'action éducative collective qui, quoi que l'on fasse, est inévitablement mal taillée si l'on raisonne dans l'absolu. Le tout est de s'efforcer de focaliser le pilotage sur une sorte de plus grand commun dénominateur culturel. Et donc d'être bien conscient des limites de l'entreprise. Il en va ainsi de toutes les actions importantes qui impliquent l'homme. Le dénominateur commun doit être prioritairement recherché au niveau des valeurs.

Une chose reste sûre, il n'appartient pas à l'école seule de fixer les fins et les buts de l'éducation. C'est à la communauté éducative globale - à la limite, elle se confond avec la société - qu'il appartient de les déterminer lors de débats démocratiques *réels*. J.-E. Charlier (communication personnelle) conclut que la question majeure se situe donc en amont de la mise en place

du pilotage; il faut d'abord être capable d'élaborer un projet cohérent pour l'enseignement. L'expérience montre que c'est loin d'être facile. En effet, les conditions économiques, politiques, culturelles et sociales actuelles rendent pratiquement impossible l'émergence d'un projet éducatif pleinement acceptable pour tous. Il faut, toutefois, se méfier d'une sorte de perfectionnisme qui empêche d'avancer, et donc pouvoir accepter des actions qui ne sont que partiellement satisfaisantes, pour autant qu'elles s'inscrivent dans ce qui semble être la bonne direction.

Sur un plan plus technique, on observe que la formulation de certains buts fait référence à un critère externe de réussite. Par exemple, l'un des six buts que le président Bush voulait que son pays atteigne était : "Les performances des élèves des Etats-Unis, en sciences et en mathématiques, doivent être les meilleures du monde en l'an 2000." Autre exemple, en Angleterre : "La main-d'oeuvre britannique doit posséder les connaissances et les compétences qui lui permettent d'être compétitive sur les marchés européens et mondiaux" (*National Commission of Education*, 1990).

Comme le relève T.N. Postlethwaite (1993), "Dans le passé, les buts de l'éducation étaient formulés dans un contexte national, sans référence comparative explicite à des standards nationaux ou internationaux? (...) En revanche, depuis le milieu des années 80, on observe une tendance à formuler les buts en termes de cibles à atteindre." On fait de même dans les entreprises.

□ Développer les services de statistiques

Pour piloter le système éducatif et rendre des comptes à la nation, la collecte régulière de données et de statistiques aux niveaux local, régional, national et international est indispensable. Ces statistiques doivent porter non seulement sur les ressources et, plus généralement, sur tout ce qui entre dans le système, mais aussi sur la façon dont le système fonctionne (processus) et sur ce qu'il produit.

Pratiquement tous les pays disposent de services de statistiques de l'éducation plus ou moins développés. Toutefois, les données recueillies varient considérablement tant dans leur nature que dans leur qualité. Ainsi s'explique notamment les difficultés considérables qu'éprouvent les grandes organisations internationales telles que l'UNESCO ou l'OCDE pour réunir des statistiques suffisantes, fiables et comparables. D'énormes efforts sont actuellement consentis pour améliorer cette situation.

□ Piloter le pilotage...

Le pilotage est une technique : il s'agit, en effet, d'un procédé appliqué systématiquement pour obtenir un résultat déterminé, en l'occurrence l'optimisation de l'action éducative.

La conception même de l'optimum soulève une question cruciale : qui va le définir et comment ? Autocratiquement ou démocratiquement ? Faut-il dire que nous optons résolument pour le second parti, ce qui implique la maîtrise sociale du pilotage.

Pour piloter un système éducatif, il faut le maintenir sur une trajectoire et donc réduire sa liberté en le soumettant à une norme permettant de déterminer s'il s'écarte du chemin tracé ou si le but est effectivement atteint.

Toutefois, les standards et les normes peuvent être réducteurs et freiner l'innovation. C'est particulièrement vrai dans le domaine de l'éducation, car l'apprentissage d'un même savoir, savoir-faire ou savoir-être peut se faire selon bien des voies et des modalités différentes. Par ailleurs, si le pilotage sert à conduire bon gré, mal gré vers un objectif autoritairement fixé, la norme, le standard deviennent "une arme redoutable, un instrument de vassalisation et d'exclusion" (Gonod, 1992, p.8). C'est pourquoi les normes et les standards aussi devraient toujours faire l'objet d'une négociation sociale.

Or, comme on le verra, la réalisation d'un pilotage - surtout si l'approche est macroscopique - soulève des problèmes de grande technicité, face auxquels beaucoup d'enseignants ne sont pas armés. "La démocratisation du contrôle postule la démocratie participative couplée avec l'émergence d'une culture technique de masse par la démocratisation des connaissances et de l'information technologique" (Gonod, p.8). On se retrouve ainsi, une fois de plus, devant la nécessité d'élever le niveau scientifique de la formation des enseignants.

Par ailleurs, la politique et la pratique du pilotage doivent faire l'objet d'une discussion permanente entre toutes les parties intéressées. On trouvera, dans l'annexe, une évocation de différentes méthodes de prévention des dangers que peut représenter une application incontrôlée des technologies et une évocation des méthodes de recherche de consensus à leur propos.

u pilotage

appliqué
de l'opti-

ale : qui
? Faut-il
e la maî-

ajectoire
le déter-
int.

s et frei-
ucation,
t se faire
pilotage
, la nor-
vassali-
es et les
iale.

l'appro-
ité, face
ation du
ce d'une
ces et de
e fois de
n des en-

re l'objet
n trouve-
tion des
nologies
opos.

Les indica

□ La notion d'indicateurs

En raison de l'usage considérable que le pilotage en fait, il importe de préciser la notion d'indicateur. V. de Landsheere (1992, p.478) écrit :

“Mesure destinée à servir le pilotage, un indicateur est une statistique directe et valide informant sur l'état et les changements d'ampleur et de nature, au cours du temps, d'un phénomène sociétal jugé important. En éducation, cette mesure informe notamment sur la santé et la qualité du fonctionnement du système, sur ce que connaissent les élèves, ce qu'ils sont capables de faire, et sur l'évolution positive ou négative de ces conditions (d'où l'intérêt à accorder à l'observation de caractéristiques durables) et sur des différences significatives qui peuvent exister entre des aires géographiques ou des institutions à un moment donné. Un système d'indicateurs devrait aussi informer sur la façon selon laquelle les indicateurs mesurés conjuguent leur influence pour produire l'état du système éducatif et l'évolution de celui-ci.”

Une *donnée* est une information qualitative ou quantitative, statistiquement traitée et utile pour la prise de décision. Une *statistique* est une donnée quantitative. Les indicateurs sont des statistiques (résultats à des tests, taux d'abandon scolaire,...). Les systèmes de pilotage peuvent utiliser des données qualitatives et des statistiques.

Installer un système d'indicateurs de l'éducation est une entreprise ambitieuse. N'en attend-on pas plus qu'il ne peut donner, vu l'état de nos connaissances actuelles ? De fait, on est loin de savoir de façon sûre comment les composantes décisives d'un macrosystème ou d'un microsystème éducatif combinent leurs effets pour produire les résultats observés. Ce n'est cependant pas une raison suffisante pour ne rien faire, en particulier, pour renoncer à mettre en place des systèmes d'indicateurs susceptibles d'aider à décrire et, si possible, à expliquer, ne fût-ce que partiellement, ce phénomène. Aussi importe-t-il que, parallèlement à cet effort, on engage des recherches, d'une part, sur la façon de produire des indicateurs individuels tels que les mesures de performances intellectuelles d'élèves, en train d'accomplir des tâches de résolution de problèmes, et, d'autre part, sur la nature des composantes principales du système éducatif qui combinent leurs effets pour produire l'état du système (Odden, 1990, p. 34).

Dans de bonnes conditions, les indicateurs sont le reflet de la politique éducative et mettent éventuellement sur la voie de modifications à y apporter. Ils doivent être robustes, c'est-à-dire ne pas être trop sensibles à des fluctuations accidentelles. Il faut donc qu'ils se rapportent à des caractéristiques du système éducatif qui s'y manifestent de façon générale et durable. Si leur signification est facilement comprise par tous les membres de la communauté

éducative, le dialogue entre les politiques, les citoyens et les enseignants est facilité.

Un indicateur peut servir de clignotant dont il faut rechercher les causes. Il attire l'attention et invite à l'action (pilotage). Il est aussi instrument de communication.

Il est, par exemple, intéressant de suivre certains phénomènes éducatifs à travers le temps et de tenter de comprendre des modifications marquantes qui peuvent se manifester. Ainsi, le graphique suivant montre l'évolution du pourcentage d'échecs scolaires, pendant une période de quatorze ans, dans l'enseignement secondaire de la Communauté Flamande de Belgique (De Groof *et al.*, 1993, p.115).

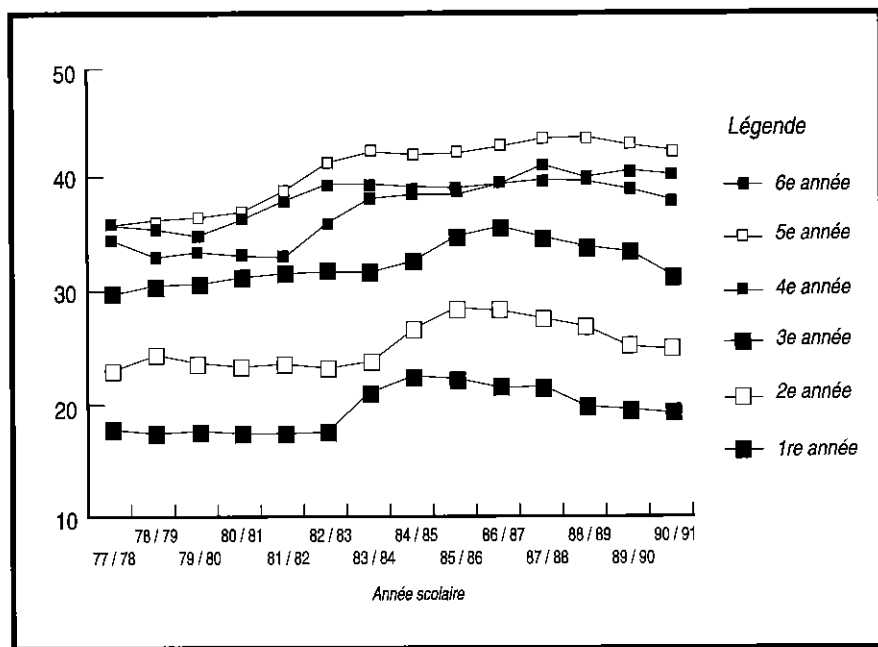


Figure 3 : Evolution du pourcentage d'élèves en retard scolaire dans l'enseignement secondaire 1977/78 - 1990/91

Ce graphique révèle deux tendances. D'une part, on constate que le pourcentage de retards scolaires s'accroît en 5e et 6e années, surtout à partir de 1981-82. C'est vraisemblablement dû au fait que moins d'élèves abandonnent prématurément l'enseignement. D'autre part, à partir de 1983-84, une augmentation des échecs se produit en première année; la situation s'améliore un peu dans les années supérieures, puis se déprécie de nouveau. Ces phénomènes peuvent être attribués à la prolongation de la scolarité obligatoire.

nants est

les cau-
ment des éduca-
marquan-
évolution
force ans,
Belgique

de

5e année

5e année

4e année

3e année

2e année

1re année

e le

out

d'é-

t, à

niè-

res,

ués

Toute statistique relative à l'éducation n'a pas valeur d'indicateur. Ainsi, connaître le nombre d'écoles que compte une région n'apprend rien sur la qualité de ce qui s'y fait. Un tel chiffre n'acquiert une signification que si un point de référence permet de l'interpréter : évolution d'une même statistique, comparaison avec des mesures similaires opérées ailleurs, normes définies par la société, ... Le plus important est de disposer du moyen d'être correctement informé sur la qualité de l'éducation.

Concevoir des indicateurs est donc une opération où interviennent des considérations philosophiques (jugements de valeur), politiques et techniques. On attend qu'ils soient importants, valides, faciles à prélever, crédibles, et mettent sur la voie d'améliorations.

Piloter, c'est plus qu'accumuler des indicateurs. Le pilotage a trois composantes nécessaires : la collecte régulière d'informations, l'évaluation de ces informations et, surtout, la traduction en actions institutionnelles ou en sanctions, c'est-à-dire en approbations ou désapprobations. Par exemple, on conclura que tel objectif a été ou non atteint de façon satisfaisante. Bref, alors que les statistiques habituellement publiées sur l'éducation apportent des informations quantitatives (budget, nombre de diplômés, etc.), les indicateurs aident à mieux percevoir le fonctionnement qualitatif des systèmes et leurs performances.

Selon Oakes (1986, p.496) les indicateurs doivent fournir au moins l'une des informations suivantes :

- Une description des performances du système éducatif qui permettent d'atteindre les buts poursuivis.
- Une information sur les caractéristiques liées aux résultats (ceci étant établi par la recherche expérimentale).
- Une description de caractéristiques fondamentales du système. Elle doit permettre une meilleure compréhension du fonctionnement.
- Une information à propos de problèmes qui se posent.
- De l'information importante pour la politique éducative, notamment pour la décision de réformes.
- Une description de caractéristiques générales, c'est-à-dire qui peuvent être observées sous une forme ou sous une autre dans tout le système, et qui peuvent être comparées dans divers contextes.
- Des informations relatives à des caractéristiques permanentes de l'ensemble ou de sous-ensembles du système, formulées de façon à permettre la comparaison à travers le temps.

De leur côté, Kaagan et Smith (1985) soulignent le rôle important que les indicateurs peuvent jouer dans le pilotage des réformes :

- Indiquer les changements d'aspects importants comme la qualité de l'enseignement et des performances des élèves, et ainsi attirer l'attention sur des problèmes éventuels.
- Evaluer l'influence des efforts de réforme.
- Encourager les efforts par la comparaison avec d'autres pays ou entités politiques.
- Attirer l'attention sur des établissements qui doivent être améliorés.
- Servir de catalyseurs d'idées nouvelles.

Pour couvrir le système éducatif, les informations doivent porter :

- sur ce qui entre dans le système (*input*) : ressources, qualité des élèves, des professeurs, etc.
- sur les processus éducatifs (*treatment*) : pédagogie pratiquée, modes d'examens, etc.
- sur les "produits", ce qui sort du système (*output*) : acquis scolaires, nombre de diplômés, etc.

Ces trois composantes entretiennent des relations étroites et multidirectionnelles. Il importe donc d'examiner comment des indicateurs s'associent pour expliquer un effet global.

En cette matière, comme dans toutes celles qui concernent l'homme, on n'a jamais affaire à une science intégralement exacte. Les indicateurs sont basés, en partie sur les connaissances scientifiques, en partie sur le bon sens, en partie sur les préoccupations politiques. Aussi, les indicateurs sont-ils loin d'être à l'abri de la critique. On leur reproche souvent de n'apporter qu'un ensemble disparate d'informations. L'exemple des indicateurs relatifs à l'économie montre pourtant combien ils peuvent être utiles, malgré cette imperfection. Les faiblesses furent bien plus graves encore dans le passé lorsque l'on tenta de définir des indicateurs sociaux.

En effet, bien avant que le mouvement relatif aux indicateurs de l'éducation ne prenne son essor actuel, **les indicateurs sociaux** destinés à évaluer les effets de la prospérité économique des années 60 sur la qualité de la vie et sur les progrès de l'équité ont suscité un intérêt considérable. On voulait piloter l'évolution socio-économique. L'ouvrage de R. Bauer, *Social Indicators* (1966) est à cet égard la référence principale. L'OCDE s'est très tôt intéressée à ce mouvement dans l'espoir de mesurer l'effet de l'éducation sur la société. C'était la grande époque de la recherche quantitative en sociologie, remarquablement illustrée par R. Boudon (*La mathématique en sociologie*, 1971) et par J. S. Coleman et ses associés, dans leur célèbre rapport (1966) sur l'égalité des chances en éducation.

La désillusion succéda rapidement à l'enthousiasme. Des indicateurs nationaux globaux noyaient les particularités de bien des situations et de bien des effets, ce qui n'aidait pas à trouver les remèdes désirés. Par ailleurs, les

relations causales restaient trop au niveau de la corrélation et non de la compréhension. De plus, un accord sur la définition des principaux concepts était loin de régner. Ni la planification, ni les décisions ne pouvaient se fonder sur des bases aussi peu sûres.

De l'expérience acquise pendant cette première époque, T.J. Wyatt (1988, p. 5) tire cinq grandes leçons :

1. Les indicateurs doivent être définis de façon simple et opératoire.
2. Il faut éviter de fondre des mesures disparates en un indicateur global qui voudrait représenter le produit de l'éducation.
3. Les indicateurs peuvent révéler des problèmes, mais ne fournissent pas les solutions.
4. L'interprétation des indicateurs est au moins aussi importante que leur construction. Pour se faire une idée d'ensemble d'une situation, il faut apprendre à articuler les indicateurs tout en examinant les tendances de chacun à travers le temps.
5. Il importe que les indicateurs aident à formuler la politique éducative.

On peut craindre, par ailleurs, que les indicateurs n'exercent un *effet de reflux* sur l'enseignement. Nous avons déjà fait allusion à ce phénomène. Contentons-nous donc de rappeler avec Porter (1988, p. 504), et de nouveau à titre exemplatif, que : "S'il existe de bons indicateurs relatifs à l'enseignement des mathématiques et des sciences, et pas pour les branches littéraires, les écoles auront probablement tendance à s'attacher plus aux mathématiques et aux sciences qu'au reste. Si les indicateurs sont relatifs à des habiletés considérées séparément et à la connaissance des faits, plutôt qu'à la compréhension profonde et à la résolution de problèmes, l'école risque de s'orienter dans le même sens."

De fait, les indicateurs peuvent devenir plus que des informations relatives à la santé du système scolaire; ils peuvent devenir des définitions de ce qu'est la bonne santé.

Il appartient à la communauté éducative de veiller à ce que pareils errements ne se produisent pas.

□ Nécessité d'un modèle

Un modèle cohérent de fonctionnement du système éducatif est nécessaire pour déterminer quels indicateurs collecter. Le choix de ce modèle détermine largement les informations que le système d'indicateurs fournira. Dans le chapitre précédent, c'est le modèle systémique *entrée - traitement - sortie - rétroaction* qui a été utilisé. Le modèle général adopté par l'OCDE

(CERI) s'inscrit dans cette ligne : dans un environnement donné, on étudie les *ressources*, les *processus* et les *effets*.

Les modèles sont toujours biaisés en faveur d'une perspective épistémologique particulière (économique, sociologique,...) relative au système éducatif. Ainsi s'explique qu'aucun modèle général du processus éducatif n'est unanimement accepté.

Scheerens (1989, cité par Wyatt, 1991, p.10) distingue trois tendances :

1. Le passage de statistiques descriptives concernant l'*input* (ressources,...) à des ensembles d'indicateurs relatifs aux performances ou aux résultats (*output*). Cette tendance répond aux vœux des décideurs politiques qui désirent savoir ce que produit le système et souhaitent que, plus que par le passé, les écoles soient amenées à rendre des comptes sur les objectifs qu'elles atteignent.
2. Une tendance à privilégier des systèmes d'indicateurs plus larges, plus centrés sur les résultats, et des mesures de contexte, tendance accompagnée d'un intérêt croissant pour les facteurs d'*input* et pour les caractéristiques des processus malléables.

Les décideurs ne veulent pas seulement savoir ce qui se passe, mais ils veulent aussi savoir pourquoi il en est ainsi et ce qu'ils peuvent faire pour remédier aux faiblesses observées. On accorde de plus en plus d'importance à ce qui se produit réellement dans les écoles (donc insistance sur le "traitement" et les résultats). On refuse de continuer à laisser la plupart des aspects du traitement dans une "boîte noire"; on veut en percer le secret, en un premier temps sans théorie préexistante. Il s'agit d'abord d'ouvrir les yeux sur la réalité des processus.

3. Une tendance à considérer la totalité du système : *input* - traitement ou processus - *output*.

Ce troisième modèle est le plus utilisé. Il est raffiné de façon intéressante par R.R. Herriott [1979, cité par Van Herpen (1992, p. 40)], qui ne se focalise pas uniquement sur le processus, mais tient aussi compte des aspects sociologiques, de la structure et de la culture, ainsi que des caractéristiques de l'environnement.

Herriott distingue six composantes :

1. *L'environnement*. Les objets et les formes de relations qui existent à tous les niveaux, en dehors du système éducatif. Celui-ci ou bien influence l'environnement ou est influencé par lui.
2. *La structure*. Patterns relativement stables d'interactions sociales qui intègrent les divers éléments d'un système éducatif, à tous les niveaux.

3. *La culture.* Les sentiments, le sens, les engagements que les différents acteurs (enseignants, élèves, membres de l'administration,...) attachent à leur vécu et aux buts, aux activités et aux problèmes du système éducatif.
4. *L'output.* Les connaissances, les habiletés et les orientations exigées par l'environnement et possédées par les élèves lorsqu'ils quittent le système éducatif à un moment quelconque.
5. *Le throughput.* Les "matériaux" bruts (fondamentalement les élèves) qui passent dans le système et sur lesquels le système opère pour produire un *output* adéquat à l'environnement.
6. *L'input.* Les instruments (manuels scolaires,...), le personnel (professeurs, services, membres de l'administration) et les informations (connaissances et valeurs) qu'un système éducatif importe de son environnement et utilise pour transformer le *throughput* en *output*.

□ Nombre d'indicateurs souhaitable ?

Il n'existe pas de réponse toute faite à cette question.

On vient de le voir, de multiples indicateurs concernent chaque aspect important de l'éducation, et il n'est pas aisé de décider combien il faut en prélever à propos de chacun d'eux. Certes, moins les indicateurs sont nombreux, mieux ils attirent l'attention, mais cette réduction peut nuire à la validité : l'ensemble peut alors ne pas être représentatif du système évalué. Or cette représentativité est nécessaire, car un indicateur n'est pas simplement une mesure de lui-même; il joue le rôle d'une variable résumant un concept plus large.

Cela étant dit, il ne faut jamais perdre de vue que les indicateurs "ne disent pas tout à propos de l'éducation. Ils ne permettent que d'entrevoir les grandes lignes d'une situation. (...) Leur degré de généralité peut dissimuler des situations très différentes parmi les sous-groupes. C'est pourquoi les indicateurs devraient pouvoir être décomposés afin de permettre l'examen de sous-populations." (T. Wyatt, 1991, p.9).

En dernière analyse, la réponse à la question doit être dictée par l'objectif que l'on poursuit. En particulier, selon que l'on désire faire du sensationnalisme ou permettre d'évaluer une situation de façon sereine et très informée, la réponse différera...

□ **Qualités souhaitées**

Bien qu'elles aient été spécialement souhaitées pour les indicateurs centrés sur l'école, les qualités suivantes concernent aussi la plupart des autres indicateurs (Gray et Jesson, 1968, cités par Nuttall, 1991):

- Evaluer directement les performances scolaires.
- Etre avant tout focalisés sur les processus d'enseignement et d'apprentissage.
- Etre quantifiables, ce qui ne veut pas dire strictement mesurables.
- Etre unidimensionnels (éviter les chevauchements).
- Couvrir l'ensemble des objectifs prioritaires.
- Porter sur les aspects majeurs du système éducatif, modifiables par des changements de politique.
- Vérifiables.
- Susceptibles d'être mis en relation avec d'autres.
- Aussi économiques que possible. La collecte doit être faisable en un temps, à un prix et avec un niveau d'expertise raisonnables.
- Etre fournis en temps utile (notamment pour la prise de décision).
- Apporter des informations relatives à des problèmes.
- Etre aisément compréhensibles par un large public.
- La validité et la fidélité des indicateurs doivent être reconnus par tous.

□ **Directives pour une utilisation constructive**

L. Darling-Hammon (1991, p. 23 sq.) donnent les indications suivantes :

1. Les chiffres ne suffisent pas. Une connaissance de l'enseignement, de l'apprentissage et du fonctionnement des institutions éducatives doivent permettre, à toutes les parties intéressées, de les interpréter correctement.

Ce n'est pas parce que les chiffres sont précis qu'ils sont valides.

2. Seule une multiplicité d'indicateurs peut donner une image correcte d'une situation. Les aspects de l'éducation qui sont les cibles directes de la politique éducative ne sont pas les seuls intéressants. Il importe, en particulier, de tenir compte des caractéristiques durables du système éducatif.

3. Les systèmes d'indicateurs doivent être redondants et continuellement révisés. Le danger de corruption d'un indicateur est réduit si l'on procède à des mesures multiples d'un même aspect du système.
4. Qu'il y ait des indicateurs ou non, une politique doit être basée sur une saine compréhension de la nature des motivations individuelles et des changements organisationnels, et de l'influence de ces facteurs sur les performances.
5. Les indicateurs doivent être utilisés comme de l'eau au moulin de l'évaluation d'une situation, et non comme des faits suffisants pour déclencher des actions qui pourraient être inefficaces. L'intérêt des indicateurs, c'est qu'ils donnent des signaux à propos de questions importantes et de problèmes qui pourraient se poser. Dans ces cas, un examen approfondi s'impose.

Cela étant dit, il importe de ne pas falsifier la signification des indicateurs. Par exemple, des indicateurs de performance peu favorables ne signifient pas *ipso facto* qu'un programme est mauvais.

Par ailleurs, étant donné la signification que les indicateurs peuvent prendre pour les responsables politiques, pour la réputation des établissements, pour certains financements, pour le succès de certains programmes, etc., toutes les précautions doivent être prises pour assurer l'intégrité, la qualité et la neutralité des analyses interprétatives des indicateurs.

L'idéal est de recourir à un organisme ou une entité apolitique non liés à des groupes de pression, institutionnellement crédibles, bien au courant des problèmes d'éducation et de gestion, méthodologiquement compétents, et capables de synthétiser les informations et de les diffuser parmi les chercheurs, les spécialistes de l'évaluation, les enseignants, les fonctionnaires "généralistes", les médias et le public (Guthrie, 1989).

L'expérience montre que l'on peut trouver une solution à ce problème épineux en se tournant vers des services universitaires spécialisés en matière de recherche expérimentale en éducation et en évaluation. Toutefois, il est impératif que ces services appuient leurs actes techniques sur une réflexion permanente relative aux options éducatives et aux débats politiques qui les concernent. Il faut, ici aussi, tenir compte des débats sur les objectifs culturels de l'école : que lui demande-t-on en matière d'éducation scientifique, historique, éthique, esthétique, sociale, civique ?... "Cette dimension d'étude 'socio-politique' du pilotage dépasse le cadre du paradigme actuel de la pédagogie, voire du paradigme dominant des sciences de l'éducation. Et pourtant, un pilotage qui négligerait ces dimensions risque de se révéler à terme - sinon très vite - comme un outil technocratique peu adéquat pour une société sans cesse en changement" (G. Fourez, communication personnelle).

Une autre solution, - qui n'exclut pas la précédente, mais plutôt la complète -, consiste en la création d'une "unité indépendante d'analyse" qui, soit au sein d'une administration de l'éducation, soit en dehors, réunit des représentants des partis politiques, des interlocuteurs sociaux, des entreprises, des parents et du public en général. Une telle unité doit jouir de beaucoup de liberté vis-à-vis du ministère de l'éducation, du parti politique dominant ou, en général, de toute autre autorité qui peut avoir un intérêt personnel en la matière. Une telle unité est normalement financée sur des fonds publics.

□ Synthèse des problèmes généraux

Des problèmes souvent difficiles se posent à propos des indicateurs. Nous en avons déjà rencontré la plupart et nous bornons soit à un rappel global, soit à l'indication de quelques aspects non encore évoqués :

- La difficulté d'un accord sur les buts de l'éducation et sur les objectifs à atteindre.
- L'utilisation appropriée des données.
- La comparabilité des mesures de rendement scolaire.

Pour assurer un maximum de comparabilité, il importe d'utiliser des instruments d'évaluation communs (questionnaires, tests,...). Cette standardisation est plus importante encore si les données à recueillir sont relatives à des comportements tels que les attitudes, les attentes, des traits de personnalité (par exemple, qualité de leader), et la façon d'enseigner.

Les catégories et la présentation doivent être les mêmes partout, surtout pour la publication des données internationales.

- Limiter les mesures aux connaissances et aux habiletés cognitives laisse dans l'ombre d'autres effets possibles de l'éducation, tant pour les individus que pour la société.
- Les indicateurs sont des statistiques. Or il est plus facile de chiffrer la quantité que la qualité. Il importe de toujours se demander si les indicateurs de rendement scolaire reflètent ce que les élèves connaissent réellement et sont effectivement capables de faire.
- Certaines caractéristiques comme la qualité des enseignants sont difficiles à évaluer massivement; des variables substitutives comme le type de formation ou l'ancienneté n'ont qu'une faible validité.
- Il est injuste de juger les enseignants à travers les résultats observés, sans tenir compte des conditions et des facteurs environnementaux. Ils peuvent varier considérablement selon les régions, les quartiers d'une ville, etc.
- Plus les indicateurs sont globaux, plus des phénomènes importants risquent d'être masqués.

- La publication d'indicateurs a une signification politique et il n'est pas rare que des pressions visant à orienter les choix ou la présentation des résultats se manifestent.
- Plus une politique ou un programme d'éducation sont imprécis, plus il est difficile de choisir des indicateurs représentatifs.
- Pour les indicateurs internationaux surtout, il ne faut pas perdre de vue que les valeurs éducatives et les options politiques, de même que les institutions et les façons de faire, varient selon les pays, selon les cultures. Cela peut même être vrai au niveau des régions.
- Les informations doivent être publiées le plus rapidement possible, car il importe de pouvoir réagir vite en cas de problèmes graves; de surcroît, à mesure que le temps passe, la signification d'un indicateur peut changer.
- Même conçus et collectés avec soin, les indicateurs n'apportent pas des informations d'une signification évidente. Celle-ci varie selon les valeurs auxquelles on adhère et aussi selon les options pédagogiques. Les attentes vis-à-vis du système éducatif diffèrent d'ailleurs selon les groupes : élèves, enseignants, familles, entreprises, autorités religieuses.

A. Porter (1988) se demande, en particulier, si le pilotage ne nuira pas à l'autonomie des enseignants. En effet, l'introduction du pilotage semble faire pencher la balance en faveur des partisans d'un contrôle central de ce qui est enseigné, de la façon de le faire, par qui et pour qui. Ne s'oppose-t-on pas ainsi à la décentralisation généralement souhaitée et, en particulier, à la responsabilisation des enseignants et aux projets d'établissements ?

Les deux phénomènes sont-ils incompatibles ? Il ne semble pas, à condition que les enseignants soient associés à la conception et au choix des indicateurs, et que leur formation leur permette d'agir en réelle connaissance de cause.

Le même auteur remarque toutefois que permettre aux enseignants de définir eux-mêmes les limites à l'intérieur desquelles ils estiment pouvoir agir ne garantit pas une amélioration de l'éducation. La participation des enseignants au développement des indicateurs pourrait déboucher sur des indicateurs qui reflètent le plus faible commun dénominateur de la pratique du moment.

En revanche, l'existence d'indicateurs de grande qualité peut être favorable à l'élévation du statut des professeurs : "Pour les enseignants, un bon système d'indicateurs peut être une arme politique permettant de renforcer la profession, car ils ne servent pas seulement au pilotage de la pratique; ils l'influencent aussi."

Exen
de réalisa

Un nombre croissant de pays ont déjà organisé ou instaurent un pilotage de leur système d'enseignement, en particulier de son rendement, et en publient périodiquement les conclusions. Les notes qui suivent portent essentiellement sur les dispositifs. Quelques exemples de résultats sont donnés, simplement pour montrer la nature des informations recueillies et des façons de les présenter. Une description d'ensemble n'était pas possible dans ce précis. Les seuls résultats des enquêtes de l'Association internationale pour l'évaluation du rendement scolaire (IEA), dont il sera question par la suite, occupent de nombreux volumes... La bibliographie permet au lecteur de localiser les ouvrages qui se rapportent aux sujets qui l'intéressent spécialement.

I. INDICATEURS NATIONAUX

□ Etats-Unis

• **La "National Assessment of Educational Progress" (NAEP)**

Décidée, dans son principe, dès 1963 et effectivement instaurée en 1969, cette *évaluation nationale du progrès éducatif* porte sur les connaissances, les habiletés et les attitudes. Les buts poursuivis sont :

- Disposer de façon continue d'informations sur les performances scolaires des élèves.
- Mesurer l'évolution de ces performances à travers le temps.
- Réaliser des sondages particuliers à propos de certains aspects des performances scolaires (par exemple, en matière de compréhension de la lecture) ou du comportement des jeunes (par exemple, leurs habitudes de consommation).
- Réunir des données, les analyser et rédiger à leur propos des rapports destinés à des publics variés : étudiants, professeurs, membres de l'administration, décideurs, pouvoirs organisateurs,...
- Encourager et aider les recherches sur les données réunies.
- Familiariser les chercheurs et les services d'évaluation régionaux et locaux avec les techniques avancées utilisées par la NAEP.
- Continuer à développer et à affiner les techniques nécessaires à la collecte et à l'analyse des données réunies.

Lors d'un nouvel appel d'offres lancé par le Congrès, en 1978, il est précisé que le but premier est d'évaluer les performances des écoliers et des

jeunes adultes dans les habiletés de base : lecture, mathématiques et communication.

A cette fin, la NAEP devra :

- collecter au moins tous les cinq ans des données permettant d'évaluer les performances d'élèves d'âges et de niveaux scolaires divers, en lecture, en expression écrite et en mathématiques;
- faire périodiquement rapport sur les changements observés, pendant une période donnée, dans les savoirs et les savoir-faire relatifs à ces disciplines;
- réaliser d'autres évaluations dans le domaine éducatif, si le besoin s'en fait sentir;
- apporter une assistance technique aux services des Etats et des communautés locales, afin de les aider à comprendre au mieux les objectifs de l'évaluation nationale, à utiliser les données relatives aux habiletés de base, et à faire des comparaisons entre les données locales et les données nationales.

L'évaluation a d'abord été faite pour quatre populations : 9 ans, 13 ans, 17 ans et adultes entre 26 et 37 ans. Par la suite, ce sont des échantillons des 4^e années primaires et des 2^e et 5^e années du secondaire qui ont été testés. Entre 75 000 et 120 000 jeunes sont évalués à chaque opération. Les évaluations ont lieu toutes les années paires. Avec le temps, le nombre de disciplines examinées a aussi augmenté. Ont été ajoutés : les sciences, le civisme, la littérature, la musique, l'histoire, la géographie, la formation professionnelle et l'éducation artistique. Actuellement, la lecture et les mathématiques sont évaluées tous les deux ans, l'expression écrite et les sciences tous les quatre ans, et les autres disciplines, de façon irrégulière.

Les sous-produits de ces opérations sont estimés très importants : développement de banques d'objectifs et de questions, création de bases de données facilitant la construction de tests, etc.

Au début, les résultats publiés concernaient le pays entier et les quatre grandes régions qu'il comprend : Nord-Est, Ouest, Centre, Sud-Est. La comparaison entre Etats était ainsi évitée; dès 1985, elle a cependant été réclamée; c'est chose faite depuis 1990.

La NAEP est financée par l'*Office for Educational Research and Improvement* du Ministère de l'Education des Etats-Unis. Plutôt que d'essayer de constituer un service administratif ad hoc, de très haut niveau scientifique, le Ministère de l'Education préfère recourir à des centres d'excellence. L'exécution des travaux est confiée, depuis 1983, à l'*Educational Testing Service (ETS)*, à Princeton, le plus grand organisme mondial spécialisé dans la construction d'instruments d'évaluation. Bien qu'elle entretienne des rapports étroits avec l'Université de Princeton, ETS est une entreprise privée. Elle a

été fondée en 1947 comme société à but non lucratif. L'importance de son chiffre d'affaires lui permet d'engager les meilleurs spécialistes de l'évaluation... et de consacrer à la recherche en éducation un budget supérieur à la totalité des dépenses que la Communauté française de Belgique y consent.

La prise en charge de la NAEP par l'ETS (qui a créé spécialement en son sein un *Center for the Assessment of Educational Progress - CAEP*) s'est immédiatement traduite par des progrès méthodologiques et techniques considérables dans les processus de collecte, d'analyse et d'interprétation des données.

L'une des critiques les plus fondamentales adressées à la NAEP est son caractère trop descriptif : "La NAEP est une sorte de thermomètre qui indique la température du corps, mais ne nous apprend pas quelle maladie cause la fièvre" (Richards, 1988, p.496). Il n'en reste pas moins que les éléments descriptifs et les signaux donnés présentent un intérêt considérable.

Les rapports de recherche

Ils sont présentés de façon très soignée et rédigés dans un langage accessible au grand public. Certains de ces rapports portent spécialement sur l'évolution constatée entre différentes évaluations. Ils indiquent comment des groupes (constitués selon l'âge, le sexe, la région, le groupe ethnique, le niveau socioculturel des parents, etc.) se comportent soit dans des tâches (*items*) isolées, soit dans des ensembles de tâches.

Les quelques exemples suivants témoignent de la riche moisson de données engrangées.

— *Lecture : premier exemple.*

Pour mesurer la capacité en lecture, en 1986, la NAEP a utilisé une grande variété de textes et accordé une attention particulière aux stratégies de lecture.

Les textes à lire allaient de phrases simples exprimant une seule idée à des articles complexes relatifs à des sujets de haut niveau en sciences naturelles et en sciences sociales. Ils comprenaient des histoires, des poèmes ainsi que des essais et des comptes rendus extraits de livres de lecture élémentaire et de manuels de l'enseignement secondaire; des horaires de chemins de fer et des factures de téléphones figuraient aussi parmi les documents à lire.

La compréhension a été évaluée de plusieurs façons, allant de questions factuelles à choix multiple à des questions ouvertes obligeant l'élève à restructurer et à interpréter la matière lue.

L'étude a été conçue de manière à pouvoir déterminer les relations entre les niveaux d'habileté et différents facteurs contextuels, considérés isolément ou en groupe.

Voici une synthèse des résultats.

Tableau 1 : Compétence moyenne en lecture des élèves de 3e primaire et de 1re et 5e du secondaire. Résultats nationaux

	3e primaire	1re secondaire	5e secondaire
Nation	38,1 (0,2) ^a	48,9 (0,1)	56,1 (0,2)
Blancs	39,8 (0,2)	50,3 (0,1)	57,3 (0,2)
Noirs	33,4 (0,4)	45,2 (0,3)	51,5 (0,3)
Hispaniques	33,2 (0,3)	44,4 (0,4)	51,3 (0,3)
Nord-Est	39,1 (0,3)	50,7 (0,3)	57,4 (0,5)
Sud-Est	37,2 (0,3)	48,1 (0,2)	54,8 (0,3)
Centre	39,3 (0,4)	49,0 (0,2)	56,5 (0,5)
Ouest	36,9 (0,4)	48,0 (0,4)	55,4 (0,4)
Milieu urbain désavantagé	31,9 (0,5)	43,8 (0,4)	51,2 (0,6)
Milieu urbain avancé	41,2 (0,5)	51,6 (0,4)	59,5 (0,5)
Garçons	37,3 (0,2)	47,5 (0,2)	54,5 (0,3)
Filles	38,9 (0,2)	50,3 (0,1)	57,7 (0,2)
Quartile supérieur	47,4 (0,1)	58,2 (0,1)	67,3 (0,1)
Quartile inférieur	25,9 (0,1)	38,1 (0,1)	43,4 (0,1)

a. Erreurs types sur la mesure entre parenthèses

Aux trois niveaux scolaires étudiés, les élèves éprouvent des difficultés marquées pour commenter ce qu'ils ont lu et justifier l'interprétation qu'ils en donnent. Les auteurs concluent que ces habiletés devraient faire l'objet de beaucoup d'attention dans l'enseignement.

Parmi d'autres observations intéressantes, on relève :

- Quand ils ont affaire à des élèves éprouvant des problèmes en lecture, les enseignants insistent surtout sur les stratégies de décodage et beaucoup moins sur la compréhension et la critique.
- Les élèves faibles ne varient guère leurs stratégies de lecture. On trouve, en fin d'enseignement secondaire, des élèves qui continuent à utiliser des approches qui dominent en 3e année primaire.
- En fin d'enseignement secondaire, les performances en lecture des élèves appartenant à des minorités ethniques et aux groupes sociaux à risques restent proches du niveau national de la première année du secondaire.

- Les bons lecteurs se retrouvent dans les filières d'enseignement les plus difficiles. Ils font régulièrement leurs devoirs à domicile et leur famille les incite à lire.

— *Lecture : second exemple*

Dans ce second exemple, les résultats sont exprimés sur une échelle d'habileté à cinq degrés. Les ancrages sont réalisés en appliquant la théorie de la réponse aux items (Voir l'article consacré à l'ancrage des échelles, ainsi que deux exemples détaillés d'échelles, dans l'annexe consacrée aux problèmes techniques). Cette théorie complexe a beaucoup progressé grâce aux recherches réalisées à l'*Educational Testing Service* de Princeton; elle continue à faire l'objet de controverses entre spécialistes.

Tableau 2 : Pourcentages d'élèves qui atteignent ou dépassent l'un des cinq niveaux de compétence en lecture - Années 1971 - 1984. - Ages : 9, 13 et 17 ans^a

Niveau de lecture	Age	1971	1975	1980	1984
		Percent			
Rudimentaire (150)	9	90,4	93,3	94,4	93,9
	13	99,7	99,6	99,8	99,8
	17	99,7	99,9	99,9	100,0
Basal (200)	9	58,3	61,7	65,1	64,2
	13	92,3	92,8	94,3	94,5
	17	96,6	97,5	97,9	98,6
Intermédiaire (250)	9	15,6	14,0	17,0	18,1
	13	57,0	57,5	59,3	60,3
	17	80,0	82,0	82,8	83,6
Apte (300)	9	1,1	0,7	0,8	1,0
	13	9,3	9,7	10,9	11,3
	17	37,2	36,1	34,8	39,2
Avancé (350)	9	0,0	0,0	0,0	0,0
	13	0,2	0,2	0,3	0,3
	17	4,9	3,5	3,1	4,9

a. The NAEP reading scale

Niveau 150 : Rudimentaire - Capacité d'accomplir des tâches de lecture simples et isolées.

Niveau 200 : Basal - Capacité de comprendre des informations isolées ou séquentielles.

Niveau 250 : Intermédiaire - Capacité de chercher des informations spécifiques, de les relier et de généraliser.

Niveau 300 : Apte - Capacité de comprendre, résumer et expliquer des informations relativement complexes.

Niveau 350 : Avancé - Capacité de synthétiser et d'apprendre par la lecture de textes spécialisés.

SOURCE : NAEP, *The Reading Report Card, Progress Toward Excellence in Our Schools* (Report N° 15-R-01), 1985.

On observe que la différence de niveau de performance entre les meilleurs lecteurs et les plus faibles demeure relativement constante à tous les niveaux scolaires. Il reste à trouver les moyens de corriger pareil écart...

La figure suivante, évoquant le bon lecteur, témoigne de la qualité graphique des illustrations destinées à communiquer des informations pédagogiques de base au grand public.

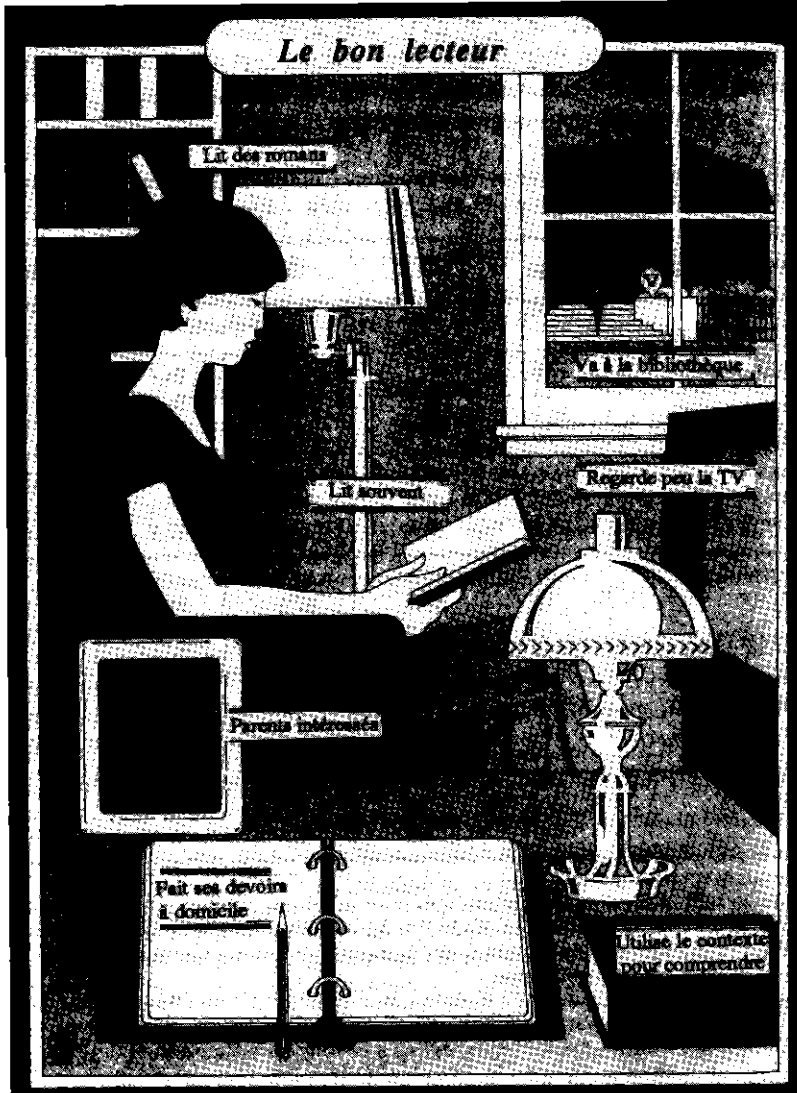


Figure 4

— Ex

Le
les élèves
conduire
que soit
de s'exp
perform
daire, m
supérieu

Et
pu s'exp
cessaires
exercer v
pables d
que leur
J. Lange

A
scolaires
tives sup
persuade

— M

L
loin d'êt

Tab

Niveau 1	Additio élémen
Niveau 2	Début compr
Niveau 2	Maîtris Début problè

— Expression écrite

Les conclusions de cette évaluation, réalisée en 1984 et portant sur les élèves de 4^e année primaire et de 2^e et 5^e années de l'enseignement secondaire, ont été jugées, elles aussi, alarmantes. La plupart des élèves, quelle que soit leur origine (Blancs, Noirs, Asiatiques, Hispaniques) sont incapables de s'exprimer adéquatement par écrit, sauf dans les cas les plus simples. Les performances s'améliorent assez bien entre la 4^e primaire et la 2^e du secondaire, mais beaucoup moins entre l'enseignement secondaire inférieur et le supérieur.

En cinquième année du secondaire, moins d'un quart des élèves ont pu s'exprimer adéquatement dans des tâches mettant en jeu des habiletés nécessaires pour poursuivre des études, entrer dans le circuit commercial ou exercer une profession libérale. La majorité des élèves américains ne sont capables de s'exprimer par écrit qu'au niveau minimal, sans même être assurés que leur message permettra d'atteindre le résultat escompté (A.N. Applebee, J. Langer et I.V.S. Mullis, 1986, p.9).

Autre conclusion grave : cette évaluation révèle qu'à tous les niveaux scolaires, la majorité des élèves sont déficients en matière d'habiletés cognitives supérieures. Ils analysent mal une problématique et ne réussissent pas à persuader.

— Mathématiques

Le tableau suivant concerne les résultats en mathématiques; ils sont loin d'être brillants.

Tableau 3 : Pourcentages d'élèves qui atteignent ou dépassent les cinq niveaux de compétence en mathématiques

	Age	1978	1982	1986
Niveau 150	9	96,5 (0,2)	97,2 (0,3)	97,8 (0,2)
Additions et soustractions élémentaires	13	99,8 (0,0)	99,9 (0,0)	100,0 (0,0)
	17	100,0 (0,0)	100,0 (0,0)	100,0 (0,0)
Niveau 200	9	70,3 (0,9) ^a	71,5 (1,1)	73,9 (1,1)
Début de l'habileté et de la compréhension	13	94,5 (0,4) ^a	97,6 (0,4)	98,5 (0,2)
	17	99,8 (0,0)	99,9 (0,1)	99,9 (0,1)
Niveau 250	9	19,4 (0,6)	18,7 (0,8)	20,8 (0,9)
Maîtrise des 4 opérations.	13	64,9 (1,2) ^a	71,6 (1,2)	73,1 (1,5)
Début de la résolution de problèmes.	17	92,1 (0,5) ^a	92,9 (0,5) ^a	96,0 (0,4)

Tableau 3 : Pourcentages d'élèves qui atteignent ou dépassent les cinq niveaux de compétence en mathématiques

	Age	1978	1982	1986
Niveau 300	9	0,8 (0,1)	0,6 (0,1)	0,6 (0,2)
Procédures de complexité moyenne. Géométrie élémentaire. Raisonnement	13	17,9 (0,7)	17,8 (0,9)	15,9 (1,0)
	17	51,4 (1,1)	48,3 (1,2)	51,1 (1,2)
Niveau 350	9	0,0 (0,0)	0,0 (0,0)	0,0 (0,0)
Résolution de problèmes en plusieurs étapes. Algèbre.	13	0,9 (0,2)	0,5 (0,1)	0,4 (0,1)
	17	7,4 (0,4)	5,4 (0,4)	6,4 (0,4)

a. Différence statistiquement significative au seuil de 0,5 par rapport à 1986.

Le niveau 150 est atteint à peu près par tous, mais, dès le niveau 200, les difficultés s'annoncent : les élèves éprouvent des difficultés pour effectuer des opérations qui exigent plus qu'un raisonnement numérique simple. Moins d'un quart des élèves de neuf ans atteignent le niveau 250 (maîtrise des quatre opérations sur des nombres entiers et raisonnement arithmétique élémentaire) et, en gros, un tiers des élèves de 13 ans n'atteignent pas le niveau. Pour les deux niveaux suivants, les observations sont pour le moins inquiétantes.

Les auteurs concluent (p.12) que les résultats sont désespérants pour deux raisons. D'abord, trop peu d'élèves, quel que soit le groupe d'âge, atteignent les niveaux correspondant respectivement aux niveaux primaire, secondaire inférieur et secondaire supérieur. Ensuite, parce que l'écart entre les performances réelles des élèves et celles que l'on pouvait attendre ne fait que se creuser à mesure que l'on avance dans la scolarité.

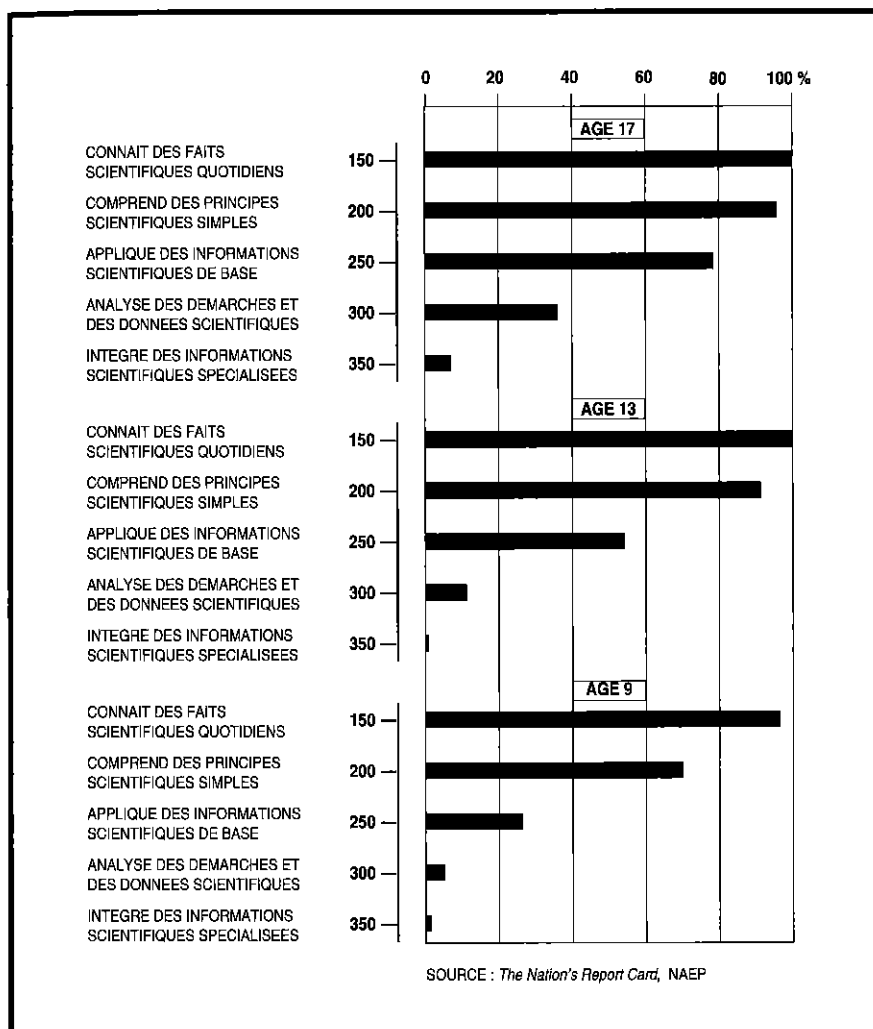
Commentant les données de 1986 et se référant aux enquêtes internationales de l'IEA et de l'IAEP, on peut lire deux observations incendiaires dans l'*Executive summary* du rapport de la NAEP, focalisé sur les résultats de 1986 (J.A. Dossey *et al.*, s.d., p.1) :

"Les données internationales les plus récentes concernant l'apprentissage des mathématiques montrent que l'élève japonais moyen atteint un niveau supérieur à celui des 5% les plus brillants parmi les élèves des Etats-Unis qui fréquentent les cours de mathématiques préparatoires à l'université. Récemment, une société de production de semi-conducteurs japonaise, installant une usine aux Etats-Unis, a dû engager des étudiants ayant terminé le premier cycle universitaire, pour effectuer des contrôles statistiques de qualité qui sont confiés, au Japon, à des élèves sortant de l'enseignement secondaire."

— Sciences

Voici d'abord la synthèse des résultats obtenus en sciences, en 1986, par les élèves de 9, 13 et 17 ans habitant la région ouest des Etats-Unis (Mullis et Jenkins, 1988, p.60).

Tableau 4 : Compétences en sciences. Pourcentages d'élèves qui atteignent ou dépassent les niveaux d'ancrage en 1986 - Ages : 9, 13 et 17 ans.



1986
 5 (0,2)
 9 (1,0)
 1 (1,2)
 0 (0,0)
 4 (0,1)
 4 (0,4)

au 200,
 effectuer
 simple.
 (maîtrise
 métrique
 pas le ni-
 moins in-

ants pour
 'âge, at-
 aire, se-
 entre les
 e fait que

tes inter-
 endiaires
 résultats

is-
 int
 les
 s à
 uc-
 les
 er
 les

Toujours à propos des sciences, et pour les mêmes âges, le tableau suivant donne trois points de repère (1977, 1982 et 1986) permettant d'estimer l'évolution en dix ans.

Tableau 5 : Pourcentages d'élèves qui atteignent ou dépassent les cinq niveaux de compétence en sciences.
Âges : 9, 13 et 17 ans.

	Age	1977	1982	1986
Niveau 150 Connaissance rudimentaire de plantes et d'animaux. Trouve comment utiliser des appareils domestiques simples	9	93,6 (0,5) ^a	95,0 (0,5)	96,3 (0,3)
	13	98,6 (0,1)	99,6 (0,1)	99,8 (0,1)
	17	99,8 (0,0)	99,7 (0,1)	99,9 (0,1)
Niveau 200 Comprend des principes scientifiques simples relatifs à la vie animale ou végétale	9	67,9 (1,1) ^a	70,4 (1,6)	71,4 (1,0)
	13	85,9 (0,7) ^a	89,6 (0,7)	91,8 (0,9)
	17	97,2 (0,2)	95,8 (0,4)	96,7 (0,4)
Niveau 250 Début de la compréhension de la démarche expérimentale	9	26,2 (0,7)	24,8 (1,7)	27,6 (1,0)
	13	49,2 (1,1) ^a	51,5 (1,4)	53,4 (1,4)
	17	81,8 (0,7)	76,8 (1,0) ^a	80,8 (1,2)
Niveau 300 Analyse des données et des démarches scientifiques	9	3,5 (0,2)	2,2 (0,6)	3,4 (0,4)
	13	10,9 (0,4)	9,4 (0,6)	9,4 (0,7)
	17	41,7 (0,8)	37,5 (0,8) ^a	41,4 (1,4)
Niveau 350 Intègre des informations scientifiques spécialisées	9	0,0 (0,0)	0,1 (0,1)	0,1 (0,1)
	13	0,7 (0,1)	0,4 (0,1)	0,2 (0,1)
	17	8,5 (0,4)	7,2 (0,4)	7,5 (0,6)

a. Différence statistiquement significative au seuil de 0,5 par rapport à 1986.

De 1977 à 1986, un léger progrès général se manifeste. Néanmoins, les connaissances en sciences et la capacité de les utiliser restent très limitées. De surcroît, des données qui ne sont pas reproduites ici indiquent que les élèves appartenant à des minorités ethniques obtiennent des résultats nettement inférieurs aux moyennes nationales. En outre, les filles restent moins performantes que les garçons.

Il est intéressant d'ajouter aux données qui précèdent, des informations qui ne figurent que très rarement dans les publications relatives aux performances scolaires. Elles concernent, d'une part, la façon dont les professeurs enseignent les sciences et, d'autre part, ce que font les élèves (I.V.S. Mullis *et al.*, 1988, p.99).

Tableau 6 : Pratique de l'enseignement en sciences, 1986.

A quelle fréquence (%) votre professeur de sciences fait-il les choses suivantes dans son enseignement ?	Tous les jours	Plusieurs fois par semaine	Une fois par semaine	Moins d'une fois par semaine	Jamais
1^{re} secondaire					
Expose la matière	31 (0,9)	21 (0,9)	18 (1,1)	13 (0,9)	18 (0,9)
Démontre un principe scientifique	15 (1,0)	22 (1,3)	22 (1,0)	22 (1,3)	19 (0,9)
Demande d'expliquer les résultats d'une expérience	14 (1,1)	19 (1,3)	20 (0,8)	24 (1,1)	23 (1,0)
Demande des hypothèses	8 (0,8)	12 (0,8)	15 (1,0)	24 (1,0)	42 (1,5)
Demande qu'on interprète des données	9 (0,9)	12 (0,7)	15 (0,7)	15 (1,0)	48 (1,2)
5^e secondaire					
Expose la matière	55 (1,0)	24 (0,8)	9 (0,8)	3 (0,3)	8 (0,7)
Démontre un principe scientifique	18 (1,0)	31 (1,6)	28 (1,2)	13 (1,0)	10 (0,7)
Demande d'expliquer les résultats d'une expérience	16 (1,1)	29 (0,9)	28 (1,0)	15 (0,9)	12 (0,8)
Demande des hypothèses	11 (0,7)	19 (0,8)	23 (1,3)	22 (1,0)	25 (1,3)
Demande qu'on interprète des données	15 (0,8)	24 (0,9)	23 (0,9)	16 (0,9)	22 (1,0)

Erreurs types entre parenthèses

Tableau 7 : Enseignement des sciences.
Fréquence (%) de certaines activités didactiques en 1986

Indiquez le nombre de fois que les activités suivantes ont lieu au cours de sciences	Tous les jours	Par semaine	Par mois	Par an	Jamais
3^e primaire					
Film	—	23 (1,8)	16 (0,9)	18 (1,3)	31 (1,3)
Excursion	—	3 (0,5)	7 (0,6)	24 (1,1)	55 (1,7)
Faire des expériences	11 (1,0)	30 (1,4)	—	26 (1,0)	19 (1,4)
Lire dans le manuel de sciences	33 (1,7)	27 (1,5)	—	11 (0,7)	15 (0,8)
Faire un rapport oral ou écrit	20 (0,9)	16 (0,9)	—	17 (0,9)	33 (1,4)

Erreurs types entre parenthèses

Indiquez le nombre de fois que les activités suivantes ont lieu au cours de sciences	Tous les jours	Plusieurs fois par semaine	Une fois par semaine	Moins d'une fois par semaine	Jamais
1re secondaire					
Film	4 (0,7)	11 (1,0)	23 (1,4)	40 (1,7)	21 (1,4)
Excursion	1 (0,2)	1 (0,2)	1 (0,3)	16 (1,5)	82 (1,5)
Faire des expériences seul	3 (0,5)	5 (0,5)	13 (0,9)	29 (1,0)	50 (1,4)
Faire des expériences en groupe	4 (0,5)	9 (0,7)	22 (1,4)	34 (1,5)	30 (1,3)
Décrire une expérience	4 (0,6)	5 (0,5)	10 (1,1)	21 (1,3)	60 (1,6)
Lire dans le manuel de sciences	43 (1,7)	24 (1,2)	15 (0,8)	8 (0,9)	9 (0,9)
Lire des articles scientifiques	12 (1,1)	14 (0,8)	23 (1,3)	25 (1,1)	26 (1,3)
Faire un rapport oral ou écrit	3 (0,4)	5 (0,5)	11 (0,9)	35 (1,9)	46 (1,9)
5e secondaire					
Film	2 (0,2)	7 (0,6)	22 (1,2)	42 (1,2)	27 (1,1)
Excursion	0 (0,2)	0 (0,1)	1 (0,4)	13 (1,2)	86 (1,2)
Faire des expériences seul	2 (0,3)	7 (0,6)	20 (0,9)	25 (1,0)	46 (1,3)
Faire des expériences en groupe	3 (0,6)	14 (0,9)	36 (1,2)	28 (1,4)	18 (1,4)
Décrire une expérience	2 (0,3)	10 (0,9)	22 (1,3)	24 (1,5)	41 (1,6)
Lire dans le manuel de sciences	28 (1,0)	26 (1,0)	16 (0,9)	12 (0,8)	18 (1,3)
Lire des articles scientifiques	7 (0,7)	10 (0,8)	18 (1,1)	26 (1,1)	39 (1,5)
Faire un rapport oral ou écrit	2 (0,3)	3 (0,3)	10 (0,8)	34 (1,3)	52 (1,6)

Erreurs types entre parenthèses

— *Un retentissement national*

Les résultats alarmants des enquêtes qui viennent d'être évoqués ne signifient cependant pas que l'enseignement des Etats-Unis est toujours de moindre qualité qu'ailleurs. D'évidence, l'explication se trouve en partie dans les conditions très défavorables dans lesquelles beaucoup d'écoles doivent travailler.

La manifestation la plus tangible de l'inquiétude qui règne dans le pays est sans doute la publication, par le Président des Etats-Unis, d'un opuscule intitulé *A Nation at Risk* (La nation est en péril). Dans son retentissant discours du 18 avril 1991, George Bush proclame que les Etats-Unis doivent devenir "une nation d'étudiants" : en l'an 2000, 90% des jeunes devraient

terminer l'enseignement secondaire long avec succès et les élèves américains devraient être devenus les meilleurs du monde en mathématiques et en sciences. L'avenir dira si pareil désir a pu devenir réalité...

Une masse de publications

La NAEP a des centaines de publications à son actif. Elles sont du domaine public et peuvent être reproduites sans autorisation. Outre celles qui ont déjà été évoquées, il faut mentionner :

- *Les Rapports techniques*, publications de haut niveau scientifique, décrivant notamment les techniques statistiques utilisées. La NAEP 1983-84 *Technical Report* N° 15, dû à A.E. Beaton (1987), est un modèle du genre.
- *Les Fascicules de données* objectives préparés par des commissions composées de chercheurs, de spécialistes des disciplines, de membres de l'administration, d'enseignants, de spécialistes de l'évaluation, de représentants des entreprises et de citoyens concernés.
- *Les Ensembles d'items de tests*. Après leur utilisation, environ la moitié des items de tests construits pour les évaluations sont ainsi mis à la disposition des Etats et des communautés locales. Les items appartiennent alors au domaine public. Ils concernent les diverses disciplines et présentent un large éventail de difficulté. Ils sont accompagnés du pourcentage d'étudiants de l'échantillon national qui ont répondu correctement aux questions.
- *Les Etudes*. Ces documents de haut niveau, produits par des spécialistes, traitent d'aspects particuliers de l'évaluation et de questions connexes.
- *Les Publications méthodologiques*. Il s'agit, dans certains cas, de véritables traités de plusieurs centaines de pages.

The National Education Goals Reports

En 1989, le Président des Etats-Unis a réuni les gouverneurs de tous les Etats du pays pour définir *les buts éducatifs majeurs de la nation* pour l'an 2000. Afin de piloter les actions destinées à poursuivre ces objectifs, une commission intitulée *National Educational Goals Panel* a été constituée. Elle publie chaque année un rapport public (*The National Education Goals Report*) informant des progrès accomplis dans la poursuite de ces objectifs par le système éducatif.

Les grandes rubriques de ce rapport sont les suivantes :

1. Les habiletés et les dispositions favorables à l'étude.

2. Le pourcentage du groupe d'âge terminant l'enseignement secondaire supérieur.
3. Le rendement scolaire en mathématiques et en sciences.
4. L'habileté en lecture.
5. Une vie saine, éloignée de la drogue.

Le rapport de 1992 insiste spécialement sur le contexte international : les performances observées aux Etats-Unis sont comparées à celles d'autres nations.

Autres évaluations nationales importantes : des études longitudinales

A côté des enquêtes nationales réalisées par la NAEP, d'autres recherches évaluatives de grande envergure sont réalisées dans le pays. Des études longitudinales à grande échelle méritent une attention particulière.

Normalement, les indicateurs ne sont pas collectés, à travers le temps, sur les mêmes élèves, mais bien sur des échantillons représentatifs de diverses populations d'élèves se trouvant à un niveau déterminé de leurs études. On fait l'hypothèse que ces populations successives présentent des caractéristiques équivalentes. Ce n'est pas toujours certain. Suivre les mêmes échantillons pendant plusieurs années peut évidemment apporter des informations plus fines et plus valides sur les changements de comportements et de performances, et plus généralement sur le devenir des individus.

Les études longitudinales sont difficiles. On constate, en particulier, que l'échantillon initial rétrécit en raison de changements de domicile ou d'école, de désintérêt, voire pour des raisons plus graves encore. Des moyens considérables et coûteux doivent être déployés pour maintenir les contacts. On sait aussi que les élèves qui refusent de continuer à participer peuvent présenter certaines caractéristiques psychologiques spécifiques, ce qui tend à biaiser l'échantillon. Enfin, surtout lorsqu'il s'agit de déterminer les effets de programmes éducatifs, le recours à des plans expérimentaux de grande complexité peut être nécessaire.

Il n'en reste pas moins que des suivis de mêmes cohortes, fussent-ils relativement modestes, sont hautement souhaitables.

A cet égard, et face à la crise de l'éducation qui les atteint actuellement, les Etats-Unis ne lésinent pas sur les moyens. R. C. Atkinson et G.B. Jackson (1992, pp. 46-47) décrivent les actions en cours. Elles sont impressionnantes. En effet, le Centre National des Statistiques de l'Education (NCES) des Etats-Unis finance simultanément trois grandes études longitudinales : la *National Longitudinal Study* de 1972, la *High School and Beyond Study* de 1980 et la *National Educational Longitudinal Study* de 1988.

Le Projet Talent

Avant de décrire les recherches qui viennent d'être mentionnées, il importe de rappeler qu'elles furent précédées d'une autre, restée unique dans l'histoire de l'éducation : **Le Projet Talent**. Cette recherche a été financée par l'*American Institute for Research* et réalisée par l'Université de Pittsburgh (Flanagan, J.C. *et al.* 1964; Flanagan, J.C., Ed., 1971).

Parallèlement au recensement légal de 1960, les Etats-Unis ont décidé de procéder à un inventaire des aptitudes de tous les jeunes âgés de 15 ans et de suivre leur développement pendant vingt-cinq ans, à raison d'une nouvelle évaluation tous les cinq ans. L'un des buts de ce suivi était notamment de découvrir quelles expériences éducatives auraient le plus contribué à la réalisation des potentialités décelées.

Un échantillon stratifié, nationalement représentatif, a été tiré au hasard. Il compte 500 000 jeunes, dont tous les élèves de cet âge inscrits en dernière année de l'enseignement secondaire inférieur et dans les trois années du secondaire supérieur. Des jeunes qui n'étaient déjà plus scolarisés ont aussi été testés.

Non seulement des informations générales ont ainsi été recueillies (notamment à l'aide d'un inventaire d'intérêts et d'un questionnaire de personnalité), mais les tests suivants ont aussi été subis : langue maternelle (expression, bon usage, orthographe), mathématiques (connaissance, compréhension et utilisation des concepts de base), aptitude à l'apprentissage des langues étrangères. Les jeunes ont aussi été interrogés sur leur plan d'avenir. Les chefs d'établissement et les conseillers pédagogiques ont été interviewés parallèlement.

Il n'est pas sans intérêt historique de relever que, comme les études internationales de l'IEA, ces recherches n'ont été possibles que par l'arrivée de l'ordinateur : il s'agissait de dépouiller plus de deux millions de formulaires et de traiter plus d'un milliard de données...

Tous les jeunes examinés, avec qui le contact a pu être maintenu, ont été revus 1, 5, 10 et 20 ans après l'âge normal de sortie de l'enseignement secondaire. Ils ont notamment été interrogés sur les études faites, les activités de travail et de loisirs, et sur leurs problèmes.

Les publications relatives à cette étude continuent à sortir.

National Longitudinal Study, 1972 : entrée dans l'enseignement supérieur et dans la vie active

Cette étude, portant sur un échantillon national de 16 683 élèves terminant leurs études, est relative à la transition de l'enseignement secondaire long à l'enseignement supérieur, et de l'enseignement supérieur à la vie active. Les données recueillies sont de nature démographique et scolaire : cours

suivis, résultats scolaires, projets d'études et de travail. Des données de suivi ont été collectées en 1973, 1974, 1976, 1979 et 1986, afin d'identifier les changements intervenus dans les études, le travail et le statut marital.

Il apparaît que la majorité des élèves testés en 1972 ont continué leurs études, mais n'ont pas suivi le modèle traditionnel d'études supérieures faites en quatre ans, immédiatement après la sortie du secondaire. Généralement, les hommes mettent plus de temps à faire leurs études que les femmes.

L'étude *High School and Beyond* de 1980 a, elle aussi, porté sur la transition de l'école à la vie active ; le déclin du rendement scolaire a été la principale raison de cette recherche. En 1980, des échantillons nationaux de 28 000 élèves de l'avant-dernière année et de 30 000 élèves de la dernière année de 1 100 écoles secondaires ont été testés.

En 1982, 1984 et 1986, de larges sous-ensembles de ces deux cohortes ont été réexaminés; en 1992, les élèves qui avaient été testés en avant-dernière année du secondaire, en 1980, ont été revus. Les enquêtes ont porté sur les attitudes des parents, les avis des professeurs et les scores des élèves à différents tests. D'importantes observations ont pu ainsi être faites à propos de l'abandon des études et des facteurs qui déterminent les aspirations au cours des deux dernières années du secondaire.

Des comparaisons entre les deux études qui viennent d'être mentionnées ont permis de déterminer dans quelle mesure les changements sociaux et culturels intervenus ont modifié les attitudes et les performances scolaires des étudiants observés.

National Educational Longitudinal Study, 1988 : facteurs de la réussite scolaire

Cette étude a pour objectif de déterminer les facteurs qui contribuent aux performances scolaires, à la poursuite des études et à l'entrée dans l'enseignement supérieur. Pendant dix ans, 25 000 élèves seront suivis à partir de la 4^e année de l'enseignement secondaire. Outre le recueil d'informations générales, des tests de lecture, de mathématiques, de sciences, de géographie et d'histoire ont été administrés en 1988, 1990 et 1992. On dispose ainsi d'une ligne de base concernant 36% d'élèves blancs, 44% d'élèves noirs et 39% d'hispaniques, qui présentent au moins deux facteurs de risque. Ces facteurs sont : les familles monoparentales, la pauvreté (revenu annuel inférieur à 15 000 \$), les enfants laissés seuls à la maison pendant plus de trois heures par jour, les parents n'ayant pas réussi d'études secondaires, l'abandon des études par un frère ou une soeur, la pauvre connaissance de la langue anglaise.

Des systèmes d'indicateurs propres aux Etats

Presque tous les Etats possèdent leurs tests de connaissances et beaucoup d'entre eux créent leur propre système d'indicateurs et de pilotage. La Californie, le Connecticut, l'Etat de New York et la Caroline du Sud sont particulièrement avancés à cet égard.

L'exemple de la Californie

Le système d'indicateurs PACE, actuellement en vigueur en Californie, porte sur les aspects suivants, à propos desquels on rend compte dans une publication annuelle : **L'Etat de l'éducation en Californie.**

Inputs

- Nombre d'élèves et leurs caractéristiques : appartenance à des minorités ethniques ou linguistiques, composition et revenus de la famille,...
- Ressources humaines, y compris les profils des enseignants et des membres de l'administration; diplômes des enseignants, programmes de formation continuée des maîtres (y compris pendant la première année de prise de fonction) et du staff.
- Organisation des classes et système de discipline; caractéristiques de l'école et du district.

Processus

- Curriculum et programmes spéciaux, branches suivies par les élèves, exigences pour l'attribution du diplôme de fin d'enseignement secondaire, examens, manuels scolaires et autres matériels, financement de programmes spéciaux.

Outputs

- Performances des élèves, y compris la tendance de leurs résultats aux *California Achievement Tests* (résultats scolaires globaux, selon l'ethnie, la langue maternelle, le niveau socio-économique de la famille), scores au *Scholastic Aptitude Test* (test d'aptitude aux études supérieures), informations sur les travaux à domicile, le nombre d'heures quotidiennes de TV et, en général, toutes autres variables pouvant influencer les résultats scolaires.

Pour la plupart de ces indicateurs, on dispose déjà des tendances sur dix ans.

Les **Performance Reports for California Schools** concernent cinq secteurs principaux. Les quatre premiers s'intéressent à la qualité de l'éducation, tandis que le cinquième a trait aux variables contextuelles relatives aux écoles et aux élèves :

1. Cours choisis par les élèves et compétences minimales exigées.
2. – Performances en lecture et en mathématiques des élèves terminant l'enseignement secondaire long.
 - Pourcentage des élèves se situant respectivement dans le 1er, 2e ou 3e quartile.
 - Scores moyens des écoles.
3. Scores obtenus par les élèves qui veulent faire des études supérieures :
 - Au *Scholastic Aptitude Test (SAT)*. L'utilisation de ce test d'aptitude aux études supérieures, dans tout le pays, depuis de nombreuses années, permet de déceler les évolutions, et les résultats servent souvent de point de référence dans les débats publics. (Le SAT porte principalement sur le vocabulaire, le raisonnement verbal, la connaissance du programme de mathématiques de l'enseignement secondaire, le raisonnement quantitatif. Il possède un pouvoir discriminatif élevé pour la sélection initiale et une bonne validité prédictive du succès des études universitaires).
 - A des tests de rendement portant sur cinq branches principales.
4. Taux d'abandons scolaires.

Un questionnaire auquel les élèves répondent permet de déterminer le niveau socio-économique et socioculturel des écoles, ainsi que d'autres caractéristiques de leur environnement.

□ Angleterre

• *L'Unité d'Évaluation du Rendement Scolaire (Assessment of Performance Unit, APU)*

La création de l'APU, par le Ministère de l'Éducation et de la Science, a été décidée en 1975, mais les opérations effectives n'ont commencé qu'en 1978; elles ont duré jusqu'à 1988. A l'origine, l'évaluation était moins destinée au grand public qu'aux responsables de la répartition des crédits pour l'éducation. Après peu de temps, la diffusion des résultats est devenue beaucoup plus large. Par principe, on ne rendait publics ni le nom des élèves testés, ni celui des écoles. On voulait ainsi éviter que les écoles les plus performantes n'attirent les élèves les plus doués, ce qui affaiblirait encore les écoles qui obtenaient de moins bons résultats. Pour ces dernières, n'importe-t-il pas surtout d'offrir des moyens d'amélioration ?

Des échantillons représentatifs des élèves de 11, 13 et 15 ans ont été testés dans les domaines suivants :

- Langage : habileté de communication en langue maternelle par la lecture, l'écriture, l'écoute et la parole.
- Mathématiques (insistance sur la communication).
- Sciences : administration de la preuve, vérification d'hypothèses, expérimentation, capacité de tirer des conclusions fondées.
- Développement personnel et social : apprendre à mieux se connaître, sens des responsabilités, comportement moral envers la société et l'environnement.
- Développement esthétique.
- Développement physique, y compris la psychomotricité.

A partir de 1980, les résultats des écoles, aux divers examens, ont été rendus publics. Assez rapidement, des systèmes de modération ont permis une meilleure interprétation de ces résultats.

En 1988, une réforme générale de l'éducation est intervenue et, avec elle, des dispositions de pilotage très poussées ont été prises. L'APU a été supprimée ainsi que le *School Evaluation and Assessment Council* (SEAT). Cette décision coïncide avec l'adoption d'un curriculum national.

• **Le Curriculum National et son évaluation : un pilotage systématique**

En Angleterre, il est de tradition que les Autorités Locales déterminent avec beaucoup de liberté le curriculum des écoles qui dépendent de leur juridiction. Afin de disposer d'informations sur les rendements au niveau national, une loi de réforme de l'éducation (*Education Reform Act*) de 1988 prescrit une évaluation nationale portant sur un curriculum national. Comme on vient de l'indiquer, cette évaluation se substitue à celle de l'APU. Des échantillons représentatifs des populations de 7, 11, 14 et 16 ans doivent être testés afin de déterminer dans quelle mesure les objectifs définis sont atteints.

La loi de 1988 marque un tournant dans l'histoire de l'éducation de l'Angleterre. Pour la première fois, un cadre légal destiné à élever le niveau des rendements scolaires est créé. La loi instaure aussi un *Conseil National du Curriculum*, organisme responsable de la réforme continue des programmes; ce conseil a aussi pour mission de donner des avis au gouvernement et aux écoles.

Le Curriculum National comprend trois disciplines "centrales", - la langue maternelle, les mathématiques et les sciences -, et sept disciplines "fondamentales", - la technologie, l'histoire, la géographie, une langue étrangère, les arts plastiques, la musique et l'éducation physique. S'y ajoute la religion, si les parents le souhaitent.

Quatre cycles d'études sont distingués :

1. de 5 à 7 ans.
2. de 8 à 11 ans.
3. de 12 à 14 ans.
4. de 15 à 16 ans.

Les élèves des deux premiers cycles doivent étudier toutes les branches, sauf la langue étrangère. Dans les deux cycles supérieurs, la souplesse est plus grande.

La loi de 1988 a été complétée, en 1992, par un livre blanc intitulé *Choix et diversité : un cadre nouveau pour les écoles*. Il apporte notamment les précisions suivantes concernant l'accroissement de l'autonomie des établissements et le pilotage :

- La constitution de réseaux d'écoles autonomes (*self-governing*) est encouragée, afin de mieux répondre à la diversité des besoins et des intérêts des élèves.
- Une importance encore accrue est accordée au pilotage rigoureux des écoles et aux comptes à rendre aux parents et à la communauté éducative.

Un aspect particulier du pilotage prévu mérite une mention spéciale. Que faire si les évaluations et les rapports d'inspection font apparaître qu'une école dépendant des Autorités Locales n'assure pas une éducation de qualité suffisante à ses élèves, fonctionne de façon inacceptable et doit donc être considérée comme un établissement "à risques" ?

"En un premier temps, l'Autorité Locale sera mise en demeure d'élaborer un plan de redressement. S'il échoue, le Ministre chargera une Association éducative d'assurer le fonctionnement de l'école. Une telle association doit comprendre des personnes possédant une expérience reconnue en matière de gestion et d'éducation. L'association devra, en effet, gérer l'école jusqu'à ce que les performances éducatives soient améliorées dans la mesure souhaitée. Ce n'est qu'à ce moment que des subventions de l'Etat pourront être accordées" (*The White Paper, School Update*, automne 1992, p.3).

Pour chaque cycle et pour chaque discipline, des cibles d'apprentissage (*attainment targets*) sont définies, de même que les programmes et les dispositifs d'évaluation.

Par *cible d'apprentissage*, on entend les connaissances, les habiletés et la compréhension réelle que les élèves doivent normalement atteindre à la fin de chaque cycle. Pour chacun de ces objectifs à atteindre, on distingue dix niveaux de performance; cette disposition devrait permettre d'évaluer les progrès accomplis au cours de la scolarité obligatoire.

Par *programme d'études*, on désigne les contenus, les habiletés et les processus à faire acquérir par des élèves d'aptitudes et de maturité différentes, pendant les quatre cycles.

Par *dispositif d'évaluation*, on désigne les évaluations qui sont effectuées à la fin ou presque au terme de chaque cycle, afin de déterminer dans quelle mesure les performances atteignent les *cibles*. L'évaluation résulte de la combinaison des scores obtenus à des tests élaborés à l'extérieur et de l'appréciation des enseignants. Fait important : les tests ne seront pas les mêmes pour toutes les classes; les enseignants pourront constituer les épreuves à partir d'une banque de questions. A tous les stades, les parents seront informés des résultats de leurs enfants. Les résultats d'ensemble à 11, 14 et 16 ans seront rendus publics.

Les programmes pour les divers cycles, ainsi que les dix niveaux des cibles d'apprentissage sont très détaillés. En témoigne le court extrait ci-dessous relatif aux cibles d'apprentissage en technologie de l'information, point figurant au programme de technologie.

NIVEAU 1

- 1a Pouvoir travailler avec un ordinateur.
- 1b Pouvoir expliquer que des jouets ou des appareils ménagers peuvent être commandés par des signaux ou des instructions informatiques.

NIVEAU 2

- 2a Pouvoir utiliser des images, des symboles, des phrases générés par ordinateur pour communiquer du sens.
- 2b Utiliser l'ordinateur pour stocker et retrouver de l'information.

...

NIVEAU 9

- 9a Etre capable d'évaluer un package de logiciels ou un type d'ordinateur, d'analyser la situation pour laquelle ils ont été conçus; évaluer leur efficacité, leur facilité d'utilisation, leur niveau d'adéquation et suggérer des perfectionnements.
- 9b Concevoir et réaliser un système informatique utilisable par d'autres et rédiger les instructions relatives à son utilisation.
- 9c Comprendre les conséquences de l'introduction de données inexactes dans des fichiers contenant des informations personnelles.

NIVEAU 10

- 10a Etre capable de modéliser un système, d'en concevoir la réalisation et de la tester; justifier les méthodes utilisées et les choix faits.
- 10b Discuter des problèmes que la technologie de l'information pose dans les domaines de l'économie, de l'éthique, de la morale et de la société.

L'organisation et la direction de l'évaluation sont du ressort du *Schools Evaluation and Assessment Council* (SEAC).

Cette évaluation ne se traduit jamais en termes de réussite ou d'échec. Tout élève n'atteignant pas le niveau prévu doit subir des tests diagnostiques complémentaires. Des tests spéciaux sont aussi conçus pour les enfants ayant des besoins particuliers. A cet égard, ce nouveau système se veut nettement formatif.

L'évaluation est aussi conçue pour apporter aux professeurs le feedback dont ils ont besoin pour planifier leur enseignement à un niveau qui convient à leurs élèves. En particulier, la distinction des dix niveaux, pour chaque cible d'apprentissage, devrait aider les enseignants à partir de ce que leurs élèves savent, comprennent et peuvent faire. C'est ainsi que l'on espère obtenir une élévation systématique du niveau des performances.

Enfin, les écoles sont libres d'appliquer le curriculum selon les modalités qui leur conviennent.

Comme déjà mentionné, les résultats des écoles seront accessibles au public, et les parents ou les responsables de l'école pourront aussi avoir accès aux résultats d'une classe déterminée. Ces résultats ne seront pas exprimés en termes de valeur ajoutée en fonction du niveau socio-économique du recrutement, mais, pour chaque école, un rapport sera élaboré non seulement sur la façon dont elle fonctionne, mais aussi sur les données contextuelles - socio-économiques et autres - qui peuvent l'influencer.

Il est fait obligation aux écoles d'envoyer aux parents un *rapport annuel* indiquant non seulement les résultats que leur enfant a obtenus aux tests relatifs au Curriculum National, mais aussi toutes autres informations relatives aux progrès accomplis.

Concrètement et à titre d'exemple, les parents d'un enfant de sept ans, testé dans l'année, devront recevoir les informations suivantes :

- Où l'enfant se situe en langue maternelle, en mathématiques, en sciences et en technologie par rapport aux cibles d'apprentissage.
- Comment l'enfant se situe par rapport à certains aspects particuliers de ces disciplines (profils). Par exemple, pour la langue maternelle, les niveaux atteints sont indiqués pour l'expression orale, la compréhension des messages oraux, la lecture et l'expression écrite.
- Comment ces résultats se situent par rapport à ceux des autres élèves de sept ans fréquentant l'école,
- Les progrès dans chacune des branches du Curriculum National.
- Les progrès dans les autres branches et activités.
- Une appréciation globale des progrès de l'élève.
- Le relevé des absences.

- Le nom d'un membre du personnel de l'école avec lequel le rapport peut être discuté.

Le programme d'évaluation adopté en 1992 était le suivant :

- A partir de 1993, les élèves de 7 ans seront testés en langue maternelle, mathématiques, sciences, technologie, histoire, géographie.
- A partir de 1994, les élèves de 11 ans seront testés en langue maternelle, mathématiques, sciences, technologie, et les élèves de 14 ans en histoire et géographie.
- A partir de 1995, les élèves de 11 ans seront testés en histoire et géographie, les élèves de 14 ans en langues étrangères.

Toutefois, un allègement de ce programme d'évaluation a été préconisé en 1993, en raison de l'opposition massive des enseignants.

Décision importante, on publiera, à l'avenir, des mesures standardisées des progrès que les élèves de chaque école ont accomplis dans le Curriculum National. Plus précisément :

- A partir de 1995, on établira les progrès accomplis entre le testing de 7 ans et celui de 11 ans.
- A partir de 1997, entre 11 et 14 ans.
- A partir de 1999, entre 11 et 16 ans.

Par ailleurs, toutes les écoles devront publier annuellement un *prospectus* qui, à côté d'informations diverses, devra mentionner :

- les résultats obtenus par l'école aux tests portant sur le Curriculum National;
- ce que sont devenus les élèves qui ont terminé leur scolarité obligatoire dans l'école.

Le Curriculum National, qui va à l'encontre d'une longue tradition de grande liberté locale, a fait l'objet d'une forte opposition de la part des enseignants et de leurs syndicats, mais le gouvernement a passé outre. Il souligne que d'importantes possibilités de modulations locales sont offertes. On craint aussi que la publication des résultats obtenus par les écoles ne nuisent aux établissements qui travaillent dans des contextes difficiles. L'idée libérale de la libre concurrence, y compris en matière scolaire, a cependant trouvé ici de nouveaux défenseurs.

L'opposition des enseignants s'explique notamment par le bouleversement de leurs habitudes de travail. Ils craignent la mise en place d'une grande machine administrative à caractère technocratique, susceptible d'accapa-

rer leur temps et leur énergie, en particulier pour des tâches d'évaluation, au détriment de la fonction pédagogique. En outre, la réforme n'aurait pas fait l'objet d'une concertation préalable suffisante.

Ces craintes ou ces critiques sont justifiées, au moins en partie, mais elles ne devraient cependant pas occulter la grande qualité générale de l'ensemble curriculaire mis en place, notamment des options pédagogiques fondamentales, des buts et des objectifs, de la définition des compétences de base, de l'esprit et des instruments de l'évaluation, et de la qualité du suivi des élèves, qu'ils soient spécialement doués ou qu'ils éprouvent des difficultés d'apprentissage. En fait, il est difficile de déterminer dans quelle mesure les inquiétudes et les critiques ne sont pas en partie causées par les grands efforts d'adaptation qu'exige l'élévation considérable du niveau de "professionnalisation" que les autorités cherchent à susciter, tant en ce qui concerne l'enseignement que l'évaluation de ses résultats.

• **Des indicateurs de performance**

Les Autorités Scolaires Locales (LEA) d'Angleterre s'appêtent à publier des indicateurs sur les aspects suivants :

- Taux d'encadrement (ratio maître-élèves).
- Qualification des enseignants (études faites,...).
- Habiletés pédagogiques des maîtres.
- Attitudes professionnelles des enseignants (dévouement à la tâche, satisfaction professionnelle,...).
- Qualité d'exploitation du curriculum.
- Utilisation du temps imparti.
- Niveau d'investissement des élèves dans le travail.
- Qualité des expériences d'apprentissage.
- Devoirs à domicile.
- Ponctualité, absentéisme.
- Attitudes et comportements des élèves (y compris vandalisme et graffiti).
- Coût par élève.

Concluant l'analyse de l'orientation générale que prend l'Angleterre en matière de pilotage, Gipps estime que l'évaluation liée à des objectifs spécifiques du curriculum, telle qu'elle se développe actuellement dans son pays, constituera un puissant instrument de contrôle et d'évaluation des écoles, des enseignants et des élèves.

□ France : Un développement exemplaire

Bien que la France ait fait oeuvre de pionnier en matière d'évaluation des aptitudes et des connaissances, - A. Binet et Th. Simon rendent public, en 1905, le premier test d'intelligence opérationnel portant sur les processus cognitifs supérieurs (*L'échelle métrique de l'intelligence*) et H. Piéron est le père de la docimologie (1922) -, elle a connu une époque de méfiance vis-à-vis de l'évaluation objective en matière d'éducation. Mai 1968 a exacerbé cette tendance. Après une période de transition pendant laquelle un juste équilibre entre les approches qualitative et quantitative a été recherché et, en partie, atteint, la France prend actuellement place parmi les leaders en matière de pilotage de systèmes scolaires. En effet, principalement sous l'influence de la *Direction de l'Evaluation et de la Prospective*, dont il va être question, et de quelques grands spécialistes, une "culture de l'évaluation" est apparue ces dernières années.

Par une circulaire du 24 mai 1974, le ministre J. Fontanet crée un service d'évaluation des performances éducatives, devant servir au pilotage. Ce service devient, en 1987, la *Direction de l'Evaluation et de la Prospective (DEP)*, qui joue maintenant un rôle de premier plan sur la scène nationale et internationale.

La DEP n'intervient pas dans la construction des curriculums. Toutefois, les informations qu'elle apporte constituent des éléments importants pour l'adaptation de programmes existants ou pour la conception de nouveaux.

Un document, rédigé par M. Le Guen (1991) à l'intention de l'OCDE, présente, de façon détaillée, le fonctionnement de la DEP. Les notes qui suivent en sont inspirées.

Aux fins de pilotage, la DEP réalise de vastes enquêtes normatives auxquelles tous les membres de la communauté éducative sont associés, principalement les inspecteurs, les enseignants, les chercheurs, les syndicats d'enseignants et les associations de parents. Leurs représentants respectifs constituent des comités d'accompagnement. Ils participent à la définition des objectifs des enquêtes, notamment à partir des programmes et des directives officielles.

Des tests et des questionnaires sont ensuite élaborés et essayés. A cette occasion, on recueille les commentaires des enseignants titulaires des classes où les prétests ont lieu. Dès ce moment, l'anonymat des participants est strictement respecté, la DEP ayant seule accès direct aux données.

Les échantillons d'écoles sont tirés au hasard. Ils sont stratifiés selon deux critères : l'environnement (urbain ou rural) et la taille des écoles. Dans les établissements choisis, on teste tous les élèves appartenant aux niveaux

scolaires concernés. On peut ainsi étudier l'influence de la dimension des établissements.

Fait important : grâce à l'association étroite des enseignants aux opérations, un transfert de méthodologie et de techniques se produit, ce qui permet d'espérer que des centres régionaux pourront procéder à des enquêtes particulières, conçues par des équipes locales.

Les travaux prennent trois directions principales : l'évaluation des acquis, l'évaluation de la politique éducative, la production et la diffusion d'outils d'évaluation.

• *L'évaluation des acquis pendant la scolarité*

Dès le début des années 80, la France a instauré un pilotage des acquis scolaires. Les curriculums des enseignements primaire et secondaire sont entièrement couverts et l'on tient aussi compte des apprentissages extrascolaires. Des questionnaires portent, en outre, sur les processus cognitifs, les attitudes et les opinions.

En plus de la lecture, - qui fait l'objet d'une attention particulière -, l'expression écrite, les mathématiques et des aspects de la culture générale sont explorés, dès l'école primaire. Dans l'enseignement secondaire, toutes les disciplines sont couvertes de façon récurrente.

Une grande variété d'instruments sont utilisés : QCM, questions à réponses ouvertes courtes ou longues, checklists, échelles, etc.

En outre, et surtout dans l'enseignement secondaire, la "vie dans l'école" est aussi observée. A cette fin, les élèves, les professeurs et les chefs d'établissement répondent à des questionnaires portant principalement sur l'adaptation à la vie scolaire, les relations interindividuelles, les règles de fonctionnement, l'acceptation et la perception de l'enseignement, l'aide apportée par la famille, les travaux à domicile. On examine aussi l'adaptation à l'environnement social extérieur et la qualité de leader. Un troisième aspect concerne la préparation à la vie active : informations reçues par les élèves, qualité de l'orientation.

Les écoles sont décrites et, à propos de classes particulières, on établit pour chaque élève une fiche indiquant l'âge, le sexe, la nationalité, les antécédents scolaires, la langue maternelle, les langues étudiées, la catégorie socioprofessionnelle des parents et les études qu'ils ont faites.

Enfin, des questionnaires adressés aux professeurs portent sur les disciplines enseignées, les objectifs estimés importants, l'appréciation de la difficulté des tests pour leurs élèves, et la prédiction du succès aux épreuves. Une brève description des méthodes d'enseignement est aussi demandée.

Dans tous les cas, l'anonymat reste garanti.

L'analyse des données ainsi recueillies concerne principalement :

- Le pourcentage d'élèves de chaque niveau scolaire, atteignant le critère de réussite des épreuves.

	Ensemble	10 % les plus faibles	10 % les plus forts
Français	66,7	36,7	89,7
Mathématiques	74,9	44,4	95,2

Lecture : en moyenne, les élèves ont répondu correctement à 66,7 % des questions de français.

Source : MENC DEP

Figure 5 a : Pourcentage de réussite aux questions de français et de mathématiques en fin de 5e en 1991

- La relation entre les savoirs et les savoir-faire dans une ou plusieurs disciplines.
- Les corrélations entre les performances scolaires et d'autres variables (redoublement, taille de l'école, catégorie socioprofessionnelle des parents,...).

Ces évaluations permettent notamment d'estimer la progression des connaissances au cours de la scolarité, de vérifier si des différences se manifestent entre garçons et filles, et à quels propos, et d'identifier les principaux problèmes qui se posent.

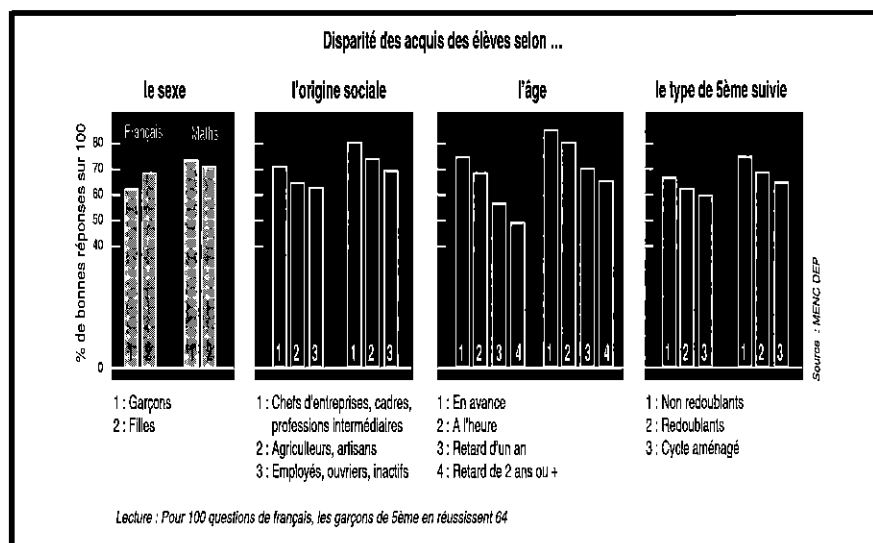


Figure 5 b : Disparité des acquis des élèves selon le sexe, l'origine sociale, l'âge et le type de 5e suivie

L'exemple qui précède, relatif à la disparité des acquis, est extrait de *L'état de l'école*, 1992, 2, p. 39. L'exemple suivant porte sur l'évolution des acquis ; il est extrait d'*Education et formations*, 1993, 34.

Tableau 8 : Evolution des acquis cognitifs et socio-affectifs des élèves entre octobre 1990 et mai 1992 (soit entre le début de la 6e et deux ans après)

Score	Octobre 1990		Mai 1992		Evolution (en proportion de l'écart-type du score de 1990)
	M	s	M	s	
Test d'ancrage en français	37,3 %	19,1 %	51,0 %	20,0 %	+ 0,717 σ
Test d'ancrage en mathématique	50,4 %	25,3 %	62,6 %	24,1 %	+ 0,482 σ
Stratégies de travail	- 0,057	0,657	- 0,064	0,700	- 0,011 σ
Organisation	0,908	0,538	0,872	0,640	- 0,067 σ
Aide de la famille pour le travail à domicile	- 0,218	0,952	- 0,574	0,964	- 0,374 σ
Temps consacré au travail à domicile :					
jours/semaine	4,5 jours	1,5 jours	4,7 jours	1,4 jours	+ 0,130 σ
durée/jour	1h17 min	52 min	1h20 min	53 min	+ 0,058 σ
Attitudes de démobilisation	- 1,017	0,539	- 0,874	0,590	+ 0,263 σ
Motivation scolaire	0,830	0,587	0,549	0,710	- 0,479 σ
Image de soi scolaire	0,249	0,692	0,36	0,704	+ 0,160 σ
Image de soi sociale	0,583	0,918	0,794	0,922	+ 0,231 σ
Goût pour la coopération	0,589	0,728	0,697	0,716	+ 0,150 σ
Assurance, sociabilité	0,207	0,239	0,163	0,288	- 0,184 σ
Défense, agressivité	- 0,047	0,617	0,072	0,634	+ 0,193 σ
Culture civique	13,66	3,99	14,41	4,14	+ 0,176 σ

Rappel :

- les scores cognitifs sont exprimés en % d'items réussis;
- les échelles socio-affectives (à l'exception de Culture Civique) sont mesurées sur une échelle allant de -2 à +2;
- le score de Culture Civique est la somme des réponses de type "démocratique" aux 24 items qui composent l'échelle.

Ce tableau présente l'évolution des acquis des élèves entre le début de sixième et la fin de la seconde année de collège (5e ou 6e selon le cas). On peut y lire, pour chacun des critères retenus, la moyenne et l'écart-type des scores relevés à chaque moment, ainsi qu'un indice d'évolution.

Ce dernier s'obtient en calculant la différence entre la moyenne des scores d'"entrée" et de "sortie", que l'on divise ensuite par l'écart-type du score d'entrée.

Cette procédure est destinée à faciliter la lecture des résultats : elle permet en effet de situer les "progrès" ou les "reculs" observés sur une même échelle, quel que soit le critère. Un indice d'évolution proche de 0 signifie ainsi qu'il n'y a pratiquement pas eu d'évolution en deux ans. Un indice positif (par exemple : + 0,250) indique que le score moyen relevé après deux ans de collège est supérieur (dans l'exemple : d'un quart d'écart-type) à celui observé en début de sixième. Un indice négatif (par exemple : - 0,500) signifie que le score moyen a baissé entre 90 et 92 (dans l'exemple : d'un demi écart-type).

- **L'évaluation de la politique éducative**

Les principales préoccupations sont l'acceptation ou le rejet des mesures décidées, l'estimation du degré auquel les objectifs estimés prioritaires sont atteints. Une attention particulière est accordée à l'efficacité des mesures prises en faveur des élèves éprouvant des difficultés dans leurs études.

Une meilleure connaissance du fonctionnement du système scolaire doit entraîner un progrès des apprentissages et permettre une plus grande efficacité du système.

- **La production et la diffusion d'outils d'évaluation**

Les enseignants qui ont été associés à la conception des instruments d'évaluation et/ou qui ont été chargés de tester leurs élèves avec ces instruments se familiarisent avec ces techniques. Il est de bonne politique de laisser, par la suite, les outils à la disposition des maîtres, afin qu'ils puissent évaluer les forces et les faiblesses de leurs élèves en début d'année scolaire et par la suite, s'ils le souhaitent. Cette familiarisation avec les méthodes et les outils d'évaluation est importante aussi pour les cadres pédagogiques, les parents, les interlocuteurs sociaux.

- **Quelques précisions sur le testage à grande échelle**

Dans le système éducatif français, l'école élémentaire comprend trois années maternelles et quatre années primaires (cours élémentaires 1 et 2; cours moyens 1 et 2), le collège compte quatre années (6e, 5e, 4e, 3e) et le lycée trois années (seconde, première, terminale).

Depuis 1989, on procède à une évaluation nationale au début du cours élémentaire 2, au début de la 6e, puis au début de la seconde.

J. Colomb (1993, pp.3-4) précise :

"A ces trois articulations très importantes du système éducatif (...), l'ensemble des élèves de chacun de ces niveaux est soumis à un ensemble d'épreuves nationales visant à évaluer ses compétences en français et en mathématiques, ces épreuves étant choisies par un groupe de pilotage national à partir des propositions du <terrain>.

Il s'agit donc d'une opération considérable : elle s'adresse chaque année à plus de deux millions d'élèves ! La première année, la totalité des données a été centralisée pour une analyse au niveau national; depuis 1990, seules les données correspondant à un échantillon représentatif de quelques milliers d'élèves, par niveau, sont centralisées et font l'objet d'analyses diverses.

Trois grandes finalités sont à l'origine de la mise en place de cette évaluation :

- Améliorer la connaissance individuelle des élèves afin de détecter, dès le début de l'année scolaire, les élèves qui présentent des lacunes dans leurs apprentissages et de pouvoir ainsi engager rapidement des procédures de remédiation, avant la mise en place de nouveaux apprentissages sur des notions mal traitées. La personnalisation de l'aide ainsi apportée aux élèves, si elle reste difficile à réaliser, constitue un moyen essentiel de la gestion de l'hétérogénéité des publics scolaires.
- Mettre à la disposition des partenaires du système éducatif des références nationales permettant de mieux situer les actions entreprises, du niveau de la classe jusqu'au niveau national. La répétition annuelle de ces opérations permet un meilleur suivi du système éducatif et donc la définition et la mise en place de procédures de régulation.
- Constituer des outils pour une "mesure" de l'efficacité de l'école qui, même si elle ne rend compte que d'une faible partie de la réalité, a le mérite de permettre une approche plus objective de la réalité du fonctionnement de l'école."

Colomb observe finalement (p.12) :

"Ces résultats sont bien reçus par l'opinion publique lorsqu'ils confirment les idées reçues et, bizarrement, souvent ignorés lorsqu'ils vont à l'encontre de ces mêmes idées reçues... Par contre, au niveau des enseignants, les outils de diagnostic qui leur sont ainsi fournis par le Ministère semblent recevoir un accueil de plus en plus positif, malgré la lourdeur des tâches matérielles qui leur sont ainsi imposées. Le risque de constitution d'une <norme nationale>, et donc d'examens de passage déguisés, accompagnés du <bachotage> correspondant, ne semble pas se manifester."

Cette dernière note optimiste est cependant contredite p.13. Observant que les évaluations nationales fournissent aux décideurs de précieux indicateurs et des outils d'analyse du fonctionnement du système éducatif (et donc de son pilotage), Colomb regrette que l'évaluation se situe de plus en plus dans ce rôle "plutôt que dans un rôle de contrôle et de régulation du système, qui devrait être le sien." Il n'y a pourtant pas incompatibilité.

• **Un remarquable ensemble de publications**

Plusieurs publications périodiques témoignent de l'activité considérable de la DEP :

— *L'état de l'école (publication annuelle)*

L'édition 1992 présente un choix de 36 indicateurs relatifs à l'activité, au coût et aux résultats du système éducatif. Cette publication annuelle rend compte au grand public de l'état du système, de son évolution et, pour certains indicateurs, de la situation de la France par rapport à d'autres pays. Le

principal souci est de répondre à deux questions : le système éducatif est-il efficace; est-il équitable ?

Pour mesurer l'efficacité, on distingue la capacité de scolariser les enfants et la capacité de leur faire acquérir les connaissances et les compétences souhaitables. Les indicateurs "d'activité" et de "coûts" apportent des éléments de réponses à ces questions. L'équité est, elle, examinée sous trois aspects importants : les établissements sont-ils équivalents, les bénéfices de l'éducation se distribuent-ils également selon le sexe et l'origine sociale ? L'édition 1992 a apporté des éclairages nouveaux sur la situation des régions dans la compétition scolaire, en publiant les taux d'accès régionaux au niveau du baccalauréat en 1991.

L'état de l'école peut être complété par des publications plus ciblées. Ainsi une *Géographie de l'école* a été publiée en 1993. Elle réunit 36 indicateurs mettant en lumière les disparités géographiques et leur évolution.

— *Repères et références statistiques sur les enseignements et la formation.*

Cette publication annuelle concerne surtout les aspects structurels du système éducatif. Chaque année, des aspects particuliers, liés à l'actualité éducative, sont traités. (Cette publication remplace les *Tableaux des enseignements et de la formation*, publiés de 1966 à 1982).

— *Vade-mecum statistique.*

Publication annuelle (disponible sur disquettes) relative aux principales données de référence sur l'éducation, sous forme de séries chronologiques.

— *Education et formation.*

Revue trimestrielle présentant, sous forme d'articles, de données ou d'études, des sujets couvrant tous les aspects du système éducatif, ainsi que son environnement économique et social. Elle traite aussi d'aspects méthodologiques dont la connaissance est nécessaire à la compréhension des résultats statistiques. En outre, des numéros spéciaux sont publiés pour diffuser des ensembles cohérents de résultats statistiques ou des études particulières.

Par exemple, le n° 22 (janvier 1990) traite des méthodes d'analyse des établissements scolaires et de leur utilité pour l'évaluation, des performances des lycéens au baccalauréat, des indicateurs d'efficacité des établissements, de l'approche sociologique du fonctionnement des lycées et de la singularité des établissements scolaires.

— *Notes d'information.*

Brefs documents, publiés une ou plusieurs fois par mois. Par exemple, les trois numéros d'octobre 1992 portent respectivement sur la taxe d'apprentissage, l'enseignement de l'éducation physique et sportive, et l'origine sociale des étudiants de 1960 à 1990.

□ Canada

Jusqu'à ces derniers temps, et en raison de l'autonomie, jalousement défendue, des différentes provinces, celles-ci ont développé leur propre système de pilotage. Le Québec servira d'exemple.

Un Conseil des Ministres de l'Éducation du Canada (CMEC) a été constitué en 1967. Il sert de forum centré sur la coopération nationale et peut lancer des projets susceptibles d'augmenter la qualité de l'éducation tant au niveau national qu'au niveau provincial. Le Projet "Indicateurs nationaux des rendements scolaires", axé sur les préoccupations communes, est l'un d'eux.

• Au Québec : Les indicateurs de l'enseignement primaire et secondaire

Depuis 1985, le Ministère de l'Éducation du Québec rend public un ensemble d'indicateurs précédés de commentaires généraux visant à mettre en évidence les points saillants.

Une section porte sur les cheminements scolaires; une autre, sur les résultats obtenus tant à des épreuves propres au Québec qu'à des épreuves internationales comme celles de l'Association Internationale pour l'Évaluation du Rendement Scolaire (IEA).

Voici un exemple de résultats publiés annuellement dans les *Indicateurs sur la situation de l'enseignement primaire et secondaire* de 1989.

Tableau 9 : Résultats à l'épreuve de français écrit de 5e secondaire selon le sexe, le réseau d'enseignement et le type de formation des élèves

	Moyenne (en %)			Pourcentage de réussite		
	1986	1987	1988	1986	1987	1988
Garçons	55,4	57,5	60,2	41,4	43,2	49,9
Filles	61,7	63,6	65,6	57,4	59,0	63,9
Réseau public	57,3	58,9	61,2	46,3	47,2	52,8
Réseau privé	65,6	68,8	71,0	66,3	71,6	77,1
Formation générale	60,2	62,1	63,9	53,3	55,0	59,5
Formation professionnelle	49,1	51,0	51,2	26,6	26,4	25,9
Ensemble	58,7	60,8	63,1	49,8	51,9	57,4

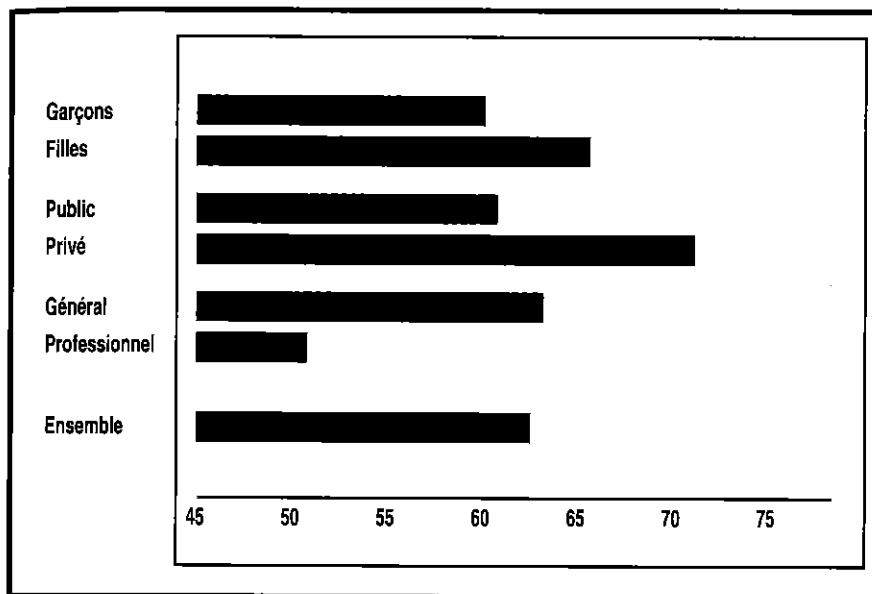


Figure 6 : Moyenne à l'épreuve de français écrit de 5e secondaire selon certaines variables, mai 1988

Ces résultats font notamment apparaître une progression régulière et générale du rendement. Celui-ci est toujours meilleur dans les écoles privées que dans les écoles publiques : les premières sélectionnent les candidats à l'admission, alors que ce n'est pas le cas des secondes.

D'autres indicateurs sont relatifs au financement, au nombre de diplômés et à l'intégration des jeunes au marché du travail.

• Les indicateurs nationaux du Canada

Ce projet, adopté en 1988 par le Conseil des Ministres de l'Éducation du Canada, a pour but premier "d'assurer une base d'informations qui permettent aux ministres provinciaux d'évaluer leurs programmes scolaires en les comparant à des normes valant pour tout le Canada." (CMEC, 1989, p. 5).

Plus spécifiquement, les objectifs généraux du projet sont les suivants :

- Permettre au Conseil des ministres de l'éducation de publier un rapport annuel présentant à la population canadienne les résultats obtenus, dans chaque province, relativement à des indicateurs de l'éducation déterminés.

- Aider les ministères à évaluer les progrès accomplis et à identifier les priorités en éducation.
- Aider les ministères à assurer aux jeunes Canadiens un enseignement de qualité, les préparant à leurs responsabilités de citoyens et de membres utiles de la société.

L'idée centrale du projet est la comparaison entre les objectifs et les attentes, d'une part, et les résultats quantitatifs et qualitatifs, d'autre part. Trois composantes sont distinguées :

1. *Les indicateurs de participation* : taux de participation, de persévérance et d'obtention du diplôme. Ils expliquent jusqu'à quel point les jeunes profitent de la possibilité de faire des études.
2. *Les indicateurs d'attente*
 - Dans quelle mesure les objectifs communs aux systèmes éducatifs sont-ils atteints ?
 - Dans quelle mesure les diplômés sont-ils satisfaits de leurs études ?
 - Qu'attend le public de la part des élèves qui terminent la scolarité obligatoire ou qui obtiennent un diplôme ?
 - Dans quelle mesure le système scolaire répond-il à l'attente du public en matière de langue et de calcul ?
 - Quel niveau de langue et de connaissances en mathématiques les élèves jugent-ils devoir atteindre pour réussir dans la vie active, après l'obtention du diplôme ?
3. *Les indicateurs de résultats scolaires*
 - Mesure dans laquelle les élèves de 13 et de 16 ans atteignent les objectifs formulés dans les programmes, et attentes du public en matière de compréhension de la langue écrite et des mathématiques. Pour la première fois, des tests nationaux vont notamment permettre à chaque province de se situer. Ces épreuves ne se substituent pas aux évaluations que les professeurs font de leurs élèves.
 - Mesure dans laquelle les résultats scolaires vont en augmentant entre 13 et 16 ans.

Le testing devra répondre aux exigences suivantes (Palmer, 1993, p.2) :

- Mesurer les performances des systèmes éducatifs provinciaux, et non celles des élèves considérés individuellement (on ne fournira pas de résultats par écoles ou par classes).
- Etre à référence critérielle.
- Seul le profil du rendement de chaque province sera publié (pas de moyennes ni de chiffres particuliers).

ier les

ement
mem-et les
part.rsévè-
int les

ucatifs

des ?
blarité

du pu-

es les
active,ent les
blic en
émati-
nment
e subs-
élèves.
nt en-

1993,

et non
de ré-

pas de

- La décision de tester la population de 16 ans a été aisément prise, car cet âge correspond à la fin de la scolarité obligatoire. Le choix de 13 ans résulte d'un compromis : comme le second testing a déjà lieu trois ans après, il est encore possible d'utiliser, au moins en partie, les mêmes instruments à des fins comparatives.
- Les enseignants sont étroitement associés à l'évaluation. Ils participent à la construction des tests et les essaient dans des classes choisies dans le pays entier.
- A l'instar de ce qui se fait aux Etats-Unis, cinq niveaux de performance seront distingués pour la lecture, l'expression écrite et les mathématiques. Un test préliminaire permet de déterminer auquel de ces cinq niveaux l'élève semble devoir être testé; chacun est néanmoins examiné à deux niveaux consécutifs pour déterminer plus sûrement ses limites.
- En mathématiques, deux types de tests seront utilisés : l'un portera sur les connaissances et l'autre sur la résolution de problèmes. Les élèves devront, en outre, répondre à un questionnaire relatif à leur école, à la langue parlée et à leur passé en mathématiques.
- En lecture, les tests comprendront quatre textes de difficulté différente; 40 à 50 questions (QCM et questions ouvertes) sont posées à leur propos.
- Les tests sont validés au point de vue du contenu par rapport au curriculum et les enseignants sont consultés à propos de la qualité des questions.

Selon les plans, les tests de rendement scolaire devront aussi porter ultérieurement sur les connaissances scientifiques et culturelles (connaissance et prise de conscience du patrimoine canadien : géographie, histoire, littérature) et sur le sens critique. Selon la tendance actuelle, le recours aux tests de performances effectives est préconisé.

□ Suède

- ***Un système précurseur : une harmonisation libérale de la notation***

Le début

Depuis longtemps, la Suède a avancé vers le pilotage de la façon la plus démocratique qui soit. Dès les années 60, le système de *modération* suivant a été adopté pour les évaluations dans l'enseignement secondaire supérieur (S. Henrysson, 1964).

D'une part, les professeurs ont continué à évaluer leurs élèves selon les critères habituels. On demandait, toutefois, d'utiliser une échelle à sept degrés et de répartir, en gros, les notes selon une distribution gaussienne. Le professeur restait cependant libre de tenir compte des caractéristiques de sa classe (par exemple, tête très forte ou presque tous élèves moyens, etc.).

Par ailleurs, on mettait à la disposition des enseignants une batterie de tests de connaissances portant sur les branches principales du programme. Cette batterie était construite chaque année par un service de recherche où collaboraient des spécialistes de l'évaluation et des professeurs expérimentés. Actuellement, les tests dont il va être question par la suite sont construits par quatre services de recherche universitaires.

Les tests, administrés à des échantillons nationaux représentatifs, étaient, eux aussi, étalonnés en sept classes, ce qui permettait aux professeurs de voir où leurs élèves se situaient par rapport à la moyenne nationale (les professeurs devaient se demander s'ils étaient plus ou moins sévères que leurs collègues d'autres écoles ou s'ils avaient affaire à des classes exceptionnellement fortes ou faibles).

Sans changer l'ordre du classement initialement établi, les professeurs étaient libres de rectifier leurs notes en fonction des comparaisons faites. Ils n'étaient pas tenus de rendre des comptes à quiconque à propos des décisions prises.

Cette attitude particulièrement libérale à l'égard des enseignants a considérablement contribué à leur donner confiance dans les tests; ils leur sont devenus de plus en plus familiers. Lorsqu'en 1993, l'utilisation des nouveaux tests nationaux, dont il va être question, n'a été rendue obligatoire que pour l'enseignement secondaire supérieur, 95 % des professeurs du primaire et du secondaire inférieur ont décidé de les utiliser aussi (Wedman, 1993, p.7).

Aujourd'hui

Actuellement, le système de notation est, en gros, le suivant (S. Marklund, 1991).

Pendant les neuf années de scolarité obligatoire, les professeurs notent leurs élèves de façon tout à fait libre et ne font jamais échouer (pas de redoublements). Pour les 4^e et 7^e années, ils disposent de tests diagnostiques pour la langue maternelle et les mathématiques (évaluation purement formative). En 8^e année, les élèves subissent un test standardisé en anglais, langue étrangère. En 9^e année, des tests standardisés en langue maternelle et en mathématiques sont administrés. Les résultats obtenus à ces tests n'ont qu'une valeur indicative pour les enseignants : ils voient comment leurs élèves se situent par rapport à ceux d'autres écoles, à propos d'apprentissages jugés fondamentaux.

A la fin de l'enseignement secondaire supérieur (12e année, 19 ans, pour l'enseignement général pré-universitaire; 11e année, pour l'enseignement "général technique, professionnel ou social"), l'évaluation finale se déroule de la façon suivante:

1. Les professeurs font passer des examens de façon traditionnelle à leurs élèves. Les épreuves sont notées, mais les résultats ne sont pas communiqués.

Les maîtres utilisent une échelle de notation à cinq degrés, en répartissant *grosso modo* les notes de façon gaussienne :

Notes	1	2	3	4	5
Pourcentage	7 %	24 %	38 %	24 %	7 %

2. En avril ou en mai, les élèves de l'enseignement général subissent des tests nationaux, étalonnés aussi en cinq classes, en langue maternelle, mathématiques, physique, chimie et langues étrangères (anglais et, au choix, allemand ou français); les élèves du groupe technique sont testés en langue maternelle et en anglais.
3. Un conseil des professeurs (ayant autorité légale) se réunit pour examiner les notes que chacun de leurs collègues proposent en fonction des résultats obtenus aux deux épreuves mentionnées ci-dessus. En principe, les notes ne peuvent pas s'écarter de plus 0,2 points, en plus ou en moins, des scores obtenus aux tests nationaux. Si le professeur demande un écart supérieur à 0,2, - ce qui est rare -, il doit déposer une justification écrite. Le conseil décide finalement des notes attribuées pour les branches sur lesquelles les tests nationaux ont porté. Pour les autres branches, le professeur reste libre.

• *Evaluation nationale*

Après dix ans d'études préparatoires, la Suède a, elle aussi, instauré une évaluation nationale, à partir de 1989. Les savoirs et les savoir-faire relatifs à la totalité du curriculum de la scolarité obligatoire sont évalués, de même que la méthode de travail des élèves. Cette évaluation se veut formative et est donc aussi destinée à aider les professeurs dans leur tâche. Deux questions centrales sont posées : "Les écoles suédoises travaillent-elles comme elles devraient le faire ? Les objectifs assignés aux écoles sont-ils atteints ?".

L'évaluation de 1989 a porté sur la 2e année primaire (langue maternelle, mathématiques, musique et autres arts), sur la 5e année (comme en 2e, plus l'éducation civique, les sciences et l'anglais). En 1992, l'évaluation a por-

té sur la 3e année de l'enseignement secondaire. De telles évaluations seront répétées tous les trois ans.

Par comparaison avec des programmes de testing prenant peu de temps, - comme ceux de la NAEP -, le programme suédois en consomme beaucoup. Chaque école doit consacrer environ vingt heures pour faire passer les épreuves, et il faut à peu près autant de temps à chaque professeur désigné pour les corriger. Selon Wedman (1993, p. 10), "on estima qu'un tel investissement était raisonnable tant les profits que tous les participants à l'opération en retireraient seraient élevés; un tel effort pédagogique en valait la peine."

Tous les enseignants reçoivent les résultats de leurs élèves ainsi que les résultats nationaux.

En plus de cette évaluation quantitative, on procède tous les ans à une évaluation qualitative des écoles.

L'ensemble des évaluations réalisées se traduit en un rapport triennal au gouvernement sur l'état de l'enseignement. Le gouvernement dispose d'une année pour faire connaître les mesures qu'il a décidé de prendre en fonction de ce rapport.

Ce riche programme d'évaluation et de pilotage porte à la fois sur les processus et sur les résultats, et allie harmonieusement les approches quantitative et qualitative. Comme le conclut A. Ruby (1991, p.19) : "Ce programme bien intégré, où les responsabilités et le jeu de l'autorité sont nettement définis, permet une grande liberté; elle aide à tenir compte des variations locales et des différences, tout en permettant d'évaluer les progrès vers les objectifs nationalement assignés".

□ **Australie**

L'Australie possède une grande expérience des enquêtes évaluatives sur l'éducation. Ce pays fut l'un des premiers du monde à se doter d'un centre national de recherche pédagogique, le *Conseil australien pour la recherche en éducation (ACER)*, fondé en 1930. Très tôt, l'ACER a organisé de vastes *surveys* du rendement de l'enseignement. Le premier (1931) a porté sur la lecture et l'arithmétique; quelque 40 000 élèves, appartenant à 700 écoles différentes, ont été testés.

Les Etats qui constituent le Commonwealth d'Australie jouissent d'une large autonomie, notamment dans le domaine de l'éducation. Pratiquement, tous ont maintenant développé leur propre système de pilotage. En se fondant sur l'expérience ainsi acquise, la *Conférence australienne des directeurs généraux de l'éducation* a formulé un projet national permettant d'évaluer les progrès de l'éducation, spécialement celui des performances.

• **Le pilotage des écoles et des systèmes scolaires dans les Etats.**

- L'Etat de Nouvelle-Galles du Sud teste tous les élève de 4e année du secondaire en langue maternelle, en mathématiques et en sciences. Depuis 1989, un échantillon des élèves de 3e et tous les élèves de 6e primaire sont testés en lecture et en arithmétique.
- L'Etat de Victoria teste, depuis 1988, un échantillon de tous les élèves de 10 et de 14 ans, en lecture, en expression écrite et en mathématiques. Les élèves répondent aussi à un questionnaire relatif à la qualité de la vie à l'école.
- En Australie Occidentale, des tests de rendement en lecture, en expression écrite, en mathématiques et sur le curriculum de l'enseignement secondaire inférieur sont construits en tenant compte des curriculums locaux, afin de permettre un pilotage formatif.

Dans le même Etat, des indicateurs destinés à informer sur la santé du système sont collectés (T. Wyatt, 1988, p.15) :

1. *Indicateurs contextuels*

- Pourcentage du temps scolaire consacré à diverses composantes du curriculum (langue maternelle, mathématiques, etc.), en 3e année primaire et en 2e, 4e et 6e années de l'enseignement secondaire.
- Pourcentage d'élèves inscrits dans les écoles, selon le sexe et la situation géographique.
- Taux de mobilité des enseignants.

2. *Ressources*

- Pourcentage du budget de l'Etat consacré à l'éducation et montant du budget.
- Traitement des enseignants à différents moments de leur carrière.
- Dépenses courantes par élève.

3. *Processus*

- Nombre de jours de formation continuée offerte aux enseignants et nombre de jours effectivement utilisés à cette fin, par école, par type d'école et selon la situation géographique.
- Nombre d'établissements qui ont mis en place une structure participative communautaire (par type d'école et selon la situation géographique).

4. *Résultats*

- Distribution des notes scolaires, en 2e et en 4e années de l'enseignement secondaire, pour certains aspects cruciaux des curricu-

lums de langue maternelle et de mathématiques, et d'autres disciplines jugées importantes.

- En 6e année du secondaire, pourcentage d'élèves admissibles dans l'enseignement supérieur.
- Taux d'absentéisme dans les écoles primaires et secondaires, selon le sexe, l'origine ethnique et la situation géographique.

Afin de pouvoir disposer d'indicateurs de performance, les activités ou les projets dont le ministère prendra l'initiative devront être accompagnés :

- D'une définition opérationnelle des objectifs.
- De la liste des indicateurs de performance qui seront collectés.
- D'une définition de la compétence minimale.
- D'une spécification des données à collecter, des analyses à faire et des modalités des rapports à fournir.
- D'une description de l'utilisation des ressources disponibles.
- D'un calendrier.

Ce modèle général devrait permettre des approches à la fois formatives et sommatives.

• **Le projet national**

Ce projet a commencé par une recherche des indicateurs qui pourraient être utilisés au niveau national : leur adéquation, les problèmes que soulève leur interprétation, le coût de l'entreprise et le profit que les participants au projet pourraient en retirer ont été soigneusement examinés.

Ces différents aspects ont fait l'objet de monographies largement diffusées. Elles portent notamment sur les concepts, les définitions et les problèmes existant à ces propos, sur les leçons à tirer de l'expérience relative aux indicateurs sociaux. Différents indicateurs sont discutés, tant en ce qui concerne leur nature que leur utilité potentielle.

Depuis 1990, l'Australie publie annuellement un **Rapport national sur l'état de l'éducation**. Il doit servir au pilotage :

- du système éducatif global;
- du parcours scolaire des élèves;
- de la qualité des curriculums;
- des performances des élèves;
- de l'utilisation des ressources budgétaires.

□ Autres pays

Parmi les pays où le pilotage est de mieux en mieux réalisé, il faut citer le Danemark, la Nouvelle-Zélande, le Japon, Singapour. Cette liste s'allonge sans arrêt.

Par exemple, dans sa publication *Facts and figures. Education indicators* (1993), le Danemark présente 37 indicateurs (17 relatifs aux ressources, 11 à la participation et 9 aux résultats). Des séries s'étendant sur dix ans indiquent les grandes tendances. Les données relatives à l'entrée des jeunes dans la vie active sont aussi d'un grand intérêt.

II. Indicateurs internationaux

Prendre le monde comme laboratoire pédagogique où, dans une vision expérimentale, les caractéristiques des systèmes éducatifs seraient les variables indépendantes et les rendements scolaires les variables dépendantes, est l'idée générale qui sous-tend les recherches comparatives sur ce que produisent des systèmes existant dans des contextes socioculturels parfois très différents. De telles études relèvent soit de la recherche fondamentale, soit de recherches destinées à aider à la formulation des politiques de l'éducation. Jusqu'à présent, ce second espoir a été souvent déçu.

"La liaison entre les descriptions de performances scolaires et les problèmes de politique éducative est l'un des aspects les moins satisfaisants de la littérature relative aux enquêtes évaluatives internationales. Tout en respectant les contraintes inhérentes à la méthodologie des *surveys* à grande échelle, les chercheurs devraient s'efforcer de concevoir les études et les analyses de façon à répondre à des questions que les décideurs en matière d'éducation se posent effectivement. Par exemple, si l'on s'intéresse aux programmes couronnés de succès, - c'est-à-dire à la politique éducative et aux conditions qui semblent associés à des performances élevées -, des méthodes appropriées et des questions adéquates doivent être introduites dans la conception de l'enquête. Toutefois, les pratiques couronnées de succès dans un pays donné ne produisent pas nécessairement les mêmes résultats dans un autre. Des différences culturelles et les façons de concevoir l'apprentissage et l'enseignement sont de puissants déterminants. Les *surveys* internationaux peuvent et doivent aider à découvrir des aspects favorables de l'enseignement et de l'apprentissage qui ne sont pas nationalement liés et dont, par conséquent, les politiques éducatives peuvent tenir compte. Il peut aussi exister des aspects qui n'intéressent que des sous-ensembles de pays." (U.S. Department of Education, 1992, pp. 45-46).

Quoi qu'il en soit, on sait aujourd'hui que les études internationales et interculturelles de rendements scolaires peuvent produire des données qui constituent des indicateurs de plus en plus importants de la mesure où les nations réussissent à éduquer les individus et à les préparer à jouer leur rôle dans le monde du travail et de la citoyenneté. Comme le soulignent N. Bottani et H. Walberg (1992, p.9), le but n'est pas de classer les pays entre eux, mais bien de fournir aux décideurs politiques une compréhension plus profonde des facteurs qui influencent la qualité de l'éducation, et de leur permettre d'explorer une gamme plus large d'options politiques.

Arrivé à ce point, rappelons que collecte d'indicateurs et pilotage ne sont pas synonymes. Les indicateurs sont des informations utiles au pilotage, c'est-à-dire, en dernière analyse, à la prise de décisions qui doivent se traduire dans des actions. Rien ne garantit que l'existence d'indicateurs, fussent-ils de la meilleure qualité, entraînera des actions.

□ L'Association Internationale pour l'Evaluation du Rendement Scolaire (IEA)

Née à la fin des années 50, à l'initiative d'une douzaine de spécialistes de la recherche expérimentale en éducation, l'IEA est aujourd'hui, avec plus de cinquante pays membres, une organisation mondiale de recherche évaluative.

L'objectif initial était de contrôler objectivement les assertions, sinon les clichés, relatifs aux vertus et aux faiblesses des divers systèmes d'éducation. D'où l'idée d'une évaluation comparative des rendements scolaires à l'aide d'enquêtes normatives (*surveys*). L'essor de l'éducation comparée qui, avec l'IEA, prend une orientation expérimentale, et l'arrivée de l'ordinateur ne sont pas pour rien dans cette entreprise pédagogique impliquant le traitement de millions de données.

Depuis 1961, l'IEA lance une nouvelle évaluation internationale à peu près tous les trois ans, de façon à décrire un cycle couvrant les disciplines principales tous les douze ans. L'intérêt de ces données comparatives pour le macropilotage est capital.

Une étude utilise les instruments suivants :

- Des tests de connaissances, construits après une analyse comparative des curriculums des pays participants. Ce qui leur est commun et est estimé important constitue la matière des épreuves internationales. Les participants sont libres de tester, à la même occasion, des matières qui leur sont propres ou qu'ils privilégient.

- Des tests de performances effectives : expériences à réaliser en laboratoire, manipulations, observations, productions,...
- Un questionnaire international conçu pour collecter un grand nombre d'informations sur les élèves : sexe, âge, année d'études, milieu familial, nombre d'heures de cours suivis dans la branche en question, durée des travaux à domicile, temps consacré à la télévision,...
- Des tests d'aptitude verbale, numérique,...
- Des échelles d'attitudes permettant d'apprécier comment les élèves perçoivent la discipline et la façon dont elle est enseignée, et l'importance qu'elle a, à leur avis, pour la société.
- Les enseignants répondent à un questionnaire portant sur leur âge, leur formation initiale et continuée. Ils sont invités à indiquer, pour chaque question des tests, si la matière sur laquelle elle porte a été enseignée (récemment ou non, de façon intensive ou non, ...)
- Les chefs d'établissement répondent à un questionnaire relatif à la population de leur école, aux équipements, au financement, aux relations avec la communauté éducative,...
- Un questionnaire adressé au pouvoir organisateur.

• **Les débuts de l'IEA : Une étude de faisabilité**

Un projet pilote a été lancé en 1959 et a duré trois ans. Y participèrent l'Angleterre, la Belgique, l'Écosse, les États-Unis, la Finlande, Israël, la Pologne, la République Fédérale d'Allemagne, la Suède, la Suisse et la Yougoslavie. Il a porté sur les enfants de 13 ans, âge où la scolarité était encore obligatoire dans tous ces pays. Ainsi, 9918 élèves parlant huit langues différentes subirent les mêmes tests de compréhension de la lecture, de mathématiques, de sciences, de géographie et d'intelligence non verbale. Les résultats de cette recherche furent assez riches pour justifier des projets plus ambitieux encore.

Enquête sur le rendement de l'enseignement des mathématiques (1961-1967)

Douze pays, dont la France et la Belgique, ont participé à cette première enquête normative portant sur trois populations : élèves âgés de 13 ans, élèves étudiant au niveau scolaire normalement prévu pour les enfants de 13 ans, élèves fréquentant la classe précédant l'enseignement supérieur.

Au moment où l'enseignement des mathématiques se modifiait profondément en raison de l'introduction de la mathématique ensembliste, il était d'un grand intérêt de disposer de données permettant des comparaisons ultérieures.

La moisson s'avéra riche, qu'il s'agisse des conclusions relatives :

- aux systèmes d'enseignement : durée de la scolarité, nombre d'élèves par classe, programmes d'études, effets de la sélection, avance de certains pays par rapport aux autres;
- à la formation des enseignants;
- aux facteurs sociaux de la réussite scolaire;
- aux différences entre sexes.

C'est à cette occasion que le Japon se révéla le pays de loin le plus performant, surtout dans l'enseignement secondaire long, même dans les sections où les mathématiques n'étaient pas branche principale.

Le tableau suivant, qui montre comment les pays se situent les uns par rapport aux autres, est éloquent (G. de Landsheere et T.N. Postlethwaite, 1969, p.36). On observe une grande homogénéité dans les résultats obtenus par les quatre populations, dans tous les pays.

Tableau 10 : Moyennes nationales situées par rapport à la moyenne générale
(en unités d'écart type)

	Elèves de 13 ans	Classe où se trouve la majorité des élèves âgés de 13 ans	Fin du secondaire supérieur	
			Mathématiques fortes	Mathématiques
1.25	<ul style="list-style-type: none"> • Angleterre • Japon • Hollande • Belgique 	<ul style="list-style-type: none"> • Angleterre • Japon • Ecosse 	<ul style="list-style-type: none"> • Japon 	<ul style="list-style-type: none"> • Japon
1.00	<ul style="list-style-type: none"> • Ecosse • Australie • Etats-Unis • France 	<ul style="list-style-type: none"> • Israël • Belgique • Etats-Unis • France • Australie • Hollande • Allemagne 	<ul style="list-style-type: none"> • Angleterre • Belgique • Etats-Unis • Suède • France • Australie 	<ul style="list-style-type: none"> • Angleterre
0.75	<ul style="list-style-type: none"> • Suède • Finlande 	<ul style="list-style-type: none"> • Suède • Finlande 	<ul style="list-style-type: none"> • Ecosse • Allemagne • Finlande • Israël • Hollande 	<ul style="list-style-type: none"> • Ecosse-Belgique • Etats-Unis • Finlande • Allemagne
0.50				<ul style="list-style-type: none"> • Suède

L'observation de l'accroissement du rendement de l'enseignement des mathématiques en section "forte", entre 13 ans et la fin de l'enseignement secondaire, est aussi intéressante.

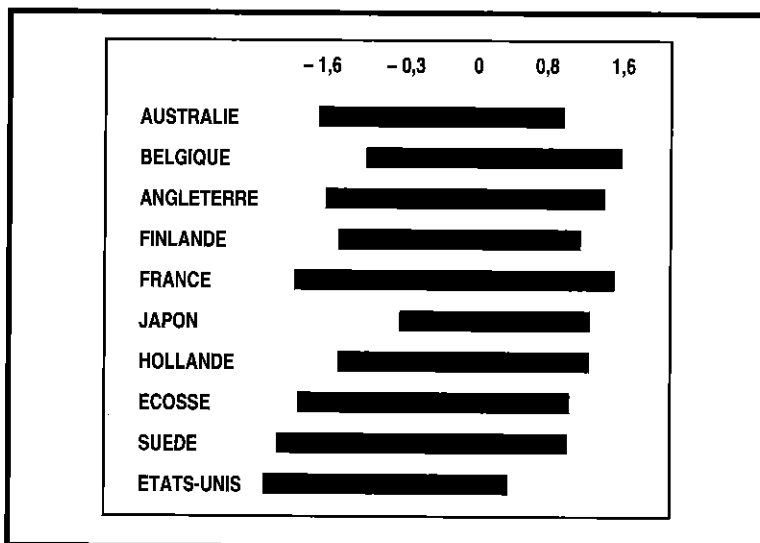


Figure 7 : Mathématiques fortes

Le Japon semble moins progresser que les autres pays, mais c'est parce qu'il part d'un niveau plus élevé (effet de plafonnement). La Belgique enregistre un progrès spectaculaire, dû au degré accusé de la sélection spontanée caractérisant le choix des classes de "mathématiques fortes".

Une enquête menée au Japon pour tenter d'expliquer ses performances brillantes, malgré un nombre d'élèves généralement très élevé par classe (souvent plus de 40), a fait apparaître deux caractéristiques majeures :

- Un enseignement impositif très exigeant.
- Une pression continuellement exercée sur les élèves par leurs parents, les maîtres et la société civile.

Les retombées de cette recherche furent considérables : affinement des techniques d'échantillonnage et d'analyse statistique, émergence d'une éducation comparée expérimentale, constitution d'un vaste réseau coopératif international dynamisé par les chercheurs en éducation les plus avancés à l'époque, stimulation de la recherche expérimentale en éducation dans les pays participants.

Sur cette recherche, voir T. Husén (1967), G. de Landsheere (1967), G. de Landsheere et T.N. Postlethwaite (1969), G. Henry (1970).

Le survey de six disciplines (1966-1973)

Cette enquête, à laquelle 22 pays ont participé, a porté simultanément sur la compréhension de la lecture, la littérature, les sciences, l'anglais ou le français enseignés comme langues étrangères, l'éducation civique. La Belgique francophone a pris part aux enquêtes sur les sciences, la lecture, la littérature, l'anglais langue étrangère, l'éducation civique.

Des échantillons nationaux des élèves de 10 et 14 ans, et des élèves fréquentant l'année pré-universitaire ont été testés. Pour chaque discipline et pour chacune des populations, 200 à 500 variables ont été étudiées, et la quantité de données a dépassé 150 millions. Elles ont, par la suite, constitué une banque mise à la disposition du monde de la recherche.

Le retentissement de ces enquêtes fut et reste considérable. Dans plusieurs pays participants, l'opinion publique se saisit des conclusions et, à leur lumière, des réformes de programmes scolaires ont été entreprises, notamment en Hongrie et en Suède.

— La recherche sur l'enseignement des sciences Un exemple d'observation grave pour la Belgique

Lors de cette recherche, un phénomène inquiétant s'est manifesté : au point de vue des connaissances, les élèves belges apparaissent comparativement faibles et, surtout, leur intérêt pour les sciences et les activités scientifiques diminue à mesure qu'ils avancent dans leurs études.

Dans l'échelle conçue pour estimer l'intérêt, un score de +15 correspond à un intérêt très élevé, un score de -15, à un très faible intérêt. En consultant le tableau ci-dessous, il semble, au premier abord, que l'intérêt est plutôt faible partout. La comparaison entre pays révèle cependant de grandes différences.

Tableau 11 : Intérêt pour les sciences

Pays	Population I		Population II		Population IV	
	Moyenne	Sigma	Moyenne	Sigma	Moyenne	Sigma
Allemagne	-1,5	-4,7	-3,7	-4,4	-4,5	-4,4
Angleterre	-3,6	-4,2	-5,0	-4,5	-5,1	-4,6
Australie	—	—	-4,1	-4,5	-4,6	-4,5
Belgique (Fl.)	-3,0	-4,3	-5,2	-4,2	-5,3	-4,1
Belgique (Fr.)	-3,8	-4,1	-4,7	-4,2	-6,1	-4,5
Ecosse	-3,4	-4,1	-5,9	-4,7	-5,7	-4,7
Etats-Unis	-1,6	-4,7	-3,9	-4,6	-4,7	-4,7

Tableau 11 : Intérêt pour les sciences

Pays	Population I		Population II		Population IV	
	Moyenne	Sigma	Moyenne	Sigma	Moyenne	Sigma
Finlande	-4,2	-3,8	-6,2	-3,9	-5,5	-4,1
France	—	—	—	—	-3,6	-4,4
Hollande	-3,1	-4,1	-5,6	-4,1	-5,6	-4,3
Hongrie	-2,2	-4,2	-1,9	-4,2	-3,0	-4,3
Italie	-3,8	-4,0	-3,5	-4,2	-4,2	-4,3
Japon	-1,2	-4,7	-4,2	-4,7	—	—
Nouvelle-Zélande	—	—	-4,3	-4,2	-4,0	-4,7
Suède	-3,9	-3,4	-4,9	-3,7	-5,2	-4,1

G. Henry (1976, pp.155-157) conclut : ici encore, la Belgique francophone est en position peu enviable : 9e sur 12 en Population I, 8e sur 14 en Population II et 14e sur 14 en population III.

— La recherche sur la littérature

Cette recherche internationale comparative sur l'étude de la littérature est une première. Dix pays y ont participé. Les populations examinées ont été les élèves de 14 ans et ceux qui fréquentaient l'année scolaire pré-universitaire.

Les deux groupes ont subi un test à choix multiple portant sur deux courtes nouvelles (l'une commune à tous et l'autre tirée au hasard entre trois). Les élèves ont aussi répondu à un questionnaire portant sur leur intérêt pour la littérature, sur leurs préférences et, plus généralement, sur leur intérêt pour la lecture.

La diversité des cultures et des patrimoines littéraires n'a pas permis l'élaboration d'un test commun. C'est pourquoi les mesures ont essentiellement porté sur les aspects cognitifs et affectifs des réactions des élèves vis-à-vis de la littérature, et sur leurs intérêts, leurs attentes.

Voir A.C. Purves, 1973.

— La recherche sur l'éducation civique

Dans la période de crise grave et de remise en cause des valeurs traditionnelles que les sociétés occidentales traversaient déjà, il a semblé important d'évaluer les connaissances et les attitudes sociopolitiques des jeunes citoyens en devenir. Deux populations ont été testées, celles des élèves fréquentant les 3e et 6e années de l'enseignement secondaire.

Des tests à choix multiple portant sur les sujets suivants ont été utilisés :

- Fondement et nature de la citoyenneté.
- Institutions politiques - matières nationales.
- Processus politiques - matières nationales.
- Processus et institutions politiques - matières internationales.
- Processus et institutions économiques.
- Institutions et processus sociaux.

Une version différente des questionnaires a été conçue pour chacune des deux populations. Certains items étaient cependant communs afin de disposer d'un ancrage permettant les comparaisons.

Des échelles affectives ont aussi été proposées. Elles portaient sur des aspects comme : Populaire / Impopulaire; On peut faire confiance / On ne peut pas faire confiance; Puissant / Faible; Sympathique / Antipathique.

Enfin, les élèves ont répondu à un questionnaire de *background* détaillé.

Parmi les observations ainsi réalisées, on note :

- Peu de différences selon les sexes.
- Une corrélation positive générale et élevée entre les résultats aux tests et le niveau socioculturel d'origine. Une relation du même type existe aussi avec le niveau d'aspiration des élèves (profession qu'ils espèrent exercer; durée des études qu'ils comptent faire). La fréquence des discussions relatives aux problèmes de la vie sociale et politique que les jeunes ont dans leur famille est aussi déterminante.

Cette recherche confirme l'importance du rôle que l'école peut jouer dans la préparation du futur citoyen. "L'information et la formation civiques ne pourront vraiment se réaliser que si l'on consent à aborder franchement les matières sociales et politiques, à condition, bien entendu, de garantir un véritable pluralisme idéologique. (...) Quelques directives ne suffisent pas à atteindre ce but. Il importe d'élaborer un curriculum diversifié, du point de vue tant des contenus que des méthodes. En particulier, on se souciera au moins autant des attitudes que des connaissances" (Massoz et Henry, 1980, p.202).

Sur les surveys des six disciplines, voir aussi J.B. Carroll (1975), L.C. Comber et J. Keeves (1973), A. Grisay et G. Henry (1973), A. Grisay (1974), G. Henry (1974), G. Henry (1975), G. Henry (1976), A. Grisay (1976), M.T. Loret (1980), A. C. Purves (1973), R.L. Thorndike (1973), J.V. Torney *et al.* (1976), D.A. Walker (1976).

La 2e enquête sur l'enseignement des mathématiques (1980-1986)

Les objectifs de cette étude étaient :

- Décrire les programmes de mathématiques appliqués dans les systèmes scolaires et examiner les changements intervenus depuis 1960.
- Examiner dans quelle mesure le curriculum prévu était effectivement implanté.
- Evaluer les rendements scolaires en mathématiques et étudier l'influence relative de différents déterminants.
- Evaluer l'augmentation des connaissances pendant une année scolaire et rechercher les raisons des différences d'accroissement chez les élèves considérés individuellement et dans des classes entières.
- Réunir, de façon intensive et extensive, des informations sur la didactique mise en oeuvre.

Les curriculums des pays participants ont été analysés, en particulier pour identifier les objectifs poursuivis et estimer s'ils étaient atteints. La façon dont les mathématiques étaient enseignées a aussi été examinée.

La Belgique a participé à cette recherche et s'est ainsi dotée d'un point de comparaison précieux.

Publication : N. Deltour et G. Henry (1988).

La 2e enquête sur les sciences (1984)

Les principaux objectifs de cette étude étaient d'examiner l'évolution de l'enseignement des sciences à travers le monde (surtout depuis 1970, année de la première évaluation) et de déterminer les facteurs qui expliquent les différences de rendement et d'autres produits de l'enseignement des sciences.

Faute de financement, la Belgique n'a pas pu participer à cette étude (Voir M. Rosier et J. Keeves, 1991). C'est d'autant plus regrettable que, comme le montre le graphique suivant (J. Keeves, 1992, p.14), on voit apparaître des phénomènes particulièrement importants à un moment où les sciences et les technologies explosent et pèsent lourd pour l'avenir des nations. Ce graphique permet, en effet, la comparaison du rendement de l'enseignement des sciences aux élèves de 14 ans, entre 1970-71 et 1983-84.

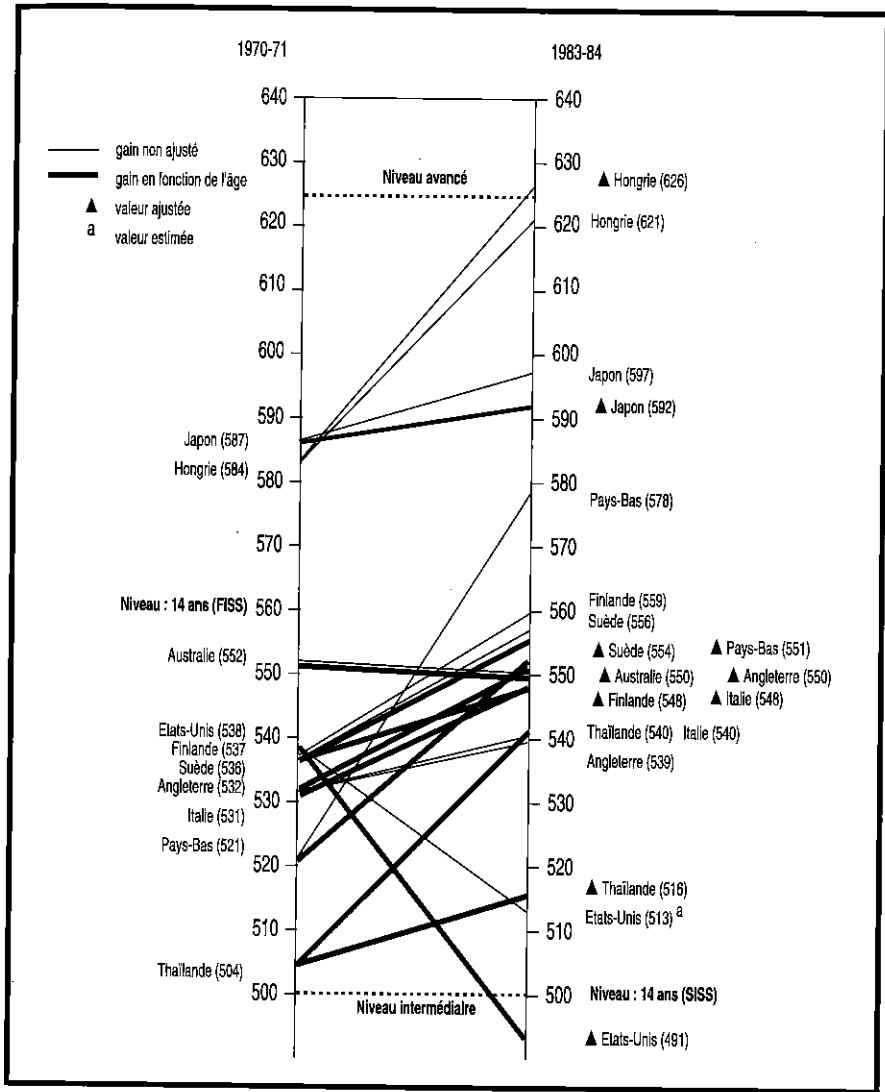


Figure 8 : Scores moyens en sciences - Elèves de 14 ans - 1970-71 et 1983-84

Que les scores soient ajustés ou non en fonction de l'âge, on constate un grand progrès en Hongrie. Ce pays a profondément remanié son programme d'enseignement des sciences en fonction des résultats observés en 1970-71; en particulier, la place des travaux pratiques en laboratoire a été fortement augmentée.

Les résultats du Japon restent très élevés, mais les progrès sont moins spectaculaires qu'en Hongrie. Peut-être voit-on s'esquisser, dès ce moment, un phénomène de plafonnement qui se produit actuellement au Japon. Il se pourrait qu'il subisse les inconvénients d'un enseignement trop impositif à un moment où le flou, le plus ou moins indéterminé, interviennent de plus en plus, et donc où la divergence prend une importance toujours accrue.

Par ailleurs, et contrairement à la tendance générale, les Etats-Unis sont en régression accusée. Cette observation a suscité une inquiétude considérable dans le pays et a fortement influencé les prises de position spectaculaires du Président Bush; elles se sont exprimées dans la publication retentissante de *A Nation at Risk*, à laquelle il a déjà été fait allusion.

Cet exemple montre bien le profit et les enseignements que peut apporter un pilotage pouvant s'appuyer sur des données internationales comparatives.

Etude sur l'utilisation pédagogique des ordinateurs (1987-1990)

OBJECTIFS :

- Déterminer comment et dans quelle mesure les ordinateurs sont utilisés, avec quels résultats.
- Observer l'évolution des matériels et des logiciels.
- Observer comment les enseignants sont formés à l'utilisation pédagogique de l'ordinateur.

Dix-neuf pays, dont la Belgique, ont participé à cette évaluation. Il apparaît que, jusqu'à présent, le recours à l'ordinateur pour les apprentissages scolaires reste généralement faible (Voir H. Pelgrum et T. Plomp, 1991).

Etude sur l'éducation préscolaire : la qualité de vie des enfants de quatre ans (1986-1996)

Cette recherche marque un tournant dans les options méthodologiques de l'IEA. En effet, l'approche qualitative domine ici. Par ailleurs, alors que les enquêtes portant sur les écoles primaires ou secondaires avaient affaire à des milieux institutionnellement homogènes, la recherche sur le préscolaire doit tenir compte non seulement de la grande hétérogénéité des

solutions apportées pour accueillir les jeunes enfants en dehors de leur famille, mais aussi du fait que le lieu d'accueil peut varier plusieurs fois au cours d'une même journée. Est-il possible de déterminer l'impact différentiel de ces divers milieux sur le développement de l'enfant? La population choisie est celle de 3:6 à 4:6 ans.

OBJECTIFS :

- Déterminer les besoins en modalités d'accueil des enfants de 4 ans et observer l'utilisation qui est faite des possibilités existantes.
- Déterminer quels facteurs (culturels, socio-économiques,...) déterminent le choix des parents.
- Quelles sont les caractéristiques des environnements d'accueil? Quelle est la qualité de vie des enfants dans ces lieux?
- Dans quelle mesure les valeurs et les attentes des adultes déterminent-elles l'organisation des lieux d'accueil et les activités qui s'y déroulent?
- Quelles différences observe-t-on entre pays?

On constate, par exemple, que le temps quotidien que le père passe seul avec l'enfant est de 0,30 h en Belgique, de 0,42 aux Etats-Unis, 0,48 en Finlande et 0,54 en Chine. Autre observation frappante : en Belgique, 40% des enfants de quatre ans passent quotidiennement par deux lieux d'accueil différents, en dehors de leur domicile (4,4% en Thaïlande, 16% aux Etats-Unis,...). Il s'agit évidemment de moyennes.

- Quelles sont les conséquences sur le développement de l'enfant?
- Dans quelle mesure les enfants sont-ils ainsi préparés à la scolarité?

Cette enquête est réalisée en trois phases :

- Phase 1 : Enquêtes sur les familles, en particulier sur l'utilisation qu'elles font des différentes possibilités d'accueil de leurs enfants.
- Phase 2 : Un échantillon des différents milieux d'accueil, identifiés au cours de la Phase 1, sera étudié de façon approfondie, afin d'évaluer la qualité de vie que les enfants y trouvent.
- Phase 3 : Etude de suivi sur les enfants de 7 ans : comment se sont-ils développés?

Voir P. Olmsted et D. Weikart, 1989.

Recherche sur la compréhension de la lecture (1990-1993)

Cette recherche a porté sur deux populations : 9 et 14 ans; 210 000 élèves ont été testés dans 32 pays. L'âge de neuf ans a paru spécialement intéressant parce que l'on est en droit de penser que quatre ans de scolarité suffisent normalement pour assurer une maîtrise élémentaire de la lecture. Il



faut insister sur le mot *élémentaire*, car l'enseignement systématique de la lecture devrait continuer pendant toutes les études secondaires. L'âge de quatorze ans a été retenu parce qu'à ce moment, l'enseignement est encore obligatoire dans tous les pays participants.

Les objectifs généraux étaient :

- Produire des instruments valides pour l'évaluation de l'habileté en lecture (niveau de littérisme).
- Estimer le pourcentage de chaque population qui atteint un niveau déterminé en lecture : élémentaire - intermédiaire - élevé.
- Déterminer l'importance des activités de lecture dans les deux populations.
- Estimer la qualité des programmes d'enseignement de la lecture dans les pays participants.
- Déterminer la relation entre le niveau d'habileté en lecture et les indicateurs économiques.
- Etablir, en 1990, une ligne de base pour le pilotage dans les pays participants.
- Comparer les pratiques d'enseignement.

Les tests ont porté sur trois types de textes :

- Narrations : courtes histoires faisant intervenir des animaux.
- Exposés. Par exemple, description du sable mouvant, suivie de questions portant sur la façon de le reconnaître.
- Documents. Lecture d'informations structurées sous forme de graphiques, de tableaux, de cartes, de modes d'emploi.

Les résultats obtenus pour la population de Belgique francophone de 9 ans se situent dans la bonne moyenne; en revanche, ceux qu'obtient la population de 14 ans sont alarmants pour les trois types de textes (Voir W.B. Elly, 1992).

Le graphique ci-dessous montre la distribution des résultats pour la lecture de textes narratifs à quatorze ans.

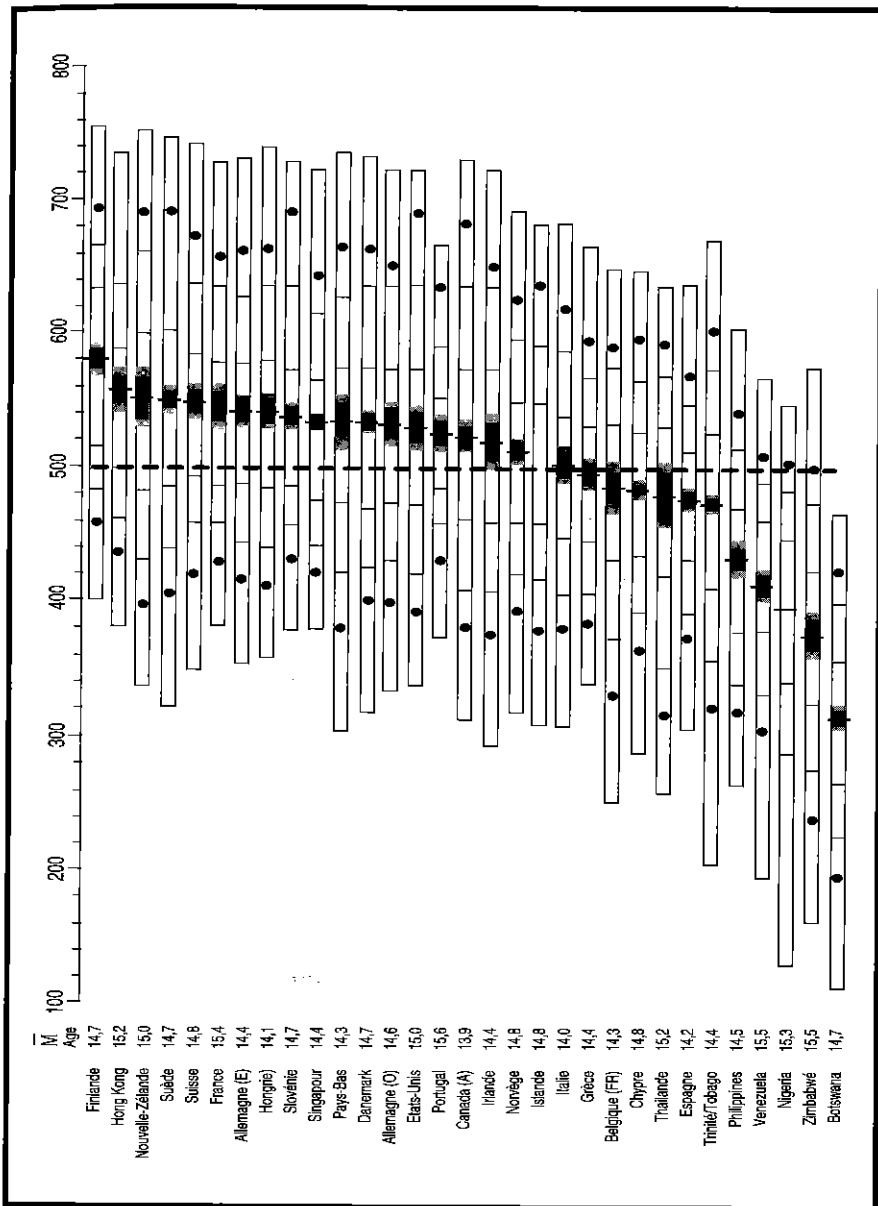


Figure 9 : Distribution des scores pour la lecture de documents informatifs (tableaux, figures, cartes, ...) - Elèves de 14 ans

Pareilles observations restent à expliquer. Où se trouvent les failles éventuelles dans les méthodes d'enseignement ? Parmi d'autres explications possibles se trouve le nombre d'heures consacrées quotidiennement à la télévision. Le tableau ci-dessous indique le pourcentage de jeunes qui la regardent plus de cinq heures par jour.

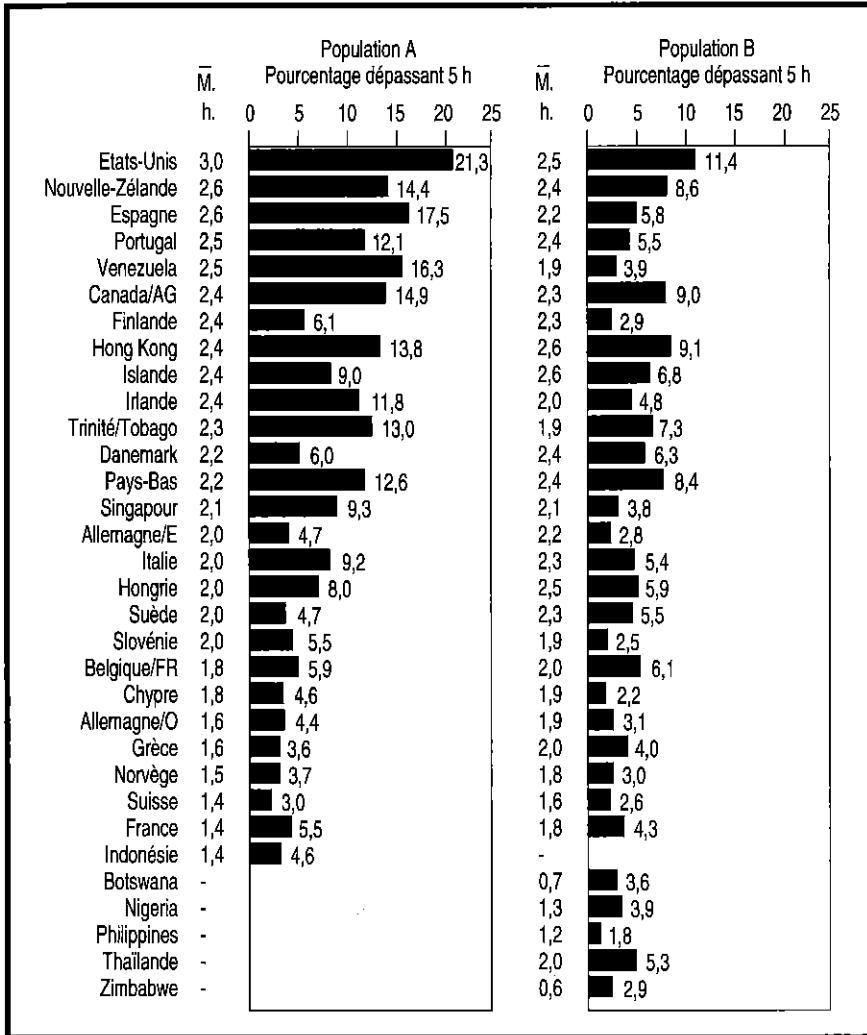


Figure 10 : Moyenne du nombre d'heures consacrées quotidiennement à la télévision et pourcentage d'élèves déclarant la regarder plus de cinq heures par jour

Les analyses statistiques révèlent une corrélation négative entre l'habileté de lecture et une fréquentation exagérée de la télévision. En revanche, il

semble que la regarder pendant moins de 3h30 par jour peut être favorable. C'est ce qui apparaît notamment en Finlande.

Dans un ouvrage prolongeant cette enquête (*Effective Schools in Reading*), T.N. Postlethwaite et K. Ross (1992) ont décrit les écoles qui enseignent la lecture de la façon la plus efficace et ont formulé un ensemble de propositions pour les planificateurs de l'éducation.

La 3e étude sur l'enseignement des mathématiques et des sciences (1991-1995)

Entre 30 et 40 pays, dont la Belgique, participeront à cette recherche, dont les instruments sont en cours de construction.

• Rôle historique et apports méthodologiques de l'IEA

L'IEA a fait oeuvre de précurseur. Fille de son temps, elle a accumulé une grande quantité de données comparatives et d'études dont les pouvoirs organisateurs et le monde de l'éducation en général n'ont pas su tirer pleinement profit, loin s'en faut. Ce phénomène s'explique, en bonne partie, par la technicité des rapports et par l'insuffisance de leur diffusion dans la plupart des pays. Par ailleurs, au moment où l'IEA a commencé ses enquêtes, peu de centres nationaux dominaient les connaissances et les techniques statistiques ou autres, nécessaires à l'exécution des travaux. A cet égard, l'IEA a accompli une action de formation et de stimulation d'une valeur inestimable.

Le réseau scientifique et social développé par l'IEA à l'échelle mondiale a beaucoup contribué au progrès de la recherche évaluative en éducation et du pilotage des systèmes éducatifs. Il est d'ailleurs symptomatique que c'est à deux anciens présidents de l'IEA, T. Husén et T.N. Postlethwaite, qu'a été confiée la direction de la rédaction des deux éditions successives de l'*International Encyclopedia of Education : Research and Studies* (Oxford, Pergamon), monumentale synthèse, en douze volumes, de la recherche scientifique en éducation à l'échelle mondiale (2e édition en 1994).

Parmi les apports plus pointus de l'IEA à la méthodologie de la recherche évaluative, on peut citer en exemples :

- La prise en compte systématique de *l'occasion d'apprendre (Opportunity to learn, OTL)* dans l'évaluation des acquis. Dans un processus éducatif, il est, en effet, contestable de faire passer des examens sur des matières que les élèves n'ont pas été mis en mesure d'apprendre (sauf si l'on est intéressé par les apprentissages dits accidentels).

A propos de chaque question posée, il est éclairant de savoir si elle porte sur une notion qui, à la connaissance des professeurs, a été en-

seignée ou rappelée pendant l'année scolaire en cours, pendant les années précédentes ou jamais.

- La notion de *productivité cognitive* (*productivity yield*). L'indice de productivité correspond au rapport entre le taux de réponses correctes et le pourcentage du groupe d'âge scolaire scolarisé. Par exemple, si un test de 30 items est administré à un échantillon représentatif d'un groupe d'âge totalement scolarisé et si tous les élèves répondent correctement, la productivité cognitive est de 100%.

T.N. Postlethwaite et D. Wiley (1991) ont analysé des scores en sciences obtenus en Thaïlande et aux Etats-Unis, lors de la "Recherche des six disciplines" (1966-1973). Le pourcentage de réponses correctes aux items a été le même dans les deux pays. Toutefois, en Thaïlande, un tiers seulement des jeunes de 13 ans étaient scolarisés, alors qu'aux Etats-Unis la scolarisation était de 100%. Le graphique ci-dessous fait bien apparaître la différence de productivité. En apparence, l'indice d'efficacité de la Thaïlande (55,99%) est pratiquement le même qu'aux Etats-Unis (55,73%), mais en Thaïlande, la productivité réelle n'est cependant que de 18,09%.

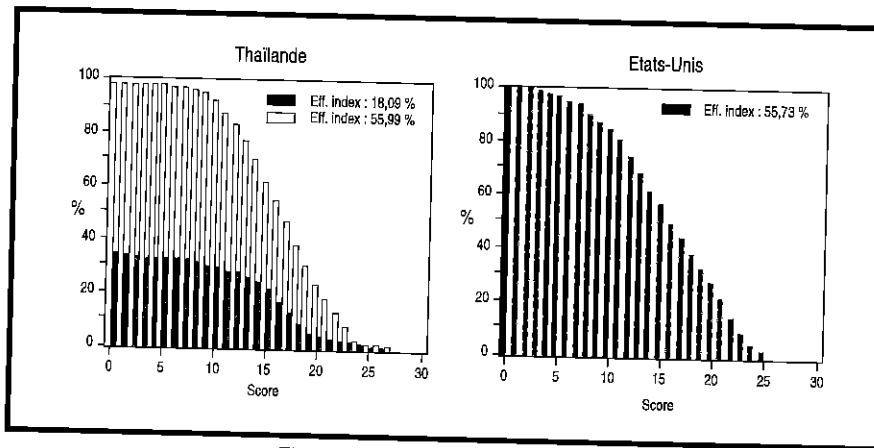


Figure 11 : Productivité cognitive

- La prise en compte de la *durée d'implication effective dans la tâche* (*time on task*). C'est pendant la recherche sur l'enseignement des langues étrangères (1966-1973) que J. B. Carroll a fait systématiquement usage de cet indice. En effet, ce n'est pas le nombre d'heures de cours prévues dans un programme pour l'enseignement d'une discipline qui permet le mieux de prédire l'apprentissage, mais bien le temps pendant lequel l'élève s'investit réellement dans le travail.

Une démarche inévitablement longue

Les recherches de l'IEA, qui viennent d'être évoquées, sont lourdes et longues.

Les causes de ce phénomène sont multiples :

- Il s'agit d'une entreprise coopérative financée (parfois avec bien des difficultés et des incertitudes) par les Etats participants, lesquels disposent tous du même pouvoir de décision lors des assemblées générales annuelles.
- La lourdeur de l'analyse comparative des curriculums, destinée à déceler les objectifs et les contenus d'enseignement communs, à propos desquels les noyaux des instruments d'évaluation seront constitués. Chaque participant est invité à proposer des items. La possibilité est toujours laissée d'ajouter des options locales aux instruments internationaux; il est ainsi possible de tenir compte des objectifs spécifiques à un pays.
- Les essais répétés et les révisions successives des instruments d'évaluation.
- Très tôt apparaissent les faiblesses scientifiques de certains participants qui n'ont ni formation approfondie en pédagogie expérimentale ou en éduométrie, ni expérience des évaluations objectives à grande échelle. Une assistance technique est apportée et les apprentissages ainsi suscités revêtent parfois plus d'importance pour les pays que les résultats de l'enquête en cours.
- La difficulté de faire respecter les mêmes règles d'échantillonnage des populations-cibles, telles qu'elles ont été définies. Malgré l'existence d'experts en échantillonnage, des pays, - même parmi les plus avancés -, commettent des erreurs qui biaisent les résultats.
- La lourdeur de l'essai général des instruments, de la collecte des données et, en général, de toutes les opérations à exécuter. Les leçons de cet essai doivent être tirées.
- Comme le travail se fait à l'échelle du monde, des décalages pouvant atteindre six mois existent entre les années scolaires. L'organisation doit en tenir compte.
- Quand les données définitives parviennent au centre international, il importe de les "nettoyer", ce qui, à l'échelle de millions de données, n'est pas une mince affaire. Des logiciels très sophistiqués ont été récemment conçus pour faciliter cette opération.
- Viennent alors les traitements statistiques qui permettent de dégager les dominantes sûres, de mettre des variables en relation. L'identification des variables à considérer est aussi en soi un acte important et souvent complexe.

- Les résultats sont ensuite envoyés à chaque pays participant pour en faire une interprétation aussi éclairée que possible.
- Après bien des va-et-vient, les rapports sont enfin rédigés. Ils sont soumis aux participants avant publication.
- Les publications revêtent plusieurs formes, allant de gros volumes destinés au monde scientifique et aux spécialistes attachés à l'administration supérieure des systèmes éducatifs, à des présentations simplifiées conçues pour la grande diffusion.

Ainsi s'explique qu'une recherche de l'IEA prend toujours plusieurs années.

□ **International Assessment of Educational Progress (IAEP)**

L'*Educational Testing Service* (ETS), qui avait déjà charge de la *National Assessment of Educational Progress* (NAEP) des Etats-Unis, a pris l'initiative d'enquêtes internationales rapides. Bien que, méthodologiquement, elles ne soient pas comparables à celles que l'IEA réalisent, certains y ont vu une concurrence, une lutte d'influence. C'est, en effet, au moment où l'OCDE a décidé de lancer son projet relatif aux indicateurs de l'enseignement (INES) et allait, pour cela, avoir besoin de données internationales sur les rendements scolaires qu'un marché potentiel a semblé apparaître. Cet espoir sera déçu : le projet INES prendra bien son essor, mais l'OCDE laissera, au moins provisoirement, à chaque pays, le soin de collecter les données nécessaires.

• **L'Enquête IAEP I (1987-1988)**

L'IAEP poursuit explicitement deux objectifs :

- Réduire la durée et le coût d'études internationales comparatives, en recourant aux instruments et aux méthodes mis au point pour la NAEP des Etats-Unis.
- Permettre aux pays intéressés de se familiariser avec les techniques utilisées par la NAEP, afin de voir si elles conviennent pour leurs propres projets d'évaluations locales.

Des propositions en ce sens ont été faites en 1987 et une première enquête portant sur le rendement des élèves de 13 ans, en mathématiques et en sciences, a été décidée et exécutée en 1988. Cinq pays (Corée, Irlande, Canada, Espagne, Etats-Unis, Royaume-Uni) y ont participé. Les frais de coordination et de traitement des données de cette enquête ont été pris en charge par le *Department of Education* et la *National Science Foundation* des Etats-Unis.

Les questions d'évaluation ont été empruntées aux pools de 281 items de mathématiques et de 188 items de sciences, utilisés dans l'enquête de la NAEP en 1986. Des experts provenant des pays participants ont estimé dans quelle mesure ces questions correspondaient à leurs curriculums nationaux et étaient traduisibles ou devaient être culturellement adaptées (changement des unités de mesure, des prénoms des enfants, des plantes et des animaux,...).

Les techniques éduométriques utilisées pour le traitement des données sont d'avant-garde et ont permis de gagner beaucoup de temps.

En janvier 1989, ETS a publié les résultats de ce premier survey (A.E. Lapointe *et al.*, 1989). Parmi les données comparatives, on remarque notamment une intéressante tentative de mise en évidence d'une relation entre la durée des devoirs à domicile en mathématiques et les rendements observés dans cette branche.

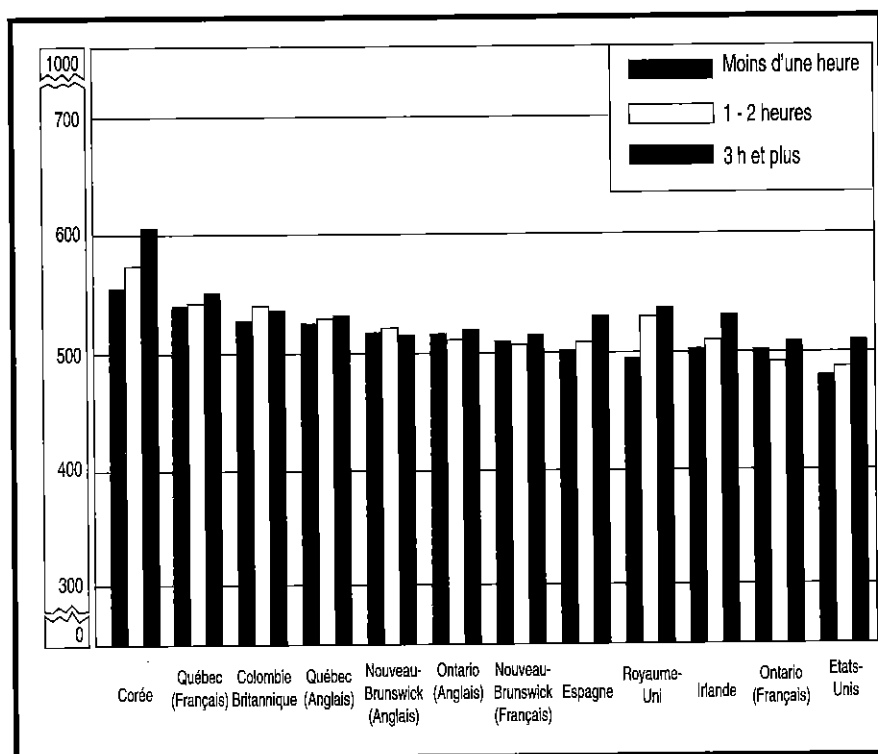


Figure 12 : Compétence moyenne en mathématiques en relation avec la durée hebdomadaire des devoirs à domicile dans cette branche

- **L'enquête IAEP II (1990-1991)**

Cette enquête, réalisée dans le même esprit que la première, concerne les populations de 9 et de 13 ans et porte sur les mathématiques, les sciences et la géographie. Vingt-huit systèmes éducatifs (nationaux ou provinciaux) y ont participé.

En conclusion de cette seconde étude, le Ministère de l'Éducation des États-Unis espère certains progrès dans les procédures d'échantillonnage et d'exécution de l'étude, des informations plus précises sur la comparabilité des échantillons prélevés dans les différents systèmes éducatifs, et la clarification de questions qui subsistent à propos de l'interprétation de données.

- **Une critique fondamentale**

Une critique fondamentale continue à être adressée à l'IAEP : même si elle procède à diverses consultations, elle n'en exporte pas moins des questions conçues en fonction des programmes des États-Unis. Pour sa deuxième opération, en 1991, ETS a partiellement modifié cet état de choses en invitant les pays participants à proposer des items de tests reflétant directement leurs curriculums respectifs. L'approche reste néanmoins du type "centre-périphérie". Elle contraste fondamentalement avec la démarche "périphérie-centre" de l'IEA et avec l'intensité des relations humaines et des actions de formation multidirectionnelles qui l'accompagnent.

Contrairement à l'IEA, l'IAEP ne procède pas à une analyse détaillée des facteurs associés aux différences observées entre les pays. Jusqu'à présent, elle n'a pas collecté largement des informations contextuelles sur les systèmes éducatifs évalués, ce qui rend difficile l'explication des différences de rendement observées.

L'IAEP a travaillé jusqu'à présent à court terme, tandis que l'IEA vise le long terme et estime que les effets secondaires de sa démarche sont au moins aussi importants, sinon plus, que les apports immédiats.

□ **Le projet de l'OCDE : International Indicators and Evaluation of Educational Systems (INES) - Indicateurs internationaux de l'enseignement.**

• **Une première expérience : les indicateurs sociaux**

Dès les années 60, l'Organisation pour la Coopération et le Développement Economiques (OCDE) a voulu produire des statistiques susceptibles d'aider à la planification de l'éducation. L'attention était surtout focalisée sur les flux, mais le souci de dresser des espèces de bilans des systèmes éducatifs, à l'instar de ce qui se faisait pour l'économie, n'était pas absent.

En 1970, l'OCDE a envisagé de produire un ensemble - qui se voulait exhaustif - d'indicateurs relatifs au rendement de l'enseignement par rapport aux objectifs proposés. En 1973, un rapport conceptuel de base, rédigé par R. Carr-Hill et O. Magnussen, fait un tour d'horizon des problèmes en cause et envisage les types d'indicateurs qui pourraient être collectés. L'intérêt historique de cette publication est grand; la qualité de ceux qui y ont collaboré permet d'assurer qu'elle reflète correctement la réflexion la plus avancée de l'époque. On trouve, en effet, parmi les contributeurs : K. Harnqvist (qui jouera aussi un rôle important dans la conception du projet INES), C. Arnold Anderson, M. Blaug et J. Coleman.

L'idée de départ est qu'il importe de compléter les indices économiques généraux tels que le PNB (*input*), qui ne donnent pas une idée exacte "du bien-être des peuples", par des indicateurs sociaux relatifs à l'*output*. L'un des buts prioritairement assignés est de suggérer "un cadre général pour un système d'information statistique permettant de contrôler la politique de l'enseignement" (p.27).

Un indicateur social est défini de la façon suivante : "Statistique d'intérêt normatif direct qui facilite la formulation de jugements concis, exhaustifs et équilibrés sur l'état des principaux secteurs de la société. Il s'agit, dans tous les cas, d'une mesure du bien-être social, pouvant susciter l'opinion que, si elle évolue dans la <bonne> direction, toutes autres choses demeurant égales par ailleurs, la situation s'améliorera, ou bien que les gens bénéficieront d'une situation meilleure. C'est pourquoi le nombre de médecins ou de policiers ne peut être un indicateur social, alors que des statistiques sur la santé ou sur le taux de criminalité peuvent l'être."

Là où semblait exister un rapport entre une mesure statistique et "ce vague concept de bien-être social", cette mesure a été appelée "indicateur" en ce sens qu'il mesure l'*output* ou le rendement (p.12). De tels indicateurs "devraient être élaborés autant que possible au sein d'un système général de comptabilité sociale" (p.12).

Cinq domaines d'objectifs relatifs aux rapports entre l'enseignement et la société sont distingués. Une présentation quelque peu détaillée de ces domaines se justifie pour deux raisons. D'abord par l'intérêt intrinsèque qu'ils présentent et parce qu'ils témoignent bien de l'esprit du temps. Ensuite, parce que cette présentation aide à mieux prendre conscience de la distance qui sépare l'actuel Projet INES de l'OCDE, des suggestions de 1973, difficiles à rendre opérationnelles.

1. *La transmission des connaissances théoriques et pratiques*

Ces connaissances doivent permettre aux individus d'agir plus ou moins indépendamment au sein des principales institutions de la société.

On distingue :

- *L'alphabétisme fonctionnel.* Les connaissances comme telles ne suffisent pas. Pour vivre de façon satisfaisante dans une société complexe, l'individu doit apprendre à comprendre ce qu'il peut et doit faire dans une situation donnée. Les auteurs du rapport soulignent immédiatement la difficulté de collecter, sur ce sujet, un ou plusieurs indicateurs permettant notamment la comparaison entre pays.
- *La participation à la vie politique.* Les auteurs reconnaissent qu'il est pratiquement impossible de déterminer la part qui revient à l'enseignement dans l'aptitude à exercer cette activité.
- *L'étendue du pouvoir réel et apparent des élèves.* Il s'agit notamment de la place qui leur est donnée dans les prises de décision.
- *L'efficacité des consommateurs.* Aptitude à bien choisir ce que l'on va acheter. Le nombre d'heures passées à suivre des cours d'économie domestique et l'affiliation à des sociétés de défense des consommateurs figurent parmi les indicateurs envisagés.
- *Les rapports sociaux.* Réserver la fréquentation scolaire aux jeunes seulement pourrait les priver de l'expérience du travail avec des adultes. D'où la proposition d'admettre des adultes parmi les élèves de l'enseignement secondaire. L'égalité entre filles et garçons est un autre aspect retenu.
- *Les citoyens de demain.* Education sexuelle, initiation au contrôle des naissances, écologie des sociétés humaines.

2. *L'enseignement et l'économie*

- *Contribution à la croissance économique.* Connaissance des fonctions de production et des comptes nationaux. L'indicateur envisagé ici est la part de la différence entre le revenu par habitant d'un pays p et celui d'un pays de référence, qui peut s'expliquer par les différences entre leur capital humain respectif.
- Qualité de la main-d'oeuvre instruite.

- Différence de gains en fonction du niveau d'instruction et de la profession.
- Chômage en fonction du niveau d'instruction.
- Volume de la formation sur le tas.
- Différences entre le niveau d'instruction des nouveaux diplômés pénétrant sur le marché du travail et le niveau moyen d'instruction de la population active.

3. L'égalité des chances

- *Taux de scolarité* à tous les niveaux et dans tous les types d'enseignement, en fonction du sexe, de la race, du Q.I., de l'âge, de la catégorie sociale d'origine.
- *Niveau d'instruction des parents.*
- Proportion des *ressources d'enseignement* consacrée spécialement aux groupes qui sont considérés comme défavorisés.
- *Situation et revenus* en fonction des différents niveaux de réussite scolaire, par âge, race et catégorie sociale des parents.

4. Le système d'enseignement au service des individus

Satisfaction de la demande privée (ou demande sociale) et rendement de l'enseignement pour l'individu : santé (contribution apportée par l'éducation), loisirs (proportion de personnes qui continuent d'exercer une activité de loisir apprise à l'école), offre d'éducation, créativité (importance accordée aux travaux personnels lors de l'évaluation des performances scolaires), discipline (actes de vandalisme, absentéisme), aptitude à l'éducation (pourcentage d'adultes qui suivent volontairement des cours qui ne relèvent pas de la formation professionnelle).

La diversité des objectifs poursuivis, les grandes différences dans l'importance qui leur était attribuée, la difficulté de traduire les propositions en spécifications opérationnelles et la priorité accordée à l'*output*, - en laissant les processus dans l'ombre -, expliquent en grande partie l'échec de cette première tentative de l'OCDE.

• Le projet INES

En 1988, le *Centre pour la Recherche et l'Innovation dans l'Enseignement (Centre for Educational Research and Innovation, CERI)* de l'OCDE a lancé un nouveau projet coopératif de mise au point et de prélèvement périodique d'indicateurs internationaux de l'enseignement. En particulier, ils devraient informer sur l'efficacité de l'éducation et aider à définir et à adapter les politiques éducatives dans les pays membres. L'idéal est évidemment d'établir des liens entre l'ensemble des indicateurs internatio-

naux et les systèmes d'indicateurs nationaux, en prélevant des indicateurs similaires dans les deux ensembles. Le cadre théorique de ce projet, surtout en ce qui concerne l'évaluation, est plus solide et plus réaliste que celui qui fut utilisé pour la première tentative.

Que l'OCDE se soit engagée dans le projet INES est significatif de la liaison de plus en plus étroite qui existe entre éducation et économie. La publication, aux Etats-Unis, du célèbre rapport *A Nation at Risk* (La nation est en péril) a dramatisé cette liaison et a ainsi fortement impressionné le monde politique et l'opinion publique. On a vu que les données fournies par l'IEA et par la NAEP sur les connaissances des élèves américains en mathématiques et sur leur habileté en lecture étaient fort alarmantes. Le rapport de McKnight *et al.* (1987) concluait qu'une bonne formation en mathématiques "était essentielle à la santé et à la compétitivité du pays sur le marché technologique mondial".

C'est dans ce contexte que le gouvernement des Etats-Unis a fortement soutenu l'idée d'une étude internationale sur l'état de l'éducation. Des indicateurs relatifs à trois domaines devraient être collectés : les résultats, les ressources et les contextes.

En novembre 1987, le Département de l'Education des Etats-Unis a organisé, à Washington, en collaboration avec l'OCDE, une conférence internationale chargée d'examiner cette proposition. La conférence réunit des représentants de vingt-sept pays membres de l'OCDE, ainsi que des experts de la Banque mondiale, de l'IEA et de divers organismes américains.

A cette occasion, l'opposition entre partisans de l'évaluation quantitative et partisans de l'évaluation qualitative refit surface. "Néanmoins, le sentiment général majoritaire fut que, quelles que soient les limites des indicateurs de performances scolaires actuels, ils apportent néanmoins des informations permettant aux responsables de l'éducation de répondre aux demandes de données internationales qui aident à situer comparativement les données nationales relatives aux rendements des systèmes scolaires. Les délégués ont aussi évoqué les demandes fréquentes et insistantes formulées par les chefs d'Etat, les Ministres, les Secrétaires d'Etat, à propos des budgets, des résultats scolaires, de l'efficacité des systèmes éducatifs, tant dans leur pays qu'à l'étranger. Ces demandes ont été formulées avec une insistance particulière par des pays soucieux de leur compétitivité économique, comme les Etats-Unis, ou procédant à d'importantes modifications structurelles de leur système éducatif, comme la France et le Royaume-Uni" (A. Ruby, 1988, p.26).

En mars 1988, la Conférence de Poitiers compléta cette réflexion générale. L'urgence de lancer le projet relatif aux indicateurs fut reconnue très majoritairement. En conséquence, le CERI réunit immédiatement un groupe d'experts chargé d'opérationnaliser le projet et, dès le mois de mai, les propo-

sitions de ce groupe furent acceptées, avec quelques amendements, par le Comité directeur du CERI. Le projet pouvait prendre son essor. Une première phase exploratoire de dix-huit mois fut consacrée à la définition et à l'essai de la collecte d'indicateurs. Cette phase fut couronnée de succès et la réalisation proprement dite a pu être engagée.

Quatre réseaux internationaux ont été constitués; ils jouent un rôle crucial dans ce projet :

- Réseau A : Acquis des élèves.
- Réseau B : Education et insertion sur le marché du travail.
- Réseau C : Caractéristiques des établissements et des systèmes d'enseignement.
- Réseau D : Attitudes, représentations et attentes.

Ces réseaux sont très actifs. Ils aident notamment à faire évoluer le système d'indicateurs internationaux de façon à serrer au plus près la réalité et l'évolution des systèmes éducatifs.

Le sérieux de la réflexion qui accompagna ce départ est attesté par un nombre impressionnant d'études théoriques et de rapports dont le CERI a suscité la rédaction. Un moment décisif a été marqué par la publication, en 1992, des *Regards sur l'éducation. Les indicateurs de l'OCDE*, présentation d'un premier ensemble d'indicateurs internationaux de l'enseignement. Le retentissement de cette publication dans la presse et dans les organismes officiels témoignent de l'intérêt suscité.

Trente-six indicateurs ont été retenus provisoirement et, parmi eux, une dizaine sont de nature nettement expérimentale (ils figurent en italiques, dans la liste qui va suivre). L'ambition annoncée est de disposer d'un *tableau de bord*, de *voyants* concernant les aspects les plus significatifs et les plus sensibles de l'institution éducative et fournissant, par comparaison avec d'autres systèmes d'enseignement, une vue synthétique des points forts et des points faibles de chacun des pays.

Les indicateurs qui ont fait l'objet de la publication en 1992 sont les suivants :

A. Le contexte démographique, économique et social

- Niveau général de formation de la population adulte (25 à 64 ans).
- Pourcentages de femmes dans la population de 25 à 64 ans, par niveau de formation.
- Taux d'activité et de chômage.
- Pourcentage de la population de 25 à 29 ans, par rapport à la population totale.
- Produit intérieur brut (PIB) par habitant.

B. Les coûts, ressources et processus scolaires

- Dépenses publiques et privées afférentes à l'éducation, exprimées en pourcentage du PIB.
- Dépenses publiques d'éducation par rapport au total des dépenses publiques.
- *Sources des dépenses publiques d'éducation.*
- Dépenses de fonctionnement.
- Dépenses par élève.
- Investissement national en faveur de l'enseignement.
- Dépenses relatives par élève et par niveau scolaire.
- *Pourcentage de la population active employée dans l'enseignement public.*
- Taux de scolarisation dans le système d'enseignement.
- Taux de préscolarisation.
- Taux de scolarisation dans l'enseignement secondaire supérieur.
- Taux d'accès à des études conduisant à un diplôme d'enseignement supérieur.
- Taux de scolarisation dans l'enseignement supérieur non universitaire.
- Ratio élèves / enseignant.

C. Résultats de l'enseignement

- Taux de réussite à la fin de l'enseignement secondaire du second cycle.
- Taux d'obtention d'un diplôme du type universitaire.
- *Taux de survie dans l'enseignement supérieur du type universitaire.*
- Diplômes par domaine scientifique.
- *Pourcentage de la population jeune obtenant un diplôme scientifique.*
- Chômage et niveau de formation.
- Salaires et niveau de formation.
- *Résultats en mathématiques.*
- Comparaison entre écoles des résultats en mathématiques.
- Comparaison entre élèves des résultats en mathématiques.

Pour ce réseau, certains aspects devront faire l'objet d'une attention particulière en raison de leur pouvoir explicatif, notamment : le nombre d'heures d'enseignement consacrées à chaque discipline, le degré d'individualisation de l'enseignement, le nombre de jours de classe par année, le curriculum effectivement implanté, le taux d'investissement dans les tâches (*time on task*), les travaux à domicile (nature et durée moyenne).

Par ailleurs, le réseau C souhaite qu'à l'avenir, le travail se concentre plus sur les établissements scolaires que sur les systèmes globalement considérés. Dans une note de 1992, J. Scheerens, responsable du réseau, relève que, pour identifier les indicateurs de processus, on s'est surtout référé à des modèles empiriques de l'efficience d'un établissement scolaire. "Cela signifie, en gros, que l'on a sélectionné, parmi les indicateurs, ceux qui avaient montré une corrélation positive avec les résultats de l'éducation." Scheerens envisage trois autres voies possibles pour la sélection de nouveaux indicateurs :

1. Dresser, à partir de la littérature existante, un inventaire des modèles types de "l'école idéale". On pourrait ainsi identifier des aspects que la recherche empirique n'a pas fait ressortir, notamment certaines valeurs telles que la satisfaction que les enseignants retirent de leur travail, et aussi l'aptitude des établissements à gérer les pressions extérieures, à maintenir un bon niveau de recrutement, etc.
2. Analyser les principaux programmes de la politique de l'éducation relative à un aspect précis de la problématique éducative (par exemple, les programmes spéciaux en faveur des groupes minoritaires) ou à une catégorie d'écoles déterminée.
3. Examiner le fonctionnement des établissements en adoptant le point de vue des consommateurs d'éducation : parents, élèves, ... On pourrait ainsi identifier les "caractéristiques de processus" souhaitées pour certaines catégories d'écoles.

Voici, à titre d'exemples, deux extraits des *Regards sur l'éducation* (1992). Le premier concerne le rendement de l'enseignement des mathématiques d'élèves de 13 ans, en 1982. Ces données sont empruntées à l'IEA. La présentation particulièrement parlante facilite les comparaisons.

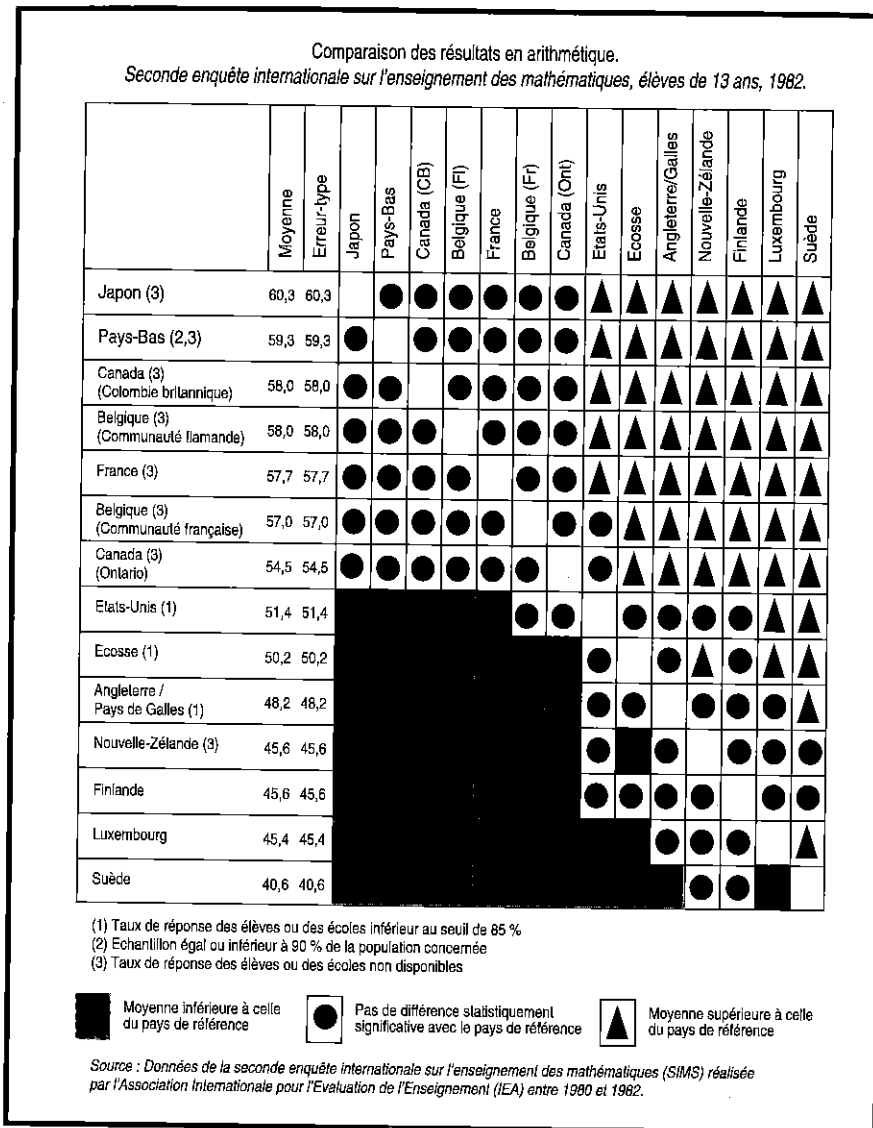


Figure 13 : Résultats en mathématiques

Le second exemple concerne le degré d'autonomie des établissements

Tableau 12 : Décisions prises au niveau de l'établissement en pourcentage du total des décisions prises dans les systèmes publics d'enseignement, par niveau scolaire (1991)

	Niveau scolaire		
	Primaire	Secondaire 1er cycle	Secondaire 2e cycle
Espagne	56	56	40
Pays-Bas	52	46	59
Portugal	40	39	44
Suède	37	37	37
Finlande	35	45	59
Belgique	35	35	35
Norvège	30	30	23
Italie	29	29	29
Etats-Unis	22	22	22
France	17	34	32
Suisse	10	10	23

**DÉCISIONS PRISES PAR LES ÉTABLISSEMENTS
PRIMAIRE ET SECONDAIRES**

Principaux résultats

Dans la plupart des pays, la part des décisions prises par l'école est la même dans le primaire que dans le secondaire.

La Finlande et la France se démarquent des autres pays : les écoles primaires ont moins de "pouvoir".

Définition

Seules les décisions prises par les établissements sont retenues. Les comparaisons portent sur les différences d'un niveau d'enseignement à l'autre du système public.

Notes explicatives

Les pays sont présentés par ordre décroissant selon la part des décisions prises par l'école primaire (CITE 1).

Dans la plupart des pays, une même législation est appliquée à l'ensemble de la scolarité obligatoire (niveaux CITE 1 et 2).

Pour quatre pays (Belgique, Etats-Unis, Italie et Suède), la situation est la même pour les trois niveaux d'enseignement.

En Finlande, on peut observer une progression dans le pouvoir de décision de l'établissement : il augmente avec les niveaux : du primaire au secondaire inférieur, puis au secondaire supérieur. On aurait pu croire que ce cas de figure se présenterait plus fréquemment.

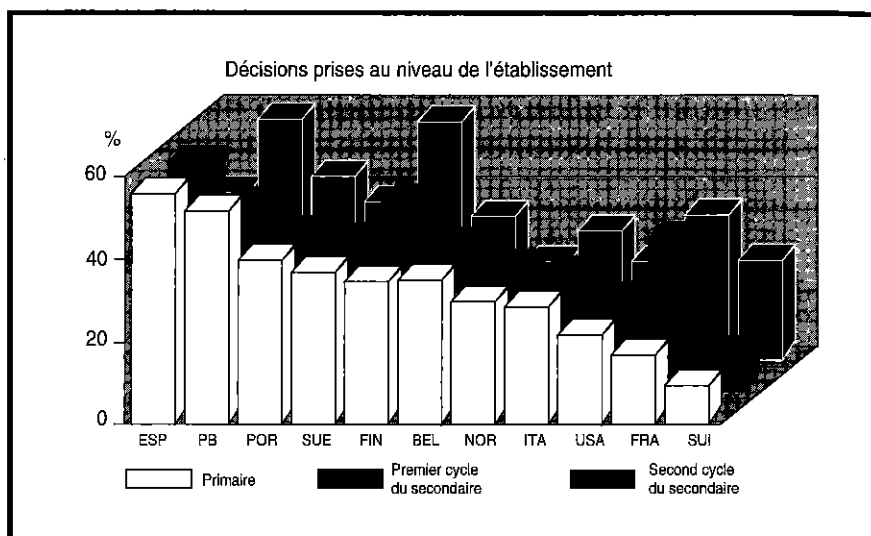


Figure 14 : *Décisions prises par les établissements primaires et secondaires*

A l'avenir, le nombre d'indicateurs sera vraisemblablement augmenté et la nature de certains d'entre eux sera modifiée. Il semble, en particulier, que plus d'insistance sur les rendements scolaires sera souhaitée; c'est, en effet, ce que les élèves apprennent, et leur développement personnel et social qui retiennent le plus l'attention du monde de l'éducation.

Pour des pays comme la Belgique, où il n'existe pas d'examens nationaux centralement organisés, il n'est pas aisé d'obtenir des indicateurs de rendement valides. Seules les grandes enquêtes normatives de l'IEA ou de l'IAEP apportent quelques points de repère, mais ils ne permettent pas de répondre pleinement aux demandes annuelles de l'OCDE. La question est d'autant plus épineuse que, dans les structures fédérales, les curriculums d'enseignement tendent à se diversifier.

La solution devra s'inspirer d'un pays comme l'Australie qui regroupe ses divers Etats au sein d'un Commonwealth, et où des statistiques nationales continuent à être publiées. Il en va de même aux Etats-Unis et au Canada. Quant à l'Angleterre, où l'indépendance des Autorités Locales était restée très grande jusqu'à ces derniers temps, elle vient, comme on l'a vu, d'adopter un curriculum national et des mécanismes d'évaluation correspondants.

Problèmes particuliers relatifs aux attitudes, aux représentations, aux attentes

Des indicateurs relatifs aux attitudes et aux attentes (Réseau D) ne figurent pas dans la première édition des *Regards sur l'éducation*. Cette absence s'explique à la fois par la complexité de l'évaluation de ces caractéristiques et par l'insuffisance des données existant dans les pays. Or la question est de grande importance et elle a retenu l'attention de l'OCDE, dès son premier projet relatif aux indicateurs de l'enseignement (voir *supra*) : bien des évaluations souhaitées portaient déjà sur les attentes (Carr-Hill et Magnussen, 1973). C'est pourquoi, en suivant notamment J. De Groof *et al.* (1992), nous accordons un développement particulier à ce sujet.

Les indicateurs relatifs aux attitudes et aux attentes concernent la façon dont les acteurs directs du système d'enseignement le perçoivent, le vivent, espèrent en bénéficiant, et se situent par rapport à lui. En particulier, les attitudes influencent directement les performances scolaires et, avec les attentes, elles déterminent dans une large mesure le contexte dans lequel le processus éducatif se déroule. Le phénomène est culturellement très chargé, ce qui explique le caractère délicat du choix d'indicateurs internationaux en la matière.

Les travaux du Réseau D sont restés, jusqu'à présent, de nature surtout exploratoire. Il s'est assigné deux objectifs principaux :

- Définir des attitudes et des attentes des communautés éducatives en matière scolaire, internationalement comparables. Cinq attitudes et cinq attentes exerçant une influence durable sur des aspects permanents du système d'enseignement doivent être retenues. Les données seront ventilées selon l'âge, le sexe et la classe sociale.
- Définir cinq attitudes et cinq attentes des enseignants et des élèves, relatives au bon fonctionnement de l'école.

Actuellement, on ne dispose guère de résultats d'enquêtes fiables et régulières en ces matières, dans la plupart des pays participants. Aussi, afin de disposer d'un cadre de référence susceptible d'aider au choix d'indicateurs, le Réseau D a dressé, à partir des recherches publiées, un inventaire des préoccupations communes et les a réparties en sept groupes :

1. Les enseignants
 - Qualités
 - Statut et moral
 - Niveau de professionnalisme
 - Formation
 - Attitudes vis-à-vis des innovations et des réformes
2. Gestion de l'école
 - Objectifs poursuivis par l'établissement

- Climat, moral, culture de l'établissement
 - Mode de leadership
 - Relations avec les parents et les employeurs.
3. Curriculum
- Qualité du curriculum
 - Curriculum prévu
 - Curriculum effectivement mis en oeuvre (curriculum implanté)
 - Performances des élèves
 - Relation entre le curriculum, la vie et le monde du travail
4. Relations maîtres - élèves
- Attentes des enseignants vis-à-vis des élèves
 - Attentes des élèves vis-à-vis des enseignants
 - Moral des enseignants
 - Moral des élèves
 - Evaluation et auto-évaluation des enseignants
 - Evaluation et auto-évaluation des élèves
5. Communication
- Circulation de l'information relative aux objectifs et aux priorités
 - Communication avec d'autres groupes professionnels
 - Modes de consultation des parents
 - Communication des résultats scolaires aux parents
6. Equité
- Discrimination raciale, sociale, sexuelle
 - Egalité des chances
 - Impartialité des enseignants
 - Mesures prises en faveur des handicapés et d'autres élèves éprouvant des besoins spéciaux
 - Respect de la langue maternelle
7. Politique de l'enseignement
- Buts poursuivis dans le système éducatif
 - Priorités budgétaires
 - Ecoles publiques - Ecoles privées
 - Liberté du choix scolaire

L'avenir

Il semble dès maintenant acquis que les indicateurs de l'éducation de l'OCDE vont prendre place à côté des indicateurs de l'économie. J. Magy (1992, p.10) estime avec raison que ces indicateurs "sont appelés à être la

toile de fond permanente des analyses fondamentales des politiques nationales d'éducation."

Il importe, par ailleurs, de souligner l'effet stimulant de ce projet sur le développement des services de statistiques et des systèmes de pilotage dans les pays qui ne s'en étaient pas encore suffisamment dotés. Cet effet se fait aussi sentir au niveau des grandes organisations internationales (OCDE, UNESCO, UE) qui s'efforcent de coordonner leurs statistiques. La *Classification internationale type de l'enseignement (International Standard Classification of Education - ISCED)*, initialement publiée par l'Unesco, va être révisée, après une large consultation des utilisateurs, pour éviter plusieurs des ambiguïtés actuelles.

La Conclusion 3 de l'Assemblée générale des pays participants au Projet INES, qui s'est tenue à Lugano en 1992, est la suivante : "Comme les indicateurs (internationaux) ne sont pas très élaborés, ils devront être complétés et pondérés par des données supplémentaires. Pour analyser les informations qu'ils fournissent et en tirer des conclusions, il est notamment essentiel de tenir compte des contextes nationaux, c'est-à-dire, notamment, des structures gouvernementales et des contextes dans lesquels s'élaborent les décisions en matière d'enseignement (importance de la décentralisation, de l'autonomie des établissements scolaires pour les décisions clés, etc.), des différences de conditions économiques et de ressources allouées à l'enseignement (illustrées par la part de l'enseignement dans le PNB), des différences de structures de la société (degré d'homogénéité, pourcentage de groupes minoritaires...), des différences dans les principales statistiques socio-économiques (démographie, tendances de la population,...), ainsi que de certaines composantes relatives, entre autres, à la culture, aux valeurs, à l'histoire, qui contribuent à façonner le cadre particulier dans lequel s'effectuent les prises de décisions et la mise en oeuvre des politiques en matière d'enseignement. (...) Il s'ensuit que, dans la mesure du possible, d'autres sources devraient être utilisées pour confirmer, corriger ou mieux comprendre les informations fournies par les indicateurs." C'est parler d'or. Cette nécessité d'éclairer la signification des indicateurs est répétée sous un autre angle dans la Conclusion 17 : "Il est certain que les indicateurs de l'enseignement peuvent largement bénéficier d'une confrontation avec les indicateurs des secteurs des sciences, de la technologie et de l'économie."

L'importance du projet INES, qui prend actuellement son véritable essor, est considérable, non seulement par son intérêt intrinsèque, mais aussi par l'effet d'entraînement qu'il provoque parmi les pays membres. Ainsi, désireuse de faire connaître les données qui lui sont propres (et non plus fondées avec celles des autres Communautés belges), la Communauté Flamande de Belgique a publié, en octobre 1993, ses propres *Regards sur l'éducation*, sous le titre évocateur *L'école fait rapport. L'enseignement flamand*

ant des

ion de
Magy
être la

dans un contexte international. La Communauté Française de Belgique va faire de même.

Toutefois, l'entreprise INES est pleine d'embûches. On peut notamment s'attendre à certaines difficultés relatives à la collecte des données concernant les performances scolaires et les aspects contextuels. Il a semblé, un moment, que l'IEA ou l'IAEP, ou ces deux organismes ensemble, pourraient se charger de ce travail. Ensuite, l'OCDE a semblé vouloir laisser, à chaque pays, le soin de collecter les données qui le concernent. Au moment où la rédaction de ce livre est terminée, des hésitations subsistent.

isations

que va

notam-
es con-
blé, un
rraient
chaque
t où la

Micropilo

Généralités

Les considérations qui précèdent concernent le pilotage à grande échelle : niveaux national et régional. Toutefois, les évaluations à petite échelle sont aussi d'un grand intérêt. Ainsi, une commune, un établissement qui ont formulé un projet éducatif, et donc des objectifs particuliers, peuvent aussi chercher à vérifier objectivement dans quelle mesure les intentions se réalisent. L'observation des rendements différentiels, expliqués par leur contexte local, a souvent des implications pédagogiques et politiques considérables.

On distingue l'*audit d'établissement*, souvent réalisé par des inspecteurs, l'*audit participatif d'établissement*, où ceux qui y travaillent coopèrent à l'évaluation avec les observateurs externes, et l'*auto-évaluation*.

Une autorégulation féconde n'est possible que si deux problèmes sont résolus : celui de la latitude que le pouvoir organisateur doit laisser à l'établissement et celui de la qualité des relations que l'établissement instaure avec son environnement.

A propos du premier problème, la conclusion que T. Wyatt et A. Ruby (1990) tirent de leur riche expérience australienne est sans ambiguïté : "Pour assumer ses responsabilités vis-à-vis du système scolaire global, le pouvoir organisateur doit être prêt à user, si nécessaire, de son autorité pour faire exécuter les décisions qui ont été prises à propos des buts principaux à poursuivre et pour assurer le bon fonctionnement du système. Mais, en même temps, il est essentiel que les écoles jouissent d'une liberté suffisante pour agir en fonction des besoins particuliers de leurs élèves. L'exercice de cette liberté d'action, devant respecter les options essentielles prises par le pouvoir, n'est pas toujours aisé. Et il n'est pas possible si l'on a affaire à des autorités autoritaires, hypercentralisatrices. Il faut donc que se développe un système complexe de relations basées sur une compréhension commune des buts et des responsabilités, ce qui n'est possible que s'il existe un mécanisme efficace de consultation et de communication".

Au sujet du deuxième problème, un établissement ne peut le mieux accomplir sa mission qu'en favorisant l'interaction avec toutes les forces vives, actives dans son environnement proche. C'est évidemment plus une question d'esprit, d'information mutuelle sur l'essentiel qu'il s'agit, que de la conduite technique de l'enseignement et des évaluations qui, d'évidence, restent de la compétence directe des maîtres.

Pour l'évaluation des performances, les services nationaux ou communautaires peuvent mettre les instruments de macropilotage, construits par les spécialistes, à la disposition de ceux qui veulent s'en servir localement, et éventuellement les aider à traiter les données. Cela se fait de plus en plus. Cependant, ces instruments ne permettront vraisemblablement pas de répondre à toutes les questions qu'une école particulière se pose à propos des facettes principales de son projet, de son action et des interactions avec son environnement. La construction d'instruments de mesure doit être l'un des objets de la formation (initiale et continuée) de tous les enseignants.

N. Bottani et H. Walberg (1992, p.73) ont dressé un inventaire des conclusions des principaux auteurs (Scheerens, 1989; Windham, 1988; Beneviste, 1987; Teauber, 1987; Oakes, 1987) qui ont proposé des listes de variables dont les recherches expérimentales sur l'efficacité scolaire ont démontré l'importance.

En voici la synthèse :

- Les différents modes de stimulation de la performance.
- Une politique éducative stimulant les performances scolaires.
- Le niveau élevé d'expectation (pourcentage d'élèves dont les professeurs espèrent la réussite; espoir qu'ont les élèves de poursuivre leurs études).
- La qualité du leadership pédagogique (temps que le chef d'école consacre à l'animation pédagogique, temps consacré aux problèmes d'éducation lors des réunions de professeurs,...).
- La clarté des objectifs prioritaires que l'établissement s'assigne.
- La coopération des professeurs à la planification de l'enseignement et à d'autres propos.
- La qualité du curriculum.
- La qualité de l'évaluation.
- La qualité des enseignants (formation disciplinaire et pédagogique initiale et continuée, flexibilité, ancienneté, etc.).
- La qualité structurelle de l'enseignement (définition claire des objectifs, répartition des apprentissages en unités cohérentes, vérification régulière des acquis, feedback fréquent, individualisation occasionnelle, ...).
- L'occasion d'apprendre.
- Le pilotage du progrès des apprentissages (utilisation de tests de connaissances; dossiers informatisés des résultats scolaires individuels, etc.).
- L'organisation pédagogique (travail en groupe, etc.).
- Le taux d'encadrement (ratio maître-élèves).
- L'utilisation du temps des enseignants (temps d'enseignement - temps de non-enseignement).
- La répartition journalière des activités scolaires.
- Le temps effectivement investi dans le travail (*time on task*).
- Le temps de travail scolaire en dehors de l'école (devoirs à domicile).
- La durée de l'année scolaire.
- Le climat de l'établissement et la façon d'exercer la discipline.
- Les formes d'indiscipline : école buissonnière, absentéisme, vandalisme, perturbations du travail.
- L'influence du groupe des pairs.

□ Des réalisations

• **En Angleterre : un exemple de checklists d'auto-évaluation**

D.L. Nuttall (1993) décrit les actions d'auto-évaluation (*School Self-Evaluation*, SSE) déployées au cours de ces dernières années au Royaume-Uni. Il apparaît que les checklists sont les instruments les mieux à portée des écoles.

Le premier grand pas a été franchi par la publication, en 1976, de *Keeping the School under Review* (Faire un examen critique permanent de l'école) (G. Rogers, 1977). Cet instrument, mis au point par les inspecteurs de l'*Inner London Education Authority*, contient des centaines de questions aidant les établissements à s'interroger sur leur organisation, leurs ressources, leur niveau de performances scolaires et les relations qu'ils entretiennent avec l'extérieur. Toute école peut évidemment se construire un tel instrument. Au lieu de vouloir tout évaluer à la fois, on peut aussi s'attaquer successivement à des aspects particuliers, de façon cyclique.

Voici les titres figurant dans liste de *Keeping the School under Review* pour l'enseignement secondaire (Nuttall, 1993, p. 43) :

- Les élèves. Le conseil d'administration de l'école, les parents, la communauté, le personnel de l'école.
- Le curriculum : mode d'élaboration. Curriculum des différentes années d'études. Documents d'enseignement. Options méthodologiques fondamentales. Evaluation. Egalité des chances filles-garçons. Dispositions en faveur d'élèves présentant des besoins particuliers.
- L'auto-évaluation des enseignants, leur formation continuée.
- Les modes de prise de décision et de communication à l'intérieur de l'école.
- L'environnement.
- Les ressources.
- Statistiques.

Tous les enseignants sont invités à s'interroger sur leur façon de travailler en classe et sur leur connaissance des élèves. De même, leurs directeurs doivent s'interroger sur leur utilisation du temps, sur les objectifs poursuivis et sur l'organisation de l'école.

- Questions-tests : (a) Recommanderai-je à un collègue de postuler un poste dans mon école ? (b) Recommanderai-je à des amis de placer leur enfant dans mon école ?

A propos de chacune des rubriques qui précèdent, des questions sont suggérées. Voici, à titre exemplatif, celles qui concernent la construction du curriculum :

- Comment les objectifs de l'établissement sont-ils définis ?
- Comment le programme est-il défini ? Est-il cohérent avec les objectifs ?
- Comment ce programme de l'établissement se situe-t-il par rapport au programme officiel et aux directives des inspecteurs ?
- La procédure de consultation associe-t-elle bien tous ceux qui travaillent dans l'établissement, ainsi que les parents, les administrateurs de l'école, les inspecteurs ?
- Quels changements importants devrait-on apporter à l'établissement au cours des prochaines années et quelle stratégie faut-il adopter à ce sujet ?
- Quelle est la politique de l'école en matière de langage ? Quelles dispositions prend-on pour améliorer l'habileté d'expression des élèves ?
- Au cours des dernières années, a-t-on beaucoup discuté, à l'école, de publications importantes en matière de curriculum ?

Nuttall signale que quarante-cinq versions de ce type de listes ont été publiées, depuis la fin des années 70, par les Autorités Locales d'Angleterre. Il observe que peu d'entre elles font des suggestions à propos de la manière de répondre aux questions, du traitement des réponses et de leur utilisation. Aussi, la façon de travailler a-t-elle beaucoup différé selon les écoles et le profit qu'elles ont retiré de cette auto-évaluation aussi.

Le Projet GRIDS (grilles) s'est avéré plus fécond. Il s'agit d'une démarche en cinq étapes qui s'est largement répandue dans le pays :

1. Phase préparatoire : plan de travail, estimation des dépenses, choix d'un coordinateur éventuel. La démarche d'auto-évaluation est décrite au personnel et l'on s'efforce d'obtenir sa coopération.
2. Evaluation initiale. Un questionnaire, auquel on répond anonymement, porte sur les forces et les faiblesses de l'établissement, telles que le personnel les perçoit. Il est invité à suggérer des aspects qui mériteraient spécialement d'être évalués. Les aspects à évaluer se dégagent d'une discussion générale des réponses.
3. Evaluation proprement dite. On s'efforce de la concentrer sur peu d'aspects. L'équipe constituée cherche à identifier les points forts et les points faibles des pratiques et fait des suggestions de changements. Seules les propositions acceptées unanimement par le personnel sont retenues.

4. Action. Les besoins de formation continuée du personnel sont définis et des mesures sont prises pour y répondre, généralement en collaboration avec l'aide du pouvoir organisateur. Cette phase est soigneusement pilotée : on vérifie si les changements intervenus entraînent les améliorations souhaitées et si des difficultés inattendues ne surviennent pas.
5. Evaluation et adoption. Le nouvel état de choses est comparé à l'état antérieur, de façon aussi objective que possible.

Fait important, la participation des écoles au projet est volontaire et les professeurs se sentent responsables de l'entreprise. La rédaction d'un rapport d'évaluation n'est pas imposée, et la communication, aux autorités, du document éventuellement rédigé n'est pas exigée. On veut ainsi prévenir le sentiment d'être mis en accusation.

Nuttall relève enfin que, dans l'expérience GRIDS, on a constaté que les écoles primaires s'intéressaient principalement à des problèmes pédagogiques (par exemple, l'enseignement de la lecture), tandis que les écoles secondaires se tournaient plutôt vers des problèmes d'organisation et de gestion. Des suggestions ont été faites dans l'espoir d'obtenir une approche plus équilibrée.

(Des grilles pour l'enseignement primaire et l'enseignement secondaire, ainsi que des manuels méthodologiques qui s'y rapportent, ont été publiées dans une collection "GRIDS" par Longman Resources Unit, York, 1988).

• **En France : la Proposition de Peretti**

A peu près à l'époque où le projet GRIDS se développe, une proposition d'A. de Peretti (1986), - qui ignore, à l'époque, les travaux britanniques (communication personnelle) -, s'inscrit dans la même direction : procéder à une évaluation *positive* de la *qualité* de l'enseignement dans chaque établissement. Un ensemble de 125 indicateurs est proposé. Ils devraient "permettre aux chefs d'établissement et aux personnels de *choisir*, en relation avec les familles et les élèves, *les indicateurs les plus adaptés* aux caractéristiques de leur établissement et aux efforts d'amélioration les plus opportuns" (p.357). Ce choix doit être *mobile* pour éviter des phénomènes d'usure et de colmatage.

Peretti suggère que des personnes de ressources accompagnent les personnels de chaque établissement dans l'évaluation de celui-ci. Des synthèses d'observations pourraient être envoyées à l'échelon régional ou national.

Cinq catégories d'indicateurs sont distinguées et, pour chacune, cinq aspects sont mentionnés :

A. Indicateurs d'organisation

- Qualité de la définition des buts (objectifs généraux, spécifiques; projet global; caractéristiques de l'établissement; connaissances des objectifs par les divers partenaires).
- Organisation différenciée des groupements d'élèves (principes de répartition des élèves dans les classes, règles de composition des groupes de niveau, critères d'organisation de groupes de méthodes, possibilité de groupements interdisciplinaires).
- Souplesse de l'emploi du temps (variations selon les objectifs et les disciplines, selon la pédagogie mise en oeuvre);
- Mesures destinées à rompre la monotonie ou à éviter des pertes de temps, facilitation du remplacement des enseignants, prise en compte des vœux méthodologiques des enseignants.
- Efficacité de la concertation des professeurs (mesures destinées à faciliter les rencontres, fonctionnement du conseil d'administration, cohérence des conseils de classe,...).
- Ordonnance des moyens (agencement et distribution des locaux, centres de documentation, laboratoires et ateliers, locaux mis à la disposition des élèves, adéquation des moyens, des horaires, du ramassage scolaire).

B. Indicateurs de fonctionnement

- Déroulement des enseignements (cohérence méthodologique, densité des innovations méthodologiques, degré d'utilisation du centre de documentation, degré de différenciation de la pédagogie, options disciplinaires offertes aux élèves).
- Ingénierie méthodologique et technologie éducative (importance et qualité des matériels, instruments d'observation et d'évaluation utilisés, audiovisuels, EAO, qualité du bulletin remis aux familles).
- Evaluation des apprentissages (fréquence de l'évaluation formative, cohérence des évaluations sommatives, unités capitalisables, auto-évaluation).
- Pratiques d'orientation et de redoublement.
- Mesures de formation (formation continuée des enseignants au sein et en dehors de l'établissement, enseignants en stages en entreprises,...).

C. Indicateurs d'activités

- Accueil des personnes (réunions de parents, accueil de visiteurs dans l'école,...).
- Aide aux élèves (études surveillées, tutorat, suivi des élèves en difficulté,...).
- Initiation culturelle.
- Productions et publications.
- Compétitions (scolaires, artistiques, sportives,...).

D. Indicateurs de relation avec :

- Le monde scolaire
 - Le monde scolaire et universitaire (relations avec d'autres établissements, etc.).
 - Nature des rapports d'inspection.
 - Pourcentage de conseillers pédagogiques parmi les professeurs.

- Les milieux familiaux (fréquence des messages adressés aux parents, ouverture de certains services aux familles, etc.)
- L'environnement (jumelage avec des entreprises, enquêtes sur le terrain,...)
- L'opinion (articles dans la presse locale, participation à des émissions de radio ou de TV, journées portes ouvertes,...)
- La collectivité intellectuelle (bibliothèques, choix des manuels, fréquentation de musées, de laboratoires de recherche; conférences de haut niveau dans l'établissement,...).

E. Indicateurs de cohérence

- Résultats (suivi du développement des élèves, maîtrise de l'expression orale et écrite, travail indépendant, réussite aux examens,...).
- Localisation des progrès potentiels.
- Régularité (image de l'établissement, taux de mutation des personnels, pourcentage de professeurs en formation,...).
- Projet d'établissement.
- Attachement à l'établissement.

Dès 1986, ce tableau d'indicateurs a été utilisé dans des journées de formation de chefs d'établissement et dans des universités d'été. Chacun des participants était invité à choisir, dans la liste, une trentaine d'indicateurs qui pourraient correspondre à son propre projet pédagogique. Ce choix servait de base de discussion entre les participants. Ils étaient amenés à expliciter les raisons de leur sélection et avaient ainsi l'occasion de découvrir pourquoi certains collègues, travaillant dans des situations proches de la leur, en avaient fait une autre.

• **Au Danemark**

En plus de programmes similaires à ceux qui viennent d'être évoqués, les écoles danoises peuvent aussi demander une expertise externe. Dans ce second cas, six mois après l'auto-évaluation, un inspecteur vient séjourner une semaine entière dans l'établissement; il rencontre les élèves, les enseignants, les parents et des représentants de la communauté locale. Un rapport provisoire est ensuite publié : seules les informations factuelles erronées peuvent être modifiées, mais non les jugements portés.

• **En Nouvelle-Zélande**

En Nouvelle-Zélande, l'autopilotage des établissements tend à se généraliser (parallèlement à l'instauration d'un pilotage national).

Les écoles sont invitées à faire le bilan de leurs performances tous les trois ans, avant de recevoir la visite d'un délégué de l'*Education Review Office* qui vient contrôler si les ressources mises à disposition ont été correcte-

ment utilisées et si les objectifs qui figurent dans le projet d'établissement ont été atteints.

La démarche de l'auto-évaluation varie selon l'âge des élèves et la grandeur de l'établissement; dans tous les cas, elle doit commencer par l'élaboration d'un plan d'évaluation précisant des indicateurs de qualité de l'éducation. La question toujours posée est : quels effets a eus la politique éducative adoptée par l'établissement sur les apprentissages des élèves ? Les indicateurs sont choisis par les autorités locales, en collaboration avec les enseignants et les autres agents d'éducation.

Selon leur nature, les indicateurs sont collectés à des moments variables et de façons différentes. Ils peuvent être empruntés aux indicateurs nationalement utilisés.

Comme tous ceux qui ont participé à la formulation du projet d'établissement sont associés à l'évaluation, il est généralement facile de trouver des personnes capables de collecter les informations et les données nécessaires. Les enseignants sont partiellement libérés de leur charge pédagogique pour collaborer au travail. Tout comme les autres personnes qui interviennent, ils doivent, eux aussi, être préparés à leur mission d'évaluation.

Un rapport est finalement rédigé. Il décrit la méthode utilisée, les indicateurs de qualité choisis et les résultats de l'évaluation. Il doit déboucher sur des propositions de mesures à prendre, tout en précisant quelles ressources nouvelles sont éventuellement nécessaires pour surmonter les obstacles identifiés et permettre des changements d'orientation dans le travail.

En un premier temps, ce rapport n'est communiqué qu'à ceux qui sont impliqués. Par la suite, il est remis à l'*Education Review Office* qui examine la façon dont les informations ont été collectées, les propositions formulées, et fait éventuellement d'autres suggestions.

Les approches qui viennent d'être évoquées, à titre exemplatif, ont tendance à privilégier l'aspect descriptif. Elles peuvent rendre de grands services à l'éducation. Commentant l'expérience anglaise en la matière, Gipps (1989) déplore, toutefois, que les rapports, souvent copieux, que les écoles rédigent sur leur propre fonctionnement, sont rarement lus par les parents et souvent négligés par les politiques.

□ Le recours à des experts extérieurs

L'autopilotage n'exclut pas le recours à l'expertise extérieure, qu'elle soit le fait de chercheurs, d'inspecteurs, de parents possédant des compétences particulières, ou de membres de l'administration. Cette aide extérieure,

qui peut être apportée à tous les stades du processus, est recommandée dans le Projet GRIDS, et Ruby et Wyatt (1988) évoquent une démarche similaire, expérimentée par le Département de l'éducation du Queensland (Australie) : toute école ressentant le besoin d'informations susceptibles de l'aider à résoudre un problème peut prendre contact avec les spécialistes du Service de recherche de l'Etat. Les chercheurs s'entendent avec l'école sur les informations à réunir, ainsi que sur la façon de les analyser et de les présenter. Les chercheurs ne se substituent pas à l'école pour résoudre les problèmes; ils se limitent à un rôle de soutien. La prise de décision et la conduite des actions restent donc de la responsabilité de l'école.

□ Conclusion

Le micropilotage procède d'une attitude générale, d'une philosophie de l'action : ne pas se contenter de belles intentions, mais toujours s'obliger à vérifier, aussi objectivement et systématiquement que possible, si elles ont vraiment été suivies d'effet, et dans quelle mesure. En fait, c'est une nouvelle "culture d'établissement" qui doit se développer.

L'une des conditions du succès de l'autopilotage réside dans la préparation des enseignants à cette nouvelle mission. Le concept d'indicateurs est relativement nouveau dans les écoles. Leur utilisation demande non seulement l'acquisition de connaissances et d'habiletés nouvelles, mais aussi une transformation globale de l'état d'esprit dans lequel les écoles ont fonctionné. Les maîtres n'ont pas toujours été accoutumés à contrôler systématiquement l'efficacité de leur enseignement.

Certes, les enseignants reconnaissent généralement l'importance de la collecte systématique des indicateurs, dont il vient d'être question, mais, en pratique, cette approche est plus souvent considérée comme une charge supplémentaire que comme une composante essentielle et intégrale du processus d'enseignement.

La réticence à évaluer rigoureusement les programmes et les pratiques peut être due à plusieurs facteurs. Le poids d'une lourde journée d'enseignement laisse peu de place pour planifier et effectuer ce qui est demandé et pour s'interroger sur les observations faites. Il ne faut pas, non plus, sous-estimer l'influence de l'état d'esprit de l'après-mai 1968, majoritairement adversaire de toute mesure quantitative des performances scolaires.

Enfin, les innovations se succèdent dans l'enseignement et l'on en attend souvent des effets presque immédiats. Or le pilotage est une action complexe aux effets plus ou moins lents à se manifester; il demande une attention et des efforts prolongés.

lée dans
imilaire,
stralie) :
à résou-
ce de re-
nforma-
ter. Les
es; ils se
actions

On ne doit pas, non plus, ignorer la crainte d'être accusé. Si les indicateurs révèlent des situations peu favorables, ne va-t-on pas automatiquement en rendre les enseignants responsables ? Ce serait évidemment une grave erreur. L'un des moyens les plus sûrs et les plus justifiés pour l'éviter consiste à plus focaliser l'attention sur ce qui explique les résultats observés que sur les résultats mêmes. L'expérience montre que la grande majorité des enseignants sont toujours heureux d'être mis en situation de pouvoir encore mieux faire leur métier.

Bref, par rapport au macropilotage, le micropilotage permet de concentrer beaucoup plus l'attention sur les aspects qualitatifs, contextuels. Il ne procède pas de l'audit, mais se veut essentiellement une analyse, une réflexion sereines sur ce qui se passe et s'est passé dans l'école, ce qui n'empêche pas la rigueur.

Réalisé dans un esprit démocratique, le micropilotage insiste beaucoup moins sur les carences que sur les façons de mieux faire.

d'une
inten-
ment et
été sui-
ouvelle

a prépa-
teurs est
on seule-
aussi une
ctionné.
quement

nce de la
mais, en
arge sup-
proces-

pratiques
nseigne-
mandé et
sous-es-
ment ad-

on en at-
ion com-
attention

Le pour et le co

□ Les bénéfiques escomptés

Ils sont de six ordres :

- Une meilleure rationalité politique.
- Une action éducative généralement plus efficace.
- Une plus grande équité. Pour l'assurer, il faut commencer par débusquer les inégalités, puis les combattre tout en veillant au respect des différences.
- Une information démocratique de tous les membres de la communauté éducative.
- Un développement de la recherche évaluative en éducation.
- Des informations précieuses pour la pratique de l'enseignement.

Analysant les effets secondaires bénéfiques du pilotage pratiqué, en Angleterre, par l'*Assessment of Performance Unit*, Gipps (1989) estime que la période (souvent plusieurs années) qui s'écoule entre deux évaluations peut être utilisée pour creuser les résultats. En particulier, l'analyse des erreurs apporte de précieuses informations aux enseignants. Et Gipps conclut : "Il est ironique que l'évaluation nationale, qui a été accueillie avec beaucoup de réserves par les enseignants, s'est avérée une précieuse source de feedback direct aux praticiens."

• Une rupture d'isolement

Dans l'étude où il situe la problématique des indicateurs et du pilotage dans le contexte de la politique de l'éducation de la Communauté française de Belgique, J. Magy (1992, pp.152-153) s'attache à un aspect généralement négligé. Il relève avec raison qu'à bien des égards les parents, les élèves, les éducateurs et les cadres pédagogiques sont largement abandonnés à eux-mêmes lorsqu'ils doivent prendre des décisions d'importance majeure. Ils sont, en effet, loin de toujours disposer d'informations suffisantes pour éclairer leur choix.

Schématiquement, J. Magy évoque les cas suivants :

- Parents et élèves devant choisir une école sans être informés des qualités réelles des divers établissements.
- Chefs d'établissement peu ou prou informés des performances et d'autres caractéristiques pédagogiques et managériales des autres établissements.
- Pouvoirs organisateurs et inspecteurs manquant de données de référence à propos d'ensembles d'établissements ou à propos de ce qui se passe effectivement dans d'autres systèmes éducatifs.

- Enseignants et formateurs manquant d'outils de référence pour définir des objectifs et des niveaux de maîtrise, et souvent mal informés à propos des instruments d'évaluation existants et de la manière de les utiliser adéquatement. Or ces outils peuvent contribuer au progrès de la "professionnalité" des maîtres et par là de l'éducation même.

D'évidence, si les diverses formes de pilotage étaient appliquées en se basant - notamment - sur des ensembles d'indicateurs significatifs, collectés de façon experte avec la collaboration de tous les membres de la communauté éducative, en particulier les enseignants et les cadres pédagogiques et administratifs, on serait en droit d'espérer non seulement une augmentation directe de la productivité (au sens noble de ce terme) du système éducatif, mais aussi, de façon indirecte, une augmentation importante de la qualification des acteurs.

L'expérience acquise lors des grands projets évaluatifs, réalisés au cours de ces dernières décennies, montre que la réflexion fondamentale qu'ils exigent et la maîtrise progressive des instruments et des techniques qu'ils suscitent, apportent latéralement un bénéfice souvent supérieur au profit immédiat permis par les évaluations.

□ Les dangers

Même s'il semble pleinement justifié de vérifier dans quelle mesure les objectifs assignés à l'éducation sont atteints (et cette évidence est plus grande encore si la communauté éducative, à commencer par les enseignants, a participé à ce choix), il n'en reste pas moins que cette vérification est ressentie par beaucoup de professeurs comme une menace. Ne va-t-on pas les rendre seuls responsables d'insuffisances observées, des sanctions ne risquent-elles pas d'être prises à leur encontre (comme cela se voit aux Etats-Unis) ? Ces questions sont redoutables. N'intente-t-on pas des procès en réparation contre d'autres membres de professions intellectuelles (y compris les médecins) qui ont nui à leurs clients ? Pourquoi les enseignants jouiraient-ils de l'immunité ? Aucune raison ne justifie cette exception, à condition que l'éventuelle faute professionnelle grave puisse être prouvée de façon indiscutable. Ce sera rarement le cas, tant les responsabilités d'un échec scolaire - a fortiori d'une simple insuffisance de rendement - sont partagées.

Il semble de loin préférable d'adopter une attitude positive en optant toujours pour une évaluation formative : les informations fournies par les indicateurs devraient essentiellement servir à mettre sur le chemin d'améliorations à tous les échelons : politique éducative, amendement des programmes, perfectionnement de la didactique, aide aux élèves en difficulté, incitations à la coopération des parents,...

Cela étant dit, l'idée de départ reste valide : oui, il faut vérifier si les objectifs choisis sont bien atteints.

Il importe, toutefois, de tenir compte de déformations plus spécifiques dont Madaus (1988) a dressé un inventaire :

- Si les élèves, les enseignants et les cadres pédagogiques croient, même erronément, que les aspects touchés par les indicateurs sont importants, ils leur accorderont de l'importance. En particulier, si les maîtres pensent que de graves décisions dépendent des résultats aux tests de rendement, ils insisteront, dans leur enseignement, sur les matières sur lesquelles les épreuves portent.
- A mesure que l'on connaît les matières sur lesquelles les tests portent et la façon dont les acquis à leur propos sont évalués, les programmes enseignés dans les classes se modifient en conséquence. Cette modification n'est pas nécessairement blâmable : certains responsables de l'enseignement ont vu dans ce phénomène un moyen d'amener les enseignants à concentrer leurs efforts sur des objectifs particulièrement importants.
- Les maîtres sont très attentifs à la façon dont leurs élèves sont interrogés (réponses ouvertes, QCM, tests plus centrés sur les savoirs que sur les savoir-faire,...) et adaptent leur enseignement pour préparer à faire face à ces exigences, même s'ils les désapprouvent.
- On risque de considérer les indicateurs comme d'infailibles informations sur ce que produit l'école.

Plus fondamentalement encore, les critiques adressées à l'évaluation opérée à l'aide de tests de connaissances ne manquent pas. Ceux-ci portent le plus souvent sur les habiletés intellectuelles de base et sur des disciplines scolaires privilégiées : sciences, géographie, etc.

Ces apprentissages sont certes importants, mais ils ne correspondent qu'à quelques-uns des buts de l'éducation. Les préoccupations d'ordre moral, esthétique et de culture générale ne devraient jamais être perdues de vue. Apprendre à appliquer un esprit critique à tout ce qui fait notre civilisation reste l'un des objectifs majeurs.

Or ce qui est le plus difficile à évaluer, ce sont précisément les qualités humaines auxquelles on doit être le plus attaché : la capacité de résistance et le courage devant l'agression, l'investissement réel dans la vie professionnelle, la soif de justice et le sens social, la solidarité, le dévouement au bien public, ...

Autre aspect dont on ne tient généralement pas assez compte, même si l'on en parle beaucoup dans les discours : le contexte social dans lequel l'éducation s'opère. Nous avons besoin de beaucoup mieux comprendre les

conditions familiale et de savoir dans quelle mesure l'élève est prêt à apprendre quand il arrive à l'école. Plus généralement, il faut s'interroger sur le soutien sociétal apporté aux jeunes. Si nous souhaitons que le public comprenne bien la signification des performances scolaires et soit conscient des conditions environnementales dans lesquelles les écoles accomplissent leur mission, il est essentiel que les responsables soient les premiers à en mieux comprendre la nature.

Enfin, une objection d'ordre pratique vient encore d'être soulevée par les enseignants anglais qui, on l'a vu, font actuellement face à l'instauration d'un système de pilotage très développé. Les évaluations ne vont-elles pas consommer du temps aux dépens de l'action éducative ? Les maîtres ne vont-ils pas être confrontés à des techniques qu'ils ignorent ?

Sauf cas exceptionnels, les évaluations sommatives ne prennent que quelques heures. Quant aux évaluations formatives, elles font partie intégrante du processus d'enseignement et, bien conduites, elles favorisent la réussite scolaire; elles permettent donc plutôt de gagner du temps.

Quant aux aspects techniques inhérents aux tâches évaluatives demandées aux professeurs, ils sont fondamentaux et doivent faire partie de leur formation docimologique de base. Dans toute la mesure du possible, les enseignants doivent pouvoir construire les instruments d'évaluation dont ils ont besoin, et être aidés dans cette tâche, là où c'est nécessaire. Il n'en reste pas moins que la construction de certains instruments sophistiqués, leur étalonnage, leur validation, ainsi que certaines analyses statistiques complexes sont de la compétence de spécialistes, à qui il appartient de présenter les dossiers et les interprétations de façon compréhensible par toutes les parties intéressées.

□ A condition de s'en servir...

Dans leur audit du système d'enseignement belge, demandé à l'OCDE, R. Haby, R. in't Veld et J. Tschoumy écrivent (1993, p.114) : "Il ne suffit pas d'avoir un minimum d'instruments à sa disposition. Encore faut-il que se développe, parmi les divers acteurs concernés, un état d'esprit favorable à la circulation des données, à des échanges sur leur signification et leur éventuelle utilisation."

On distingue trois types fondamentaux de prise de décision : le processus rationnel, le processus organisationnel et le processus politico-bureaucratique.

Le problème de la rationalité du comportement est bien connu. Appliquée à notre propos, la question peut se formuler de la façon suivante : l'information apportée par les indicateurs suffit-elle seule à faire prendre les décisions les plus sages et à les faire exécuter ? Cette question est discutée avec beaucoup de clairvoyance par A. Ruby (1991, pp.15-16).

L'information, souligne en substance Ruby, devrait presque nécessairement permettre de raisonner en "professionnel" et de susciter ainsi les actions les mieux adaptées. Or ceci est un acte de foi : on croit que les personnes agiront raisonnablement si elles disposent d'informations suffisantes pour étayer leurs décisions.

On ne peut ignorer l'importance des données qualitatives et l'aspect toujours en partie subjectif du choix des buts de l'éducation et de l'adhésion à certaines valeurs. Ces options jouent un rôle fondamental dans la définition d'une politique de l'éducation, dans la gestion des écoles et dans la conduite des apprentissages. On ne peut non plus faire abstraction du rôle que jouent, dans la résolution des problèmes, les connaissances accidentelles, non scientifiques, nées de la vie quotidienne.

Par ailleurs, il existe presque toujours plusieurs façons de résoudre un même problème. Enfin, une information rationnellement élaborée n'est qu'un élément dans un processus de décision. Cette information ne doit pas être tenue pour preuve ou pour vérité, mais bien pour une hypothèse à essayer d'appliquer, à tester et finalement à rejeter ou à accepter, selon qu'elle ouvre ou non une voie possible ou désirable.

Ruby conclut : celui qui analyse sans passion sait qu'une situation relative à l'école implique :

- De multiples décideurs.
- De multiples et discutables définitions du succès.
- Un degré élevé de tolérance à propos des manières de traduire les définitions du succès dans les faits.
- L'utilisation d'informations loin d'être complètes et parfaites.
- Une homogénéité apparente malgré les possibilités de variation des services offerts par les écoles.

Dans le *processus organisationnel*, l'évaluation devient une démarche de routine qui aide à atteindre les résultats souhaités. Il importe de collecter les données adéquates et de mettre en oeuvre les méthodes les plus efficaces pour obtenir les informations nécessaires au pilotage.

Quant au *modèle bureaucratique*, il fait de l'évaluation *une arme politique* que l'on utilise ou non selon l'opportunité. L'évaluation est ainsi mise au service d'intérêts particuliers. Les "faits" ne sont retenus que s'ils conviennent.

Les trois modèles coexistent le plus souvent, mais avec des dominantes variables.

Ces observations et ces nuances justifieraient-elles le rejet d'un pilotage systématique ? Loin s'en faut. Reconnaître ses limites n'équivaut pas à une négation du besoin d'indicateurs apportant une information solide. Ce qu'une telle discussion fait surtout apparaître, c'est le besoin d'un *pluralisme méthodologique* dans la gestion et l'action.

Les grandes décisions appartiennent aux pouvoirs organisateurs, plus ou moins bien à l'écoute de l'ensemble des membres de la communauté éducative. Mais, on le sait, une prise de décision *au sommet* n'a d'effet que si *la base* en comprend la logique et le sens, et est réellement motivée pour passer efficacement à l'action souhaitée. C'est pourquoi il importe de promouvoir un versant interactif du pilotage.

Une première mesure consiste à associer les principales parties prenantes au choix des aspects qui feront l'objet des observations. Les intérêts peuvent évidemment diverger. Ainsi, les pouvoirs organisateurs peuvent, dans certains cas, avoir pour souci premier les coûts et les rendements globaux, tandis que les professeurs et les parents privilégieront sans doute la qualité de l'enseignement et les moyens concrets de l'assurer sur le terrain.

L'association des enseignants à l'organisation du pilotage est essentielle. Non seulement parce qu'ils connaissent le mieux le terrain, mais aussi parce que, s'ils ont joué un rôle dans la définition, la collecte et l'interprétation des indicateurs, ils seront plus aisément persuadés de leur utilité. Comme toute autre technologie, le pilotage doit être aimable pour l'utilisateur.

M. Crahay propose qu'à la collecte et à la première analyse des données communautaires ou régionales succède immédiatement la création d'un *groupe de réflexion* composé d'inspecteurs, de membres de l'administration, d'enseignants, de représentants des parents. Avec les évaluateurs, ce groupe examinerait les résultats de l'étude afin d'en circonscrire les apports les plus importants et de définir des axes d'amélioration de l'enseignement. Le groupe pourrait aussi suggérer de nouvelles priorités d'investigation. Les rapports d'évaluation rédigés par les spécialistes seraient amendés ou complétés en fonction des réactions des partenaires.

A cette première analyse communautaire ou régionale succéderait une seconde phase de pilotage au niveau local. Elle aurait pour objectif de renforcer l'investissement des enseignants, des chefs d'établissement et des inspecteurs dans le processus. Dans les différentes circonscriptions inspectoriales, des écoles seraient invitées à participer à l'opération. Des *groupes de pilotage local* seraient constitués. On en revient ainsi à la problématique du micropilotage, qui a déjà été rencontrée, et aux dispositions pratiques envisagées à cette occasion. Nous les répétons partiellement.

Lors des premières rencontres, les principes du pilotage seraient expliqués. Les résultats des évaluations communautaires ou régionales seraient commentés et les instruments examinés.

Ces instruments, éventuellement adaptés ou complétés, seraient mis à la disposition des écoles ou des enseignants désireux de mieux connaître leur situation propre. L'adaptation aux conditions locales devrait être encouragée. Elle participe au *principe d'hybridation*, dont l'importance pour le développement est aujourd'hui reconnue. Il s'agit d'identifier les éléments les plus valables dans chaque culture et de les faire interagir afin de développer des méthodes et des solutions originales. Ceci s'applique aussi à la rencontre entre la culture scolaire des agents du terrain et celle des experts. En laissant cours à la créativité de chacun, on peut penser que les chances de maximiser les profits du pilotage sont accrues.

Par ailleurs, il ne faut pas perdre de vue que le pilotage à grande échelle se fonde sur des échantillons d'établissements ou de populations. Même si la représentativité est assurée, les données obtenues n'ont que valeur moyenne. D'où, de nouveau, l'intérêt des actions locales, utilisant éventuellement des instruments spécialement adaptés ou conçus pour aider à répondre aux interrogations particulières.

Ce n'est qu'au moment où, à la lumière des observations et des réflexions, des actions ciblées se produisent dans les classes, que l'on peut espérer une augmentation de la qualité de l'enseignement dans le sens souhaité.

□ ... et de raison garder

La conclusion, nous la laissons à J. Nisbet qui écrit en substance, dans un récent rapport de l'OCDE-CERI (1993, p.10).

Actuellement, on insiste beaucoup sur les données dures, que ce soit dans les indicateurs de performances ou dans les méthodes utilisées pour réaliser des évaluations nationales. Mais on n'insiste pas assez sur la meilleure compréhension de la nature de l'apprentissage que nous avons acquise ces temps derniers. "Comme le management, l'éducation doit se soucier à la fois du contrôle et de la croissance. *L'Educational Testing Service* des Etats-Unis prédit que l'évaluation en matière d'éducation changera plus au cours des dix prochaines années que pendant les cinquante années précédentes. *Il est important que cette évolution ne soit pas considérée isolément, mais bien comme un élément de la réforme générale des curriculums et que l'objectif d'un meilleur apprentissage passe toujours avant celui de contrôler* " (les italiques sont de nous).

Ann
Aspects techni

□ Introduction

Parmi les différents aspects du pilotage, ce qu'apprennent les élèves, ce qu'ils sont capables de faire ou d'être, est ou devrait être la préoccupation majeure. C'est pourquoi cette annexe technique est essentiellement consacrée à ces aspects.

Evaluer l'ensemble des produits de l'enseignement est une mission impossible, non seulement parce qu'ils sont innombrables, mais aussi parce qu'on ne peut espérer un accord unanime sur leur importance respective. Seule une attitude pragmatique permet de contourner cet obstacle majeur : parmi les apprentissages généralement considérés comme fondamentaux, tant pour la résolution des problèmes de la vie quotidienne que pour un développement souhaitable de la personne, estimons au moins ce qui peut l'être de façon aussi valide que possible. C'est notamment le cas des connaissances et des habiletés intellectuelles de base.

A propos des aspects techniques de la construction des instruments d'évaluation (tests, échelles d'attitudes, opinionaires, ...), du montage des enquêtes, des méthodes d'échantillonnage, des analyses statistiques, c'est toute la science de l'évaluation qui est mise en jeu. A ce propos, nous ne pouvons que renvoyer aux ouvrages spécialisés. Seuls quelques aspects particuliers, dont certains ne sont venus à l'avant-plan que dans un passé récent, sont évoqués ici. Le point de départ est l'*évaluation de programmes*.

La *validité des mesures* et leur *fidélité* restent évidemment une préoccupation majeure. A cet égard, les retombées des progrès récents de la *psychologie cognitive* et le mouvement actuel en faveur de la *mesure des performances effectives* sont des sources de progrès majeurs.

Le concept de *validité pédagogique* prend place à côté de celui de *validité du contenu* : avant d'évaluer des apprentissages, n'importe-t-il pas d'établir si l'élève a eu l'occasion de les réaliser, si on lui a enseigné les matières en question ou s'il a eu d'autres possibilités de se les approprier ? Bref, l'élève a-t-il eu l'*occasion d'apprendre* ?

L'*administration d'épreuves* destinées à établir le bilan des apprentissages d'une vaste matière par un grand nombre d'élèves n'est guère possible sans recourir à l'*échantillonnage matriciel*. Par ailleurs, l'observation de l'évolution de la maîtrise d'une même notion à travers le temps nécessite la construction de *tests parallèles*. A cet égard, comme à plusieurs autres, la *théorie des traits latents* permet d'importants progrès.

L'*interprétation des résultats* soulève aussi de nombreux problèmes, qu'il s'agisse de définir des *standards* ou la *compétence minimale*, de l'*analyse du pourcentage de réussite*, de la *probabilité de réponses correctes* ou de l'*ancrage des échelles*.

Toutefois, avant d'aborder ces aspects, il importe de préciser la question fondamentale et générale des répercussions possibles du pilotage, en particulier de l'évaluation des risques qu'il peut faire courir et des modalités de l'indispensable recherche de consensus à son propos. On en revient ainsi à l'*évaluation sociétale des choix technologiques* à laquelle il a été fait brièvement allusion dans le chapitre consacré à la nature du pilotage.

□ Sur l'évaluation sociétale des choix technologiques

Ce type d'évaluation - idéalement anticipative - part de la constatation que la technologie est socialement déterminée : elle ne se développe ni comme une force autonome de la société, ni comme une sorte de voie de résolution de problèmes, rationnellement suivie en fonction des objectifs définis. De multiples interactions se produisent entre les groupes sociaux, et entre ceux-ci et la technologie. Ainsi s'explique que les différents groupes peuvent donner un sens différent aux phénomènes. C'est à partir de là qu'une sociologie de la technologie peut s'élaborer.

L'une des meilleures sauvegardes contre les effets néfastes des technologies réside dans la possibilité de les refaçonner à tous les stades de leur développement. Cette faculté est d'une importance évidente pour le pilotage.

Comme d'autres effets des technologies, ceux du pilotage pourraient être si subtils qu'ils risqueraient de passer inaperçus. Par exemple, le pilotage ne va-t-il pas permettre d'exercer un pouvoir occulte sur les enseignants ?

L'évaluation sociétale des technologies cherche à débusquer les problèmes qui se posent ou peuvent se poser, évalue les divers modes d'actions possibles et présente des conclusions.

Dans ce contexte, les chercheurs, les spécialistes de l'évaluation peuvent apporter une collaboration précieuse aux responsables politiques, aux décideurs qui, trop souvent, optent pour des solutions sans être assez informés de leurs conséquences majeures, ou en n'utilisant qu'une partie des données sûres apportées par la recherche.

Informers les décideurs des diverses solutions possibles est l'un des rôles importants que les scientifiques peuvent jouer. C'est ce que Vinck (1988, p.13) appelle *l'étalement des alternatives*. Il ne faut cependant pas perdre de vue que, même si les politiques sont loin de toujours pouvoir juger de la qualité réelle des solutions techniques proposées, ils disposent, eux, d'informations sociétales que beaucoup de chercheurs ignorent (acceptabilité psychologique dans le public et au sein des partis, limites budgétaires, compromis politiques et philosophiques,...).

Au négatif, cette évaluation aide donc à mettre en garde contre des conséquences indésirables. Au positif, elle peut confirmer l'intérêt des technologies, aider à les faire accepter par le public, et attirer l'attention des scientifiques sur leur responsabilités sociétales.

"L'évaluation sociétale des choix technologiques implique la participation des différents acteurs dans le processus. Il est donc nécessaire de permettre un débat permanent" (Vinck, 1988, p.15).

L'éthique *ethnographique* étudie la co-émergence des valeurs et des pratiques dans les relations de travail. Elle observe, en particulier, que des codes et des valeurs prennent forme lors de l'utilisation des technologies.

L'éthique *stratégique* est focalisée sur les stratégies développées par les acteurs pour définir les controverses comme des controverses éthiques et pour rechercher un consensus dans ce débat.

Cette recherche du consensus doit rester la préoccupation permanente.

Se référant à un rapport déposé au Parlement européen, en 1991, par la Direction Générale XII, Programme FAST, de la CEE, J. Berleur (1991) mentionne des méthodes d'évaluation qui, ensemble, doivent aider à appréhender les avantages et les dangers des technologies, et permettre à toutes les composantes de la société, et pas seulement aux spécialistes et aux décideurs, de donner leur avis à leur propos :

- Les analyses de la littérature.
- L'étude rétrospective et prospective de séries statistiques.
- Les enquêtes auprès de personnes exposées aux nouvelles technologies.
- Les études de cas.
- Les avis d'experts.
- Le développement de scénarios.
- Les *ateliers du futur (future shops)* : réflexion en ateliers de groupes d'une vingtaine de citoyens non experts, sur ce que pourrait être le futur. Cette réflexion comporte une évaluation de la faisabilité et des propositions d'actions.
- Les *conférences de consensus* : des non-spécialistes venant de divers horizons commencent par écouter une courte présentation d'une problématique par des experts. Les "non-experts" examinent leurs thèses et leurs propositions, et doivent se mettre d'accord sur un rapport final. Les experts prennent connaissance de ce rapport et ne peuvent y corriger que les seules erreurs factuelles. Le texte est ensuite présenté aux médias.
- L'expérimentation sociale en situations réelles.

Dans une étude méthodologique, à paraître en 1993, N.T. Nguyen, C. Lobet, J. Berleur et B. Kusters précisent que "l'idée d'une évaluation sociale des technologies neutre et utilisant une seule méthode semble définitivement abandonnée en faveur d'une approche plus focalisée sur le partenariat, le *dialogue* et la *concertation*". Les mêmes auteurs soulignent (p.7) que l'un des défis majeurs lancés à ceux qui évaluent les technologies est "d'être capables d'informer correctement et à *temps* toutes les parties concernées". Ceci est d'autant plus important que, surtout pendant la phase de développement de la technologie, plusieurs voies sont possibles et ne connaissant pas encore celle qui prévaudra, on ne peut pas encore dire grand-chose à propos des impacts sur la société. Plus tard, quand la technologie est bien installée, "les implications sociales sont claires, mais, malheureusement, il est trop tard pour influencer son développement."

La *Cellule interfacultaire de Technology Assessment (CITA)* des Facultés Notre-Dame de la Paix, à Namur, qui focalise de plus en plus ses efforts sur l'explicitation des controverses sociales qui surviennent lors d'une innovation technologique, distingue cinq tâches principales (pp. 9-11):

1. L'identification des conflits entre les *diverses perceptions sociales* et les *représentations* de la technologie.
2. La *description fine* des controverses concernant le choix de la technologie et de leur évolution.
3. L'élaboration de *cartes sociales* montrant les diverses catégories d'acteurs directement ou indirectement concernés par l'innovation technologique.
4. La proposition, à partir des hypothèses formulées lors des tâches précédentes, de *scénarios alternatifs* et la promotion du processus de *participation publique*.
5. De façon plus empirique, la réalisation d'*études de cas* approfondies ou d'*expérimentations sociales*, en collaboration avec les acteurs directement impliqués dans le développement technologique.

Le mot clé de cette approche est la *participation* et, méthodologiquement, les approches qualitatives et quantitatives se combinent.

□ Sur l'évaluation de programmes

La technique du pilotage procède de l'évaluation de programmes.

Un programme se définit de façon générale comme un ensemble intégré d'activités conçues pour atteindre un but commun. L'évaluation a pour objet d'établir dans quelle mesure les effets observés correspondent aux buts poursuivis et, éventuellement, si ces buts mêmes ne devraient pas être modifiés.

En matière d'évaluation de programmes, on retrouve toutes les grandes tendances de la recherche évaluative contemporaine : approches quantitative ou qualitative, nomothétique, historique ou éclectique. L'analyse "coût - bénéfice" constitue une catégorie particulière.

Appliquant des plans expérimentaux ou quasi expérimentaux, l'évaluation quantitative étudie les relations entre des variables d'entrée (*input*) et des variables de sortie (*output*) prédéfinies, afin de déterminer la productivité globale et aussi celle des différents "investissements". Le produit est le plus souvent exprimé en termes de rendement scolaire, de continuation des études ou d'emplois obtenus à la sortie du programme.

Beaucoup d'indicateurs procèdent de simples dénombrements : montant des budgets, nombre d'élèves ou de professeurs, durée de l'année scolaire, taux d'absentéisme, pourcentage d'un groupe d'âge obtenant un diplôme de fin d'études secondaires longues, etc. Il importe évidemment que ces relevés soient correctement effectués et que l'on ne fasse pas dire aux chiffres ce qu'ils ne signifient pas. Par exemple, la signification statistique de différences de pourcentages doit être systématiquement établie. Même si elles sont grandes, il reste à les interpréter correctement.

Avec l'approche qualitative, le but poursuivi est d'abord de comprendre à quoi sont réellement dus les effets observés. Elle tient compte du fait que les objectifs de l'évaluation ne se précisent souvent qu'en cours de route. Une large place est réservée à la description aussi fine que possible des circonstances de départ et des événements qui se produisent dans la phase de mise en oeuvre. Dans l'arsenal de l'évaluateur, l'ethnométhodologie occupe une place croissante. On s'efforce de découvrir comment, dans un environnement particulier, chacun des acteurs a vécu le programme, ce qu'il en a effectivement retiré et les questions qu'il se pose à son propos. D'où le recours fréquent à l'étude de cas, à l'observation participante.

L'un des principaux reproches fait à cette démarche est de ne pas placer, au premier rang des préoccupations, les raisons qui ont conduit à lancer le programme. Or, en vraie démocratie, la décision a résulté d'un accord de nature politique et sociale, pris par l'ensemble des représentants de la com-

munauté éducative. Il est difficile d'admettre que, lors de l'évaluation du programme, on puisse faire abstraction des intentions de départ.

Dans la recherche en éducation, l'opposition entre l'évaluation qualitative et l'évaluation quantitative a alimenté un débat passionné pendant près d'un siècle. Même si la controverse subsiste, elle s'est au moins apaisée. Toute qualité n'existe-t-elle pas en une certaine quantité ?...

Comme son nom l'indique, *l'approche éclectique* cherche à tirer simultanément profit des deux approches précédentes : la recherche des relations causales (analyse quantitative) est complétée par des analyses de processus et de contexte (analyse qualitative). On tient compte non seulement des objectifs assignés initialement, mais aussi de ceux qui sont venus s'ajouter ou se substituer, en cours de travail, quand l'intérêt s'en est présenté.

L'analyse coût - bénéfice, issue de l'économie, spécialement de la science de la gestion, est à dominante quantitative. En général, ce type d'évaluation n'a guère réussi dans le domaine de l'éducation, en raison même de la forte composante affective de celle-ci. Par exemple, outre le bénéfice matériel immédiat qu'entraîne la suppression de l'échec scolaire, celle-ci a d'autres conséquences importantes : meilleure image de soi, accroissement de la motivation à faire des études, meilleure réussite dans la vie adulte, etc. Or ces "bénéfices", qui sont probablement les plus importants, sont difficilement quantifiables.

Actuellement, la réalisation des travaux de pilotage est le plus souvent confiée à des organismes centraux. C'est notamment le cas aux Etats-Unis, en Angleterre et en France. Ces organismes sont soit des services administratifs (par exemple, la Direction de l'Evaluation et de la Prospective française dépend du Ministère de l'Education nationale), soit des centres de recherche et de développement (par exemple, *Educational Testing Service, ETS*, aux Etats-Unis). On peut cependant concevoir une décentralisation, à condition d'assurer une coordination rigoureuse.

Pratiquement, aucun organisme de pilotage ne fonctionne sans recourir à des experts extérieurs et à des services de recherche en éducation (qui interviennent, le plus souvent, en réponse à des appels d'offres).

Sur l'évaluation qualitative, on consultera l'ouvrage fondamental de M.B. Miles et A.M. Huberman, *Analyse des données qualitatives*, Bruxelles, De Boeck, 1990.

□ Psychologie cognitive et perspectives nouvelles

Tant pour le pronostic que pour les épreuves de rendement ou de diagnostic, les progrès récents de la psychologie cognitive ouvrent des perspectives nouvelles. Même si les notions théoriques qui vont être évoquées - de façon très simplifiée - ne peuvent pas encore déboucher aisément sur des pratiques quotidiennes, elles aident au moins à prendre conscience de problématiques fondamentales.

Au lieu de se borner à constater qu'un élève n'a pas trouvé la solution d'un problème, n'a pas été capable d'accomplir telle tâche, la psychologie cognitive veut savoir quelle démarche mentale explique le phénomène. Grâce surtout aux travaux de Piaget, tout éducateur sait que les possibilités d'un élève sont liées au stade de développement cognitif, affectif, psychomoteur qu'il a atteint. Toutefois, la seule connaissance du niveau de développement cognitif ne permet pas de tout expliquer.

Par comparaison avec l'ordinateur (métaphore que la psychologie cognitive a privilégiée pendant un moment), le traitement de l'information par le cerveau peut être décrit de la façon suivante :

- Les structures du système cognitif peuvent être activées par des stimuli reçus par les organes des sens ou par d'autres structures de connaissances (pensée); des informations sont ainsi *entrées*.
- La *mémoire à court terme* retient un nombre limité d'informations pendant une courte période de temps. Les faits isolés sont oubliés les premiers.
- La *mémoire à long terme* stocke de façon codée et permanente les concepts et les constructions mentales dérivées de l'expérience passée de la personne (événements, relations, processus, affects,...) et dirige les opérations de tout le système de traitement de l'information. Cette *base de données* est le coeur du système cognitif. La capacité de cette mémoire est illimitée.
- Le *système musculaire* exécute les actes moteurs : parler, courir,...

L'extraordinaire performance du cerveau ne peut s'expliquer que par l'existence :

- de procédures tellement rapides qu'elles ne demandent plus beaucoup d'attention (automaticité);
- de stratégies permettant de sélectionner, planifier, piloter et, au besoin, modifier les actions (habiletés métacognitives);
- de structures cognitives (*schémas*) qui spécifient les propriétés générales d'un objet ou d'un événement. Il s'agit donc d'abstractions qui re-

présentent des ensembles significatifs d'informations stockées en mémoire (par exemple, tout le déroulement normal d'une opération chirurgicale).

L'apprentissage consiste en automatisation de procédures, en acquisition et en développement d'habiletés métacognitives, et en construction, révision et remplacement des schémas.

Par rapport aux "novices", les "experts" se caractérisent par le nombre croissant d'interconnexions qu'ils établissent entre un nombre de faits de plus en plus grand, par l'ampleur et la qualité des schémas qu'ils organisent et par une automatisation poussée des procédures, ce qui leur permet de s'attaquer à de nouveaux aspects des problèmes, de se placer dans des perspectives différentes.

Quant aux *représentations*, ce sont des "théories personnelles", des systèmes de croyances relatifs à des phénomènes. C'est la façon dont un individu ou un groupe les comprend, se les explique. Par exemple, croire que l'électricité coule dans les fils comme l'eau dans des tuyaux, parce que, dans les deux cas, on parle de *courant*... Souvent tenaces, les représentations ne sont pas faciles à débusquer, ni à évaluer. Plus elles sont incorrectes et inadéquatement articulées, plus elles rendent difficiles de nouveaux apprentissages.

• **Implications pour l'amélioration des tests de performances scolaires**

A la lumière des apports de la psychologie cognitive, l'évaluation scolaire devrait remplir les fonctions suivantes (H. Walberg, 1991) :

- Mesurer la quantité de connaissances déclaratives (faits, concepts, formules, ...) que possède l'élève et identifier le type d'organisation qu'il utilise pour les stocker.
- Mesurer la vitesse d'exécution des tâches cognitives. Le taux d'attention requis peut être manipulé par l'introduction de tâches concurrentes ou par une augmentation de la difficulté des tâches.
- Apporter rapidement un feedback aux élèves et attirer leur attention sur les erreurs de processus ou de contenu.
- Diagnostiquer le type des erreurs et leurs sources.
- Déterminer, par l'analyse des erreurs, le degré d'adéquation, d'étendue et de flexibilité des schémas.
- Identifier les stratégies de résolution de problèmes.
- Observer dans quelle mesure l'élève est capable de varier ses stratégies de résolution de problèmes lorsqu'il ne trouve pas directement les solutions.

- Identifier les stratégies métacognitives utilisées par l'élève; repérer celles qui l'aident et celles qui ne l'aident pas. Il importe de distinguer les habiletés métacognitives générales et les habiletés métacognitives spécifiques à un contenu.
- Pour évaluer la qualité des stratégies de résolution de problèmes, prendre pour référence la façon dont les experts perçoivent le problème, élaborent et choisissent un plan de résolution et mettent la stratégie arrêtée en oeuvre.
- Evaluer la qualité des réseaux sémantiques.
- Evaluer le degré d'automatisme des habiletés intellectuelles.

• **En conclusion**

- Des techniques d'évaluation permettent une certaine connaissance des processus cognitifs de l'élève.
- L'utilisation de tests de connaissances à facettes multiples s'impose (voir ci-après la note consacrée à la généralisabilité). En particulier, la façon de questionner doit être variée, de même que le niveau de difficulté. On peut ainsi estimer dans quelle mesure la façon de poser les questions influence les performances observées.
- L'évaluation doit être dynamique. On ne doit pas se contenter d'observer ce que les élèves sont capables de faire par eux-mêmes, à un moment donné. Il faut aussi déterminer jusqu'où ils peuvent aller lorsqu'on les met sur le chemin de la solution (étude du *développement proximal*, selon l'expression de Vygotsky, 1978).

Il s'impose, toutefois, de rester prudent à propos de toutes ces propositions. On ne dispose que de peu d'études statistiques à leur sujet. La validité de contenu, la généralisabilité des épreuves doivent être établies. De surcroît, les démarches proposées sont coûteuses.

□ **Sur la validité et la fidélité des mesures**

Comme toute autre mesure, les indicateurs doivent posséder deux qualités essentielles, sans lesquelles les autres considérations n'offrent plus guère d'intérêt : la validité (c'est-à-dire, mesurer effectivement ce qu'ils prétendent mesurer) et la fidélité.

En particulier, la validité interne d'un indicateur est menacée si l'on n'opère qu'une mesure partielle d'un phénomène multidimensionnel. Par exemple, si l'on réduit un test d'intelligence à des épreuves de mémoire, il est clair que les scores obtenus ne refléteront pas vraiment le niveau d'intelligen-

ce du sujet. De même, un indicateur ne portant que sur un aspect marginal d'un phénomène éducatif n'offre guère d'intérêt et risque d'ailleurs de conduire à des conclusions inexactes.

Il en va de même à propos des relations que l'on peut établir entre certains indicateurs. Ainsi, prendre, sans autres considérations, le taux de redoublement comme indicateur de la qualité pédagogique risque bien d'être fallacieux aussi longtemps que la décision d'arrêter ou non un élève relève, à la limite, d'un arbitraire qui s'ignore.

La validité est donc faussée :

- si le phénomène à évaluer est défini de façon inadéquate;
- si la variable substituée à une autre pour des raisons de facilité n'est pas valide. Par exemple, le nombre d'années d'études que prend la formation initiale des maîtres ne nous apprend rien de certain sur leur qualification (ce nombre d'années ne peut donc servir de variable substitutive);
- si l'objet de la mesure a changé au cours du temps;
- s'il y a eu "corruption". Cette falsification se produit, par exemple, si l'enseignant souffle les réponses aux questions d'un test.

Quant à la fidélité, elle a évidemment partie liée avec la validité : si les pesées répétées d'un même objet indiquent des poids différents, elles manquent non seulement de fidélité, mais aussi de validité. La théorie de la généralisabilité a beaucoup à nous apprendre à cet égard.

• **La généralisabilité**

Par généralisabilité d'une mesure, on entend le degré auquel on peut, à partir d'une observation ou d'une épreuve particulières, tirer des conclusions sur la valeur théorique recherchée. Plus concrètement, les questions posées permettent-elles d'évaluer le niveau de connaissances des élèves de façon fidèle ? On retrouve donc ainsi le problème de la fidélité et de la validité.

Les résultats observés à un examen varient notamment :

- selon le contenu et la forme des questions posées (facette questions);
- selon le moment où l'épreuve est subie (facette occasions);
- selon les points de vue et la qualité des examinateurs (facette juges).

Par exemple, à propos de la facette questions, il importe de considérer :

1. *Le langage utilisé*

- Influence du vocabulaire utilisé et de la syntaxe (lisibilité psychologique).

- Mode de présentation matérielle et psychologique. Par exemple, bien que portant exactement sur le même problème, les questions suivantes peuvent donner des résultats différents :

. $3 + 5 = ?$

- . J'achète un crayon à cinq francs et un bonbon à trois francs, combien dois-je payer ?

2. La nature de la question

J. Cardinet aime à donner l'exemple suivant à propos du calcul de l'aire du triangle.

Dans un groupe qui a étudié cette question pendant l'année scolaire, on obtient des résultats différents selon que l'on demande :

- Quelle est la formule de calcul de l'aire du triangle ?
- Quelle est la surface d'un triangle de 8 m de base et de 5 m de hauteur ?
- Voici un triangle et une règle graduée. Calculez la surface de ce triangle ?
- Même question que la précédente, mais la forme du triangle est telle que la perpendiculaire abaissée à partir du sommet tombe en dehors de la base.

Bref, la connaissance d'une formule et de son utilisation dans des cas stéréotypés ne suffit pas à témoigner d'une capacité réelle. Seules des performances concrètes, suffisamment variées dans leurs conditions pour constituer un échantillon représentatif des cas qui peuvent se présenter, permettent une évaluation valide.

A propos de cette facette "occasions", on peut aussi poser les questions suivantes :

- L'élève a-t-il eu effectivement l'occasion d'apprendre ce dont il va être question ?
- Cet apprentissage est-il récent ou déjà ancien ? Dans ce second cas, la connaissance a-t-elle été réactivée ou est-elle restée dormante ?
- Quel pouvait être le degré de fatigue de l'élève quand il a subi cette épreuve ?
- Une même question est-elle posée sous des formes diverses ?

Quant à la facette "juges", elle est souvent évoquée en docimologie : inconsistance des appréciations, effet de halo, effet de contraste, stéréotypie, effet de fatigue du correcteur ou de l'interrogateur,...

• **L'estimation de l'occasion d'apprendre**

Avant de se prononcer sur une performance en apparence insuffisante, il s'indique de vérifier si l'élève a effectivement eu l'occasion d'apprendre les connaissances ou les habiletés testées.

Pour estimer quel pourcentage des élèves ont reçu un enseignement adéquat, on peut soit interroger séparément les enseignants, soit demander à l'ensemble des professeurs concernés de se mettre d'accord sur une estimation. Dans les enquêtes de la NAEP, on ne retient les questions que si au moins 75 % des élèves ont eu l'occasion d'apprendre les matières sur lesquelles elles portent. De nouveau, des traitements statistiques complexes (dont la statistique de Mantel-Haenszel) sont utilisés pour nuancer les résultats obtenus initialement.

• **Epreuves de performance ou épreuves de récitation ?**

Si les systèmes d'indicateurs doivent aider à focaliser l'action éducative sur les savoir-faire importants, il importe de les mesurer vraiment. La validité est ce prix. Le débat à ce propos est particulièrement aigu aux Etats-Unis où beaucoup ne jurent plus que par les *épreuves de performance*, c'est-à-dire d'épreuves qui obligent l'élève à produire effectivement le comportement dont il est question ou, au moins, à construire sa réponse, au lieu de n'avoir qu'à la choisir toute faite dans sa mémoire ou en appliquant une démarche de résolution de problème routinière et stéréotypée.

Ne relèvent donc pas des épreuves de performance des questions telles que :

- Quelle est la formule de l'acide sulfurique ? (Pure mémoire).
- Sachant que l'aire d'un triangle se calcule par la formule $\text{base} \times \text{hauteur} / 2$, quelle est l'aire d'un triangle de 25 cm de base et de 30 cm de hauteur ? (Compréhension élémentaire).
- Calculez l'aire d'un triangle de 25 cm de base et de 30 cm de hauteur. (Application stéréotypée).

Dans la taxonomie de B.S. Bloom (domaine cognitif) et selon le sens conventionnel qui y est attribué aux termes *mémoire*, *compréhension*, *application*, ces trois questions font appel à des processus cognitifs inférieurs.

Ne relèvent pas non plus des épreuves de performance des examens où l'on peut répondre, pour l'avoir appris de mémoire, ce qu'il faudrait faire dans un cas donné, mais où l'on ne le fait pas effectivement. Exposer oralement ou par écrit la pédagogie progressiste que l'on pratiquerait si l'on était professeur ne prouve ni qu'on est capable de la mettre en oeuvre, ni qu'il est possible de le faire étant données les limites circonstancielles qui existent dans la plupart des situations scolaires.

Les questions à choix multiple excluraient-elles vraiment toute performance relevant des processus cognitifs supérieurs? Assurément non. Par exemple, la question suivante exige manifestement la mise en oeuvre de processus cognitifs supérieurs et relève donc bien des épreuves de performance.

Le parquet d'une chambre doit être fait de lattes entières. A cet effet, M lattes de a cm sur b cm sont nécessaires. Combien faudrait-il de lattes mesurant x sur y cm ?

A	B	C	D	E
$\frac{Mab}{xy}$	$\frac{ab}{Mxy}$	$\frac{(a+b)M}{x+y}$	$\frac{ab \cdot xy}{M}$	$\frac{Mxy}{ab}$

La difficulté de construction des épreuves de performance réside donc dans la conception de questions ou de tâches qui permettent d'évaluer de façon valide ce que le sujet est effectivement capable de faire.

Trois grandes voies s'offrent en matière de testing de performance et elles commencent à être exploitées dans le domaine des indicateurs, notamment dans plusieurs Etats des Etats-Unis (Vermont, New York, Connecticut) et au Royaume-Uni :

1. Pour les disciplines théoriques (philosophie, logique, algèbre, physique,...), poser des questions qui exigent la mise en oeuvre de processus cognitifs supérieurs.
2. Faire effectuer la performance-cible (par exemple, faire une leçon de langue étrangère devant les examinateurs ou, autre exemple, demander aux élèves de réaliser une expérience scientifique qui peut être longue et d'en rapporter le déroulement).
3. Examiner des productions réelles à travers lesquelles la qualité de la performance peut être estimée. Il s'agit alors d'un *examen sur dossier* (le *portfolio* des Anglo-Saxons). Par exemple, pour faire valoir ses capacités pédagogiques, un professeur peut constituer un dossier qui réunit des préparations de leçons, du matériel didactique qu'il a confectionné, des enregistrements vidéo de ses leçons, des travaux d'élèves, des enregistrements de conversations avec des parents d'élèves, des articles publiés, etc.

Le recours aux dossiers pour évaluer les élèves offre aussi des pistes intéressantes, mais les problèmes ne manquent pas. En effet, la qualité du dossier varie non seulement en fonction des acquis des élèves, mais aussi se-

lon leur motivation, leur plus ou moins bonne compréhension de l'usage qui sera fait du dossier et de la façon dont il sera évalué. On observe aussi des différences selon les styles cognitifs.

Même si elles répondent à un idéal, les épreuves de performances réelles soulèvent d'importants problèmes pratiques, à commencer par le temps considérable que prend une observation directe suffisante de l'exécution de tâches complexes comme, par exemple, la réalisation d'expériences scientifiques, suivies de leur interprétation. Une seule expérience ne suffit pas. Il en faut plusieurs pour obtenir un échantillon réellement représentatif permettant une évaluation fidèle et valide du savoir ou du savoir-faire des élèves.

Nécessité de procédures de modération

L'utilisation de tests de performance à des fins d'évaluations comparatives (niveau régional, national, international) pose aussi des problèmes mal résolus. En effet, les différences de tâches rendent les comparaisons difficiles. De surcroît, le degré de familiarité des élèves avec le type d'épreuves influence les résultats. L'occasion d'apprendre varie aussi.

Au Royaume-Uni - qui a accumulé une grande expérience en la matière - on recourt systématiquement à des procédures de "modération" des notes (ces techniques sont décrites dans notre *Précis de docimologie*, 1992).

Des enseignants et des spécialistes de la construction des curriculums sont réunis pour analyser les travaux des élèves, pour déterminer dans quelle mesure les tâches proposées, les exigences et les démarches d'évaluation sont comparables.

L'association des maîtres aux procédures de modération entraîne une augmentation de leur qualification professionnelle et rend possible la comparaison des tâches et des critères de correction.

Sur les portefeuilles de compétences professionnelles, voir J. Aubret (1992).

□ Sur l'administration des épreuves

• L'échantillonnage multimatriciel

Pour établir dans quelle mesure les élèves ont réalisé les principaux apprentissages prévus dans les programmes, la première idée qui vient à l'esprit est de tester la population entière sur l'ensemble des matières. Ce n'est pratiquement pas possible.

L'échantillonnage multimatriciel consiste à échantillonner à la fois les élèves et les tâches. Par exemple, on divise l'ensemble des questions que l'on

voudrait poser en 20 groupes de tâches (items). Si l'on estime, par ailleurs, que l'examen d'un échantillon de 30 classes permet de se faire une idée correcte du rendement pour un ensemble de tâches, il faudra donc tester 600 classes.

Les résultats regroupés pour l'ensemble informent sur le pourcentage de réussite par groupe de tâches et sur le pourcentage d'élèves atteignant des niveaux de performance déterminés. C'est donc le rendement de groupes et non d'individus qui est évalué.

En conservant une partie commune de tâches dans les évaluations qui se succèdent, on peut aussi déterminer les évolutions.

• **La mise en parallèle des tests**

Pour comparer, à mesure que le temps passe, les progrès réalisés dans l'apprentissage d'une même matière, on peut, comme cela se fait couramment dans les études longitudinales, garder un noyau de questions communes dans les épreuves administrées à différents niveaux scolaires (*parallélisation verticale*).

La *parallélisation* est dite *horizontale* quand les tests sont subis, à différentes occasions, à un même niveau scolaire, éventuellement par les mêmes élèves, ce qui nécessite la construction de tests parallèles, c'est-à-dire équivalents au point de vue du contenu et de la difficulté.

La mise en parallèle (*equating*) basée sur l'hypothèse de l'existence de traits latents est une méthode récente. Par exemple, A. Beaton (1987) l'a utilisée pour la parallélisation verticale des épreuves de lecture de la NAEP (9 ans, 13 ans, 17 ans).

Sur la théorie des réponses aux items (IRT) et sur la notion de traits latents, voir ci-après la note consacrée à l'ancrage des échelles.

□ **Sur l'interprétation des résultats**

• **La notion de standard**

Certains indicateurs résultent de simples dénombrements. Dans les cas élémentaires, on se borne à constater l'existence ou non d'un phénomène. Exemple : pourcentage d'un groupe d'âge poursuivant des études au-delà de l'enseignement secondaire long.

Mais, dès qu'il s'agit du progrès de l'éducation, du développement des individus, on doit pouvoir prendre des repères descriptifs, définir des niveaux à atteindre. Ces repères qualitatifs sont généralement appelés *standards*.

Un *standard éducatif* consiste en la définition non ambiguë du corpus de connaissances qui devraient être enseignées ou en une spécification des savoirs, des savoir-faire et des savoir-être à acquérir par les élèves, à différents moments de leur scolarité (fin d'année scolaire, fin de cycle,...).

La notion de standard éducatif peut être étendue à la qualité des différents aspects de l'entreprise éducative : formation des enseignants, manuels, locaux, équipement,...

Pour les évaluations qualitatives, les standards sont souvent exprimés sous forme d'échelles descriptives. En voici un exemple rudimentaire :

Indicateur : Qualité de l'organisation pour l'enseignement de la lecture

1. Médiocre : Tous les élèves suivent la même progression. Pas de travail en groupe.
2. Assez bien : Comme au degré 1, mais un élève très lent reçoit parfois un peu moins de travail que les autres.
3. Moyen : Constitution de deux ou trois groupes de niveau. Peu de flexibilité dans le groupement.
4. Très bien : Groupement selon les niveaux d'habileté. Flexibilité.
5. Excellent : Groupement après une étude approfondie des aptitudes et des difficultés éprouvées. Grande flexibilité.

Par ailleurs, on peut aussi définir des niveaux d'apprentissage qu'il faudrait au moins atteindre. Dans ce cas, on parle de *standards de performance* et l'épineux problème de la définition objective de la *compétence minimale* se pose.

Jusqu'à ces derniers temps, les programmes et le discours concernant l'éducation ont beaucoup plus porté sur les intentions, sur ce que l'on souhaitait voir apprendre que sur les niveaux effectifs des apprentissages à réaliser. Le pilotage ne permet plus de négliger cet aspect.

• **La détermination de la compétence minimale**

Toute épreuve finit par un constat de réussite ou d'échec. La *compétence minimale* marque la limite entre les deux. La définir objectivement n'est pas une mince affaire.

Depuis quelques années, la question de la compétence minimale suscite des débats passionnés. En effet, malgré une prolongation importante de la scolarité dans les pays industrialisés, on constate que maints élèves quittent l'école sans maîtriser des habiletés de base comme la lecture, l'expression écrite et l'arithmétique élémentaire. Or une maîtrise suffisante de ces habiletés est, plus que jamais, nécessaire tant dans le monde des entreprises que dans la vie quotidienne des citoyens. C'est pourquoi apparaissent dans de nombreux pays des dispositions réglementaires subordonnant, par exemple,

l'attribution d'un diplôme d'enseignement secondaire à la réussite d'une épreuve préliminaire établissant que l'élève maîtrise de façon suffisante les habiletés jugées essentielles.

La question brûlante est : comment fixer valablement le point ou la zone à partir desquels on peut affirmer qu'un individu est "minimalement compétent" ?

Fixation de la note de réussite

Un examen ou un ensemble d'examens sont réussis ou non. La note de succès correspond au minimum de performance jugée acceptable, autrement dit à la compétence estimée minimale. Dans une perspective de pédagogie de la maîtrise, on rencontre un problème similaire quand on se demande à partir de quel moment l'élève a atteint cet état.

Plusieurs appellations désignent ce point critique : *note de réussite*, *note de coupure*, *seuil de réussite*, *niveau d'exigence minimale*,... La fixation de cette note soulève presque toujours un problème redoutable. Comment, par exemple, déterminer à partir de quel moment quelqu'un connaît une langue étrangère ?

La question n'est simple que si l'on juge - là où c'est possible - en termes de tout ou rien. Ainsi, on peut décréter qu'un élève n'aura pas réussi, aussi longtemps qu'il commet une ou plusieurs erreurs dans la multiplication des dix premiers nombres. On est bon pilote d'avion qu'à partir du moment où l'on réussit *tous* les atterrissages.

Dès qu'une part d'erreur est jugée admissible et si l'acquisition à évaluer est complexe, la note de coupure devrait être nuancée sous forme de fourchette à l'intérieur de laquelle on réussit. On parle alors de *zone de coupure*.

La note de réussite est tantôt *absolue* et déterminée *a priori* (on fixe la note minimale avant l'épreuve, par rapport à un objectif à atteindre) ou *a posteriori* (on prend, par exemple, pour référence les résultats d'un groupe extérieur ou ceux du groupe même).

I. LES DECISIONS EMPIRIQUES

I.1 NOTE DE REUSSITE AUX EXAMENS TRADITIONNELS

L'une des règles les plus fréquentes dans les écoles belges était (et est parfois encore) d'obtenir au moins 50% dans chaque branche et 60% pour l'ensemble. Aucune justification objective de ces seuils n'était fournie, pas plus d'ailleurs que les raisons de placer ou non toutes les branches sur le même pied.

Parmi les démarches plus systématiques pour fixer la note de coupure, Ebel (1979) propose les suivantes :

a) *Pour les tests à choix multiple*

Partant du principe que, dans un test à choix multiple bien construit, un élève ignorant peut obtenir un score correspondant à des réponses données au pur hasard (50% pour des réponses juste-faux; 25% pour des réponses à quatre choix, etc.) et qu'un élève très fort devrait arriver à peu près au maximum, Ebel estime que la note de coupure peut, dans ce cas, être calculée de la façon suivante :

CHOIX	4 choix	Juste-Faux
Nombre d'items	100	100
Score attendu si on répond au hasard	25	50
Score moyen idéal	$(100 + 25)/2 = 62,50$	$(100 + 50)/2 = 75$
Score de coupure	$(25 + 62,50)/2 = 43,75$	$(50 + 75)/2 = 62,50$

Le score moyen idéal se situe donc à mi-chemin entre le score maximum possible et le score attendu si l'on répond au pur hasard.

Le score de coupure se situe à mi-chemin entre le score moyen idéal et le score attendu si l'on répond au pur hasard.

b) *Pour les examens traditionnels*

Le score de réussite est fixé arbitrairement à un pourcentage donné; ce score est ensuite ajusté en fonction des performances observées.

Exemple : on estime que 3/4 des élèves devraient fournir des réponses correctes, à condition qu'au moins 60% et qu'au plus 80% d'entre eux dépassent ce score.

Si au moins 60% des élèves arrivent à 75% de réussite, le score de réussite est fixé exactement à la moitié de la distance entre le score correspondant à 75% et le score qu'au moins 60% des élèves dépassent.

Si plus de 80% des élèves dépassent le score de réussite, ce score est alors situé à mi-chemin entre le score correspondant à 75% et le score que 80% des élèves dépassent.

Ebel précise (1979, p.341) : "Les chiffres de 60 et 80 % des sujets sont pris à titre d'exemple. Selon les circonstances, on peut souhaiter les augmenter ou les diminuer. Le but poursuivi, en fixant les valeurs, est de main-

tenir le niveau de connaissances et le pourcentage de réussite dans ce qui semble des limites raisonnables, tout en tenant aussi raisonnablement compte de l'erreur sur la mesure. Cette méthode n'est pas souvent utilisée et, pourtant, elle offre le mérite d'apporter une solution rationnelle à un problème difficile."

I. 2 NOTE DE REUSSITE AUX EPREUVES DE COMPETENCE MINIMALE

Dans la pédagogie de la maîtrise, le seuil de réussite exigé se situe, en principe, à 100%. Cette exigence est souvent adoucie, ne fût-ce que pour tenir compte des erreurs fortuites. En général, on accepte 90, voire 80%, mais pas moins.

En évaluation normative, il arrive que le seuil fixé corresponde à la performance moyenne d'un groupe de référence. Par exemple, V. de Landsheere (1988, p.124) rapporte que, dans plusieurs Etats américains, l'attribution du diplôme de fin d'études secondaires est subordonné à l'obtention, à un test de lecture, d'une note au moins égale au niveau moyen des classes de 3e année.

V. de Landsheere écrit très justement (1988, p.130) : "Fondamentalement, un seuil d'exigence unique est impossible à fixer parce que le degré de perfection voulu varie en pratique selon l'objectif poursuivi et la nature de l'apprentissage à réaliser. Rares sont les cas simples comme : on sait rouler à bicyclette lorsque, sauf accident, on n'en tombe jamais et lorsqu'on est capable de démarrer et de s'arrêter à volonté. Dans ce cas, le niveau d'exigence est pratiquement de 100%. Comment appliquer une règle similaire à <savoir jouer du piano> ? Comment justifier que, dans les bulletins scolaires, on demandait une note d'au moins 50% aussi bien en langue maternelle, en mathématiques, ... qu'en gymnastique et en dessin ? Que signifie d'ailleurs savoir à moitié dessiner ?"

Bref, aucune des démarches qui viennent d'être décrites n'est pleinement satisfaisante. Tout au plus apportent-elles un peu moins d'anarchie dans les décisions.

II. VERS PLUS D'OBJECTIVITE

Depuis les années 70, de nombreuses recherches ont porté sur les méthodes scientifiques de détermination de la note de coupure et sur le taux d'erreurs que l'on commet quand on les applique. En 1986, Berk relève plus de 35 de ces méthodes. On trouve dans l'ouvrage que V. de Landsheere (1988) a consacré à la compétence minimale, la présentation et une analyse critique des quinze méthodes qui ont le plus retenu l'attention. Phénomène qui doit beaucoup faire réfléchir les éducateurs : selon la méthode choisie, le taux d'échec varie de 9 à 78% !

Nous ne décrivons que la méthode d'Angoff (1971), parce qu'elle réunit un ensemble de qualités largement reconnues et qu'elle semble la plus sage.

1. On demande à un groupe d'éducateurs expérimentés d'indiquer, pour chaque question d'un examen ou d'un test, quel pourcentage de chances de répondre correctement a un élève minimalement compétent, dans le groupe à examiner.

10 - 20 - 30 - 40 - 50 - 60 - 70 - 80 - 90 - 100 %

2. Chacun des juges note ses appréciations sur le tableau suivant :

JUGE n° 1		
<i>Pour 100 élèves, à la limite de la compétence, combien, à votre avis, répondraient correctement aux questions posées ?</i>		
Questions	Pourcentage	Pourcentage / 100
1	80 %	0,8
2	40 %	0,4
3	50 %	0,5
4	0 %	0
5	20 %	0,2
Note minimale de réussite :		1,9

Pour obtenir la note de coupure sur l'ensemble des juges, on calcule la moyenne des notes minimales de réussite auxquelles ils ont abouti.

Dans cette démarche, la subjectivité ne cherche pas à se cacher, ce qui incite en permanence à la pondération dans les décisions.

Saunders (1984) propose un perfectionnement intéressant de cette méthode.

Après avoir fait les calculs en appliquant la méthode d'Angoff, les juges révisent leurs appréciations à trois occasions :

1. Après avoir pris connaissance des notes de coupure auxquelles les autres juges ont abouti.
2. Après avoir, en plus, pris connaissance de la distribution des scores obtenus par les élèves qui ont subi le test.

3. Après avoir étudié les statistiques descriptives (moyenne, médian, marge de variation, écart type), élaborées à partir des notes de coupure fixées au stade précédent (principalement pourcentage d'élèves qui réussiraient).

La note de réussite est finalement fixée sur la base des positions prises au stade 3, lors d'une discussion entre tous les juges.

CONCLUSION

Même si elles présentent un indéniable progrès par rapport à l'exigence traditionnelle d'au moins 50 ou 60% dans chaque branche, les démarches qui viennent d'être décrites restent incertaines.

Une préoccupation majeure doit subsister : la fonction première de l'évaluation n'est pas de classer les élèves, mais bien d'informer sur ce qu'ils ont appris effectivement et sur ce qui leur reste à faire pour être au moins minimalement compétents.

Cette notion de "compétence minimale" est, à la limite, aussi insaisissable que celle de la "note vraie", notamment parce qu'elle varie en fonction des contextes. Être minimalement compétent pour piloter un petit avion de tourisme ne qualifie pas pour prendre les commandes d'un gros avion de ligne... De surcroît, même s'il n'est pas sans utilité, un constat d'incompétence n'offre pas d'intérêt réel pour l'éducateur s'il ne s'accompagne pas d'un diagnostic clair ouvrant la voie à la remédiation.

• *L'analyse du pourcentage de réussite*

Dans une telle opération, on commence par calculer le pourcentage global de réponses correctes et l'erreur type pour chaque question et chaque groupe de questions.

On calcule ensuite, pour chaque population testée et pour chaque question, le pourcentage pondéré de réponses correctes (le plus souvent, on exclut des calculs les scores des élèves qui n'ont pas terminé le test).

Pour chaque pourcentage de réponses correctes, on calcule **l'erreur type sur la mesure**.

Cette erreur est due au fait que, quelle que soit la qualité des instruments, on ne mesure jamais de façon parfaite. Ceci est particulièrement évident dans les sciences de l'homme. Par exemple, quand, au terme d'un examen, on attribue une note relative à la connaissance d'une langue étrangère ou au degré de maîtrise du piano, on peut être certain que l'on n'a pas attribué **la note vraie**, celle qui correspondrait parfaitement à tout ce que l'élève sait et peut faire. Des techniques statistiques permettent de déterminer, avec

plus ou moins de probabilité, un espace au sein duquel la *note vraie* doit se trouver.

Ces techniques partent de l'hypothèse que les erreurs de mesure se distribuent au hasard autour de la *note vraie*. Si cette distribution est *normale* (courbe de Gauss), on estime que deux fois l'erreur type en plus ou en moins de la valeur observée détermine une zone à l'intérieur de laquelle il y a 95 chances sur 100 que la *note vraie* se situe.

Parmi les techniques sophistiquées utilisées pour estimer cette erreur, on recourt de plus en plus à la technique complexe du *jackknife*.

La technique du *jackknife*, introduite par Quenouille (1956) et reprise par Tukey (1958) a pour but de supprimer ou, au moins, de réduire le biais d'un estimateur. Elle consiste à construire, à partir d'un échantillon d'effectif n , n sous-échantillons d'effectifs $(n - 1)$, obtenus chacun par l'enlèvement d'une des observations à la fois, et à calculer, sur chacun d'eux, une pseudo-valeur. L'estimateur de Quenouille-Tukey est la moyenne de ces pseudo-valeurs.

Cette technique est surtout utile quand on a affaire à des statistiques biaisées (par exemple, le coefficient de corrélation), dont le biais est difficile à calculer. Le terme *jackknife* fait allusion au couteau suisse à usages multiples, que l'on utilise pour se tirer d'affaire un peu dans toutes les occasions.

Dans les exemples donnés des travaux de la NAEP, l'erreur type sur la mesure est calculée selon cette technique.

En général, on désire aussi connaître les pourcentages de réussite et les erreurs types, par population, pour chaque groupe d'items correspondant à des aspects particuliers de la discipline (aspects définis par les experts des branches).

• **La probabilité de réponse correcte**

Le modèle de Rasch exprime la probabilité de réponse correcte à une question comme une fonction de la difficulté de la question et de l'aptitude du sujet.

La théorie complexe de G. Rasch (1960) part de l'interrogation suivante : "A quelles conditions pourrait-on comparer des performances d'un même sujet, observées à différents tests, ou encore comparer les performances de différents groupes d'étudiants à un même test ?". La question se pose, par exemple aussi, lorsqu'un test est traduit en différentes langues pour réaliser une enquête internationale (G.N. Masters, 1993).

G. Rasch conteste qu'un test bien construit, administré à diverses populations, dans les mêmes conditions, permet nécessairement d'obtenir des scores comparables. La comparaison est douteuse si le test ne représente pas

la même tâche pour tous. C'est, par exemple, le cas s'ils n'ont pas suivi le même programme d'enseignement.

G.N. Masters (1993) synthétise la problématique dans les termes suivants :

"Pour qu'un test puisse être considéré comme un véritable instrument de mesure, il faut que les questions :

- portent sur la même dimension d'une performance;
- permettent de construire une échelle de mesure avec une unité bien définie;
- puissent être <calibrées> sur cette échelle, en sorte que différents choix de questions soient directement comparables;
- constituent une tâche comparable pour les différents élèves qui y répondent."

Selon le modèle proposé par Rasch, la probabilité qu'a un étudiant n de répondre correctement à une question i dépend du degré de maîtrise que l'élève a de la matière et de la difficulté intrinsèque de la question. La formule de Rasch permet de calculer si un étudiant n , répondant une question dont la réponse est notée *juste* ou *faux*, a une probabilité de 1 ou de 0 de répondre correctement.

La possibilité de comparer directement des performances accomplies à différents tests est particulièrement intéressante pour les mesures de rendement destinées au pilotage. Par exemple, il n'est plus nécessaire que tous les élèves, fréquentant divers établissements, répondent aux mêmes questions, ne fût-ce que parce qu'ils ont suivi des programmes différents.

Le prix à payer pour cette commodité est l'exercice d'un contrôle permanent de la conformité des réponses des élèves au modèle.

Quelques précisions sur la théorie des traits latents

Le modèle de construction d'échelles de rendement utilisé par Rasch suppose qu'une performance a pour assise un trait latent qui est commun à la nature de la tâche à effectuer et à une modalité de la connaissance ou de l'habileté de l'individu qui l'effectue. Les tâches ou items utilisés estiment la performance des individus dépendant de ce trait et permettent de le situer sur un continuum.

Cette approche n'exige pas que l'on dispose nécessairement d'un univers hypothétique d'items dont on peut tirer des échantillons. Plus précisément, elle postule que la réponse d'un élève à un item donné est fonction de l'erreur aléatoire et de la position occupée par l'élève sur le continuum du trait latent.

Pour un ensemble d'items, il ne doit exister d'intercorrélation ni entre les erreurs aléatoires dont chaque item est affecté, ni entre l'erreur et le trait latent. Cette assumption implique que le trait soit *unidimen-*

sionnel. Si les items choisis sont bien corrélés avec un objectif spécifique d'apprentissage ou avec une aptitude particulière, un modèle comme celui de Rasch fonctionne bien.

I.I. Bejar (cité par J. Keeves, 1992, p.116) souligne avec raison que l'unidimensionnalité n'implique pas que la performance réalisée à propos d'un item soit due à un seul processus psychologique : "En réalité, divers processus psychologiques interviennent dans l'élaboration des réponses à un ensemble d'items. Mais, pourvu que les processus fonctionnent à l'unisson, - c'est-à-dire, si la réponse produite à chaque item est liée de la même façon à un même processus -, l'hypothèse de l'unidimensionnalité tient. Elle n'est pas respectée si, par exemple, certains items d'un test de connaissances exigent des calculs et si d'autres ne demandent que la simple restitution de faits mémorisés. Si, dans la population testée, le niveau d'aptitude numérique varie, le test ne sera plus unidimensionnel".

Le modèle de Rasch utilise la fonction logistique pour établir la relation entre la position des items sur l'échelle du trait latent et la probabilité de réponse correcte à l'item.

J. Izard (1992, p.27) souligne la multiplicité des cas d'utilisation du modèle :

- Etalonnage de tests de connaissances.
- Construction de banques d'items.
- Regroupement de tests possédant les mêmes propriétés.
- Classement des sujets, même s'ils n'ont pas répondu aux mêmes questions.
- Détection de questions particulièrement sensibles à la forme d'enseignement suivi.
- Construction de tests qui posent des questions en fonction des réponses données aux items précédents (tests dits adaptatifs ou sur mesure).
- Identification d'élèves qui présentent une configuration inhabituelle de compétences.
- Identification des diverses stratégies utilisées pour la résolution de problèmes (de telles variations peuvent être liées à des biais culturels).
- Repérage des similarités entre items, ce qui aide à décrire l'habileté dont l'élève fait montre à un niveau donné.
- Analyse des réponses aux questions ouvertes, afin de voir si les notes auxquelles elles conduisent sont cohérentes avec les résultats des QCM.
- Parallélisation des tests.

Le modèle de Rasch constitue l'une des meilleures illustrations de la *théorie des réponses aux items (IRT)* dont il va être question à propos de l'ancrage des échelles.

C'est ce modèle qui est à la base des échelles que la NAEP utilise. Elles sont très parlantes, même pour un public peu averti; en outre, elles permettent de comparer facilement les performances de différentes populations à un même moment ou d'observer comment une même population évolue à mesure que le temps passe.

• *L'ancrage des échelles*

Les indicateurs étant aussi destinés à informer un large public sur les acquis des élèves et à engager un débat à ce propos, il importe de présenter les résultats de façon facile à comprendre et d'indiquer clairement quels niveaux de performance peuvent être considérés comme satisfaisants. L'un des avantages de la *théorie des réponses aux items (Item Response Theory - IRT)* est de permettre une interprétation critériée (c'est-à-dire, de préciser, par rapport à un objectif défini, quel degré de maîtrise est atteint).

"La théorie des réponses aux items est l'une des façons d'analyser les items d'un échantillon matriciel. En général, on fait l'hypothèse qu'un seul paramètre p caractérise l'aptitude d'un individu à répondre aux items figurant dans un échantillon matriciel; on suppose, par ailleurs, que chaque item possède un ou plusieurs paramètres qui décrivent la relation entre les p de tous les individus et les réponses aux items observées. On fait ainsi l'hypothèse que la probabilité qu'a un individu de répondre correctement à un item est fonction de certaines propriétés de l'item, de l'aptitude de l'individu et de la relation entre l'aptitude et l'item. On estime généralement que cette relation stochastique n'est pas linéaire. Etant donné un modèle de la relation stochastique, un programme d'ordinateur estime les paramètres individuels p et le ou les paramètres associés à chaque item" (A.E. Beaton).

Dans le système adopté par la NAEP des Etats-Unis, les résultats peuvent se distribuer entre 1 et 1000. On fait cependant l'hypothèse que 99% des scores se situeront entre 200 et 500. Les cinq niveaux choisis pour présenter les résultats, dans tous les rapports publiés, sont 300, 400, 500, 600 et 700 (échelle à cinq degrés). On a vu, au Chapitre III, des exemples simplifiés de l'application de ce type d'étalonnage.

Chaque niveau est décrit par les tâches que la majorité des élèves qui l'atteignent sont capables d'effectuer correctement; des questions types servent d'illustration.

Les résultats publiés indiquent, pour chaque niveau et pour chaque population, le pourcentage des élèves qui l'atteignent tout juste ou le dépassent.

Pour ancrer les échelles, on a repéré, lors d'enquêtes préalables, les questions qui discriminent bien entre les différents niveaux de complexité. Une question a été tenue pour repère, à un niveau donné, si les élèves qui

l'atteignent ont entre 65 et 80 chances sur 100 d'y répondre correctement. La probabilité de réponse correcte pour les élèves classés au niveau immédiatement inférieur est, par exemple, située entre 20 et 50. Les questions n'ont été finalement retenues qu'après accord des experts (ils ont dû préciser quelles habiletés prérequisées et quelles habiletés de transfert les élèves devaient posséder pour répondre correctement).

Exemples d'échelles de compétence

Les responsables de l'*Evaluation nationale des progrès de l'éducation* des Etats-Unis (NAEP) ont adopté notamment les deux échelles suivantes, - succinctement évoquées au chapitre III-, pour les populations de 9, 13 et 17 ans (*Nation's Report Cards*, 1988). Dans les publications de la NAEP, chacun des échelons est illustré par des exemples de questions d'examen.

-- Sciences

Niveau 1 : Connaissance de faits scientifiques de la vie quotidienne

Les élèves ont acquis, dans la vie quotidienne, quelques connaissances scientifiques rudimentaires à propos des animaux et des plantes. Ils sont capables de trouver la façon de se servir d'appareils familiers fonctionnant selon des principes mécaniques simples.

Niveau 2 : Compréhension de principes scientifiques simples

Les connaissances relatives aux sciences de la vie, particulièrement à propos du système biologique humain, ont augmenté.

Capacité d'appliquer quelques principes simples de la physique.

Les élèves commencent à comprendre le raisonnement scientifique, notamment en matière de classification et d'interprétation.

Niveau 3 : Application d'informations scientifiques de base

Interprétation de données figurant dans des tableaux simples ou dans des graphiques; dégagement de conclusions à partir de résultats d'expériences scientifiques. Les élèves font preuve de connaissance et de compréhension dans le domaine des sciences de la vie, notamment de familiarité avec certains comportements animaux et avec certaines relations écologiques. Connaissance de certaines informations de base en sciences physiques et en sciences de la terre et de l'espace.

Niveau 4 : Analyse de procédures et de données scientifiques

Connaissances scientifiques plus approfondies et augmentation nette de la capacité d'interpréter des textes et des graphiques en fonction des connaissances acquises. Progrès dans la compréhension des principes régissant les sciences physiques.

Les élèves sont maintenant capables de concevoir des expériences et d'en interpréter les résultats. Ils savent distinguer des relations et tirer des conclusions en appliquant la connaissance de faits et de principes, principalement dans le domaine des sciences physiques.

Niveau 5 : Capacité d'intégrer des informations scientifiques spécialisées

Capacité de dégager des relations et de tirer des conclusions par inférences, en utilisant des connaissances scientifiques détaillées dans le domaine des sciences physiques, particulièrement en chimie. L'élève prend conscience des limites des extrapolations.

Capacité d'appliquer certains principes de génétique et d'interpréter les implications sociétales de la recherche en ce domaine.

— *Mathématiques*

Niveau I : Connaissance factuelle de simples aspects de l'arithmétique

A ce niveau, les élèves connaissent quelques notions de base relatives à l'addition et à la soustraction, et la plupart des élèves sont capables d'additionner des nombres de deux chiffres sans reports. Les élèves reconnaissent des situations simples appelant des additions ou des soustractions. Ils possèdent une habileté de classement rudimentaire

Niveau II : Apparition de l'habileté et de la compréhension

Les élèves font montre d'une gamme plus large d'habiletés et d'une compréhension plus profonde qu'au niveau I. Toutefois, l'utilisation qu'ils font de ces habiletés reste imparfaite et peu flexible. On peut penser qu'à ce niveau, les élèves éprouvent des difficultés à faire des raisonnements qui exigent plus que de simples calculs. Ils ne savent résoudre que des problèmes simples n'exigeant qu'une seule opération. Ils comprennent le concept de pourcentage.

Niveau III : Maîtrise des opérations de base et début de la résolution de problèmes

Les élèves possèdent maintenant une compréhension superficielle des quatre opérations et commencent à acquérir l'habileté de résolution de problèmes. Ils ont acquis la notion de nombres pairs ou impairs, et une compréhension élémentaire de la notion de rapport.

Les élèves sont capables d'utiliser les nombres décimaux, de convertir des fractions et des pourcentages, et de calculer les moyennes.

Résolution de problèmes en plusieurs étapes. Capacité de mesurer des longueurs, d'identifier des formes géométriques, de calculer l'aire du rectangle. Utilisation d'informations se trouvant dans des graphiques ou des tableaux.

Niveau IV : Démarches et raisonnements de complexité modérée

Le raisonnement numérique devient plus sophistiqué. L'élève recourt à un éventail plus large d'habiletés mathématiques, y compris l'algèbre et la géométrie.

Multiplication de fractions et de nombres décimaux. Utilisation d'une gamme de démarches permettant de résoudre des problèmes assez complexes.

Compréhension accrue de la mesure et des concepts géométriques : mesures d'angles dans des figures simples, calcul du périmètre, de la surface et du volume.

Niveau 5 : Résolution de problèmes en plusieurs étapes; algèbre.

Connaissance des propriétés de la moyenne arithmétique. Résolution de problèmes en utilisant des données figurant dans des tableaux complexes.

L'élève est de plus en plus capable d'utiliser des notions de mathématiques apprises à l'école pour résoudre des problèmes de la vie courante.

Critique de la théorie des réponses aux items (Goldstein et Wood, 1989)

La critique la plus fondamentale porte sur la difficulté d'établir l'unidimensionnalité des échelles.

Pour établir une échelle de réponses aux items, on suppose qu'ils portent tous sur un seul et même aspect ou *trait* (par exemple, l'habileté en lecture) et qu'un élève se caractérise par une seule valeur de ce trait. S'il en est ainsi, on estime que l'échelle est *unidimensionnelle*. Il reste à prouver que c'est vrai, et c'est loin d'être aisé.

Pour chacun des items caractérisant un niveau de l'échelle, on fait l'hypothèse que toute amélioration du score, à mesure que l'élève avance dans ses études, reflète bien un changement chez celui-ci. Si les scores à un même item ne s'élèvent pas, à mesure que l'on avance dans la scolarité, l'item est considéré comme anormal et est éliminé. Par contre, si d'une année scolaire à l'autre, les résultats s'améliorent, l'item est conservé. Il reste à établir si ce progrès est dû à une amélioration effective de la connaissance et non, par exemple, à une plus grande familiarité avec le type d'épreuve employé.

Par ailleurs, il n'est pas certain que le "trait" que l'on a cru pouvoir déceler dans une population est nécessairement le même dans une autre, appartenant à une autre culture ou fréquentant un système d'éducation différent. La seule traduction dans des langues différentes est déjà source de variations de la difficulté des questions. Et si, pour ces raisons, on est amené, dans des évaluations comparatives internationales, à supprimer des ensembles d'items différant selon les pays, la comparaison peut devenir douteuse.

Par ailleurs, le problème de la valeur ajoutée se pose ici aussi. Pour estimer dans quelle mesure l'enseignement a contribué au progrès des apprentissages, il faudrait mesurer l'état initial des acquis des élèves. Dans cette perspective, l'approche longitudinale est évidemment souhaitable.

• *Le fonctionnement différentiel des items*

On considère qu'une question fonctionne différemment, si des élèves d'aptitudes égales, mais appartenant à différentes populations, ont une probabilité différente de répondre correctement. Cette différence peut, par exemple, résulter de problèmes linguistiques. Des techniques statistiques complexes sont utilisées pour déterminer ce fonctionnement différentiel.

□ Conclusion

Les considérations qui précèdent sont loin de couvrir l'ensemble des problèmes complexes que posent les évaluations destinées à donner substance à des indicateurs. Elles permettent, croyons-nous, d'entrevoir le degré de technicité de bien des démarches. Celles-ci relèvent souvent de la haute spécialisation, ce qui explique que, dans tous les pays où des systèmes d'indicateurs de l'éducation ont été mis en place, les centres de recherche universitaires jouent un rôle important.

Principales abrégés

APU	Assessment of Performance Unit (Unité d'évaluation des performances scolaires) (Royaume-Uni)
CERI	Centre for Educational Research and Innovation (Centre pour la Recherche et l'Innovation dans l'Enseignement) (OECD)
CITE	Classification internationale type de l'enseignement
DEP	Direction de l'Evaluation et de la Prospective, Ministère de l' Education (France)
ETS	Educational Testing Service (Princeton)
IAEP	International Assessment of Educational Progress (Evaluation internationale du progrès de l'éducation) (Etats-Unis)
IEA	International Association for the Evaluation of School Achievement (Association internationale pour l'évaluation du rendement scolaire)
INES	Indicateurs internationaux de l'enseignement (Projet de l'OCDE)
IRT	Item response theory
k	Facteur "Connaissance" (<i>knowledge</i>)
ISCED	International Standard Classification of Education - Français : CITE
ISCO	International Standard Classification of Occupations
NAEP	National Assessment of Educational Progress (Evaluation nationale du progrès de l'éducation) (Etats-Unis)
NOBS	Non-curriculum bounded objectives (Objectifs indépendants du curriculum)
OCDE	Organisation pour la Coopération et le Développement Economiques
OTL	Opportunity to learn. Occasion d'apprendre : indice indiquant dans quelle mesure les élèves ont eu la possibilité d'apprendre les savoirs, les habiletés ou les savoir-être sur lesquels les évaluations portent.
PIB	Produit intérieur brut
SEAC	School Examinations and Assessment Council (Conseil pour l'évaluation scolaire) (Royaume-Uni)
SSE	School Self-Evaluation (auto-évaluation des écoles).

Bibliogra

- ALLISON, G., *The Essence of Decision*, Boston, Little, Brown and Co, 1971.
- APPLEBEE, A.N., LANGER, J.A. et MULLIS, V.S., *The writing report card : Writing achievement in American schools*, Princeton, ETS (NAEP), 1986.
- APPLEBEE, A.N. et al., *Who reads best ? Factors related to reading achievement in Grades 3, 7 and 11*, Princeton, ETS (NAEP), 1988.
- ASSOCIATION FRANÇAISE DES ADMINISTRATEURS DE L'ÉDUCATION, *Le pilotage du système éducatif, Actes du XV^e colloque, Revue de l'AFAE*, 1993, 3.
- ATKINSON, R.C. et JACKSON, G.B., Eds, *Research and Education Reform*, Washington, D.C., National Academy Press, 1992.
- AUBRET, J., *Reconnaissance et validation des acquis : le portefeuille de compétences*, Paris, Université de Paris VIII, 1992.
- BAUER, R., Ed., *Social Indicators*, Cambridge, Mass., MIT Press, 1966.
- BEATON, A.E., *Implementing the new design : The NAEP 1983-84 Technical Report*, Princeton, ETS, 1987, 768 p.
- BEJAR, I.I, *Achievement testing - Recent advances*, Beverly Hills, Calif., Sage Publications, 1983.
- BERK, R.A., A consumer's guide to setting performance standards on criterion-referenced tests, *Review of Educational Research*, 1986, 56, 1, 137-172.
- BERLEUR, J., *Technology Assessment*, Namur, Faculté Universitaire N.-D. de la Paix, 1991 (syllabus de séminaire).
- BIJKER, W.E. et LAW, J., *Shaping technology / Building society*, Cambridge, Mass., MIT Press, 1992.
- BOTTANI, N. et WALBERG, H., Eds., *The OECD International Education Indicators. A framework for analysis*, Paris, OCDE, 1992. (Edition française : *L'OCDE et les indicateurs internationaux de l'enseignement. Un cadre d'analyse*).
- BURSTEIN, L., Ed., *The IEA Study of Mathematics III*, Oxford, Pergamon, 1993.
- CARR-HILL, R. et MAGNUSSEN, O., *Les indicateurs de résultats des systèmes d'enseignement*, Paris, OCDE, 1973.
- CARROLL, J.B., *The teaching of French as a foreign language in eight countries*, Stockholm, Almqvist and Wiksel, 1975.
- CLIFT, P.S., NUTTALL, D.L. et McCORMICK, R., Eds, *Studies in school self-evaluation*, Lewes, Palmer Press, 1987.

- COLLIN, R.D., *Les technologies de l'intelligence*, Genève, Neurope Lab, 1993.
- COLLINGRIDGE, D., *The social control of technology*, Londres, Francis Pinter, 1980.
- COLOMB, J., *L'évaluation des élèves et l'examen dans l'enseignement secondaire*, Rapport, Strasbourg, Conseil de l'Europe, DECS/Rech (1993) 22.
- COMBER, L.C. et KEEVES, J.P., *Science education in nineteen countries*, Stockholm, Almqvist and Wiksel, 1973.
- COMMISSION DES COMMUNAUTÉS EUROPÉENNES, *The social and Economic Implications of New Technology. First Biennial Report, Society and Technology towards the Future*, FAST-DGXII, 1991, 2, 101-117 (FAST Document 10).
- CONSEIL DES MINISTRES DE L'ÉDUCATION DU CANADA, *Mise en application des indicateurs de l'enseignement : L'expérience du Canada*, Communication à l'assemblée générale du projet INES de l'OCDE, Semmering, 1989.
- DE GROOF, J. et al. *Indicatoren voor het onderwijsbeleid. Een stand van zaken van het OESO-Project "INES"*, *Tijdschrift voor het onderwijsbeleid*, 1991-1992, 3, 142-156.
- DE GROOF, J. et al., *De school op rapport. Het Vlaams onderwijs in internationale context*, Kapellen, Pelckmans, 1993.
- de LANDSHEERE, G., *Comparaison du rendement de l'enseignement des mathématiques dans douze pays*, *Education* (Bruxelles), 1967, 106, 61-70.
- de LANDSHEERE, G., *Evaluation continue et examens. Précis de docimologie*, Paris, Nathan; Bruxelles, Labor, 1992, 7e éd.
- de LANDSHEERE, G., *Dictionnaire de l'évaluation et de la recherche en éducation*, Paris, PUF, 1992, 2e éd.
- de LANDSHEERE, G. et POSTLETHWAITE, T.N., *Rendement de l'enseignement des mathématiques dans douze pays*, Paris, Institut Pédagogique National, 1969.
- de LANDSHEERE, V., *Faire réussir - Faire échouer, La compétence minimale et son évaluation*, Paris, PUF, 1988.
- de LANDSHEERE, V., *L'éducation et la formation. Science et pratique*, Paris, PUF, 1992.
- DELORS, J., *Les indicateurs sociaux*, Paris, Futuribles, 1971.
- DEPARTMENT OF EDUCATION OF NEW ZEALAND, *Assessment for better learning*, Wellington, 1989.
- DIETEL, R., *What works in performance assessment ? Evaluation comment*, printemps 1993.
- DORON, R. et PAROT, F., Eds, *Dictionnaire de psychologie*, Paris, PUF, 1991.
- DOSSEY, J. et al., *Mathematics : Are we measuring up ? The mathematics report card, Executive summary*, Princeton, ETS (NAEP), 1988.
- DUPONT, P., DUPUIS, M. et OSSANDON, M., *A livre ouvert... dans la gestion de l'école*, Bruxelles, Ministère de l'Éducation, 1988.
- EBEL, R.L., *Essentials of educational measurement*, Englewood Cliffs, Prentice Hall, 1979, 3e éd.

- ELLY, W.B., *How in the world do students read ?* La Haye, The IEA Secretariat, 1992.
- FITZ-GIBBON, C., Performance Indicators and Examination Results, *Interchange*, 1992, 11 (Edinburgh).
- FLANAGAN, J.C. et al., *The American High-School Student : The Identification, Development and Utilization of Human Talents*, Pittsburgh, Pittsburgh University Press, 1964.
- FLANAGAN, J.C., et al., *Project Talent : Five Years after High School and Appendix II*, Pittsburgh, American Institute for Research, 1971.
- GARDEN, R.A. and ROBITAILLE, D.F., Eds, *The IEA study of mathematics II*, Oxford, Pergamon, 1989.
- GARIN, C., Les régions dans la compétition scolaire, *Le Monde*, 14 janvier 1993.
- GIPPS, C., *National Assessment and School Evaluation in the U.K.*, Communication à la Conférence annuelle de l'AERA, San Francisco, 1989.
- GLASER, R., Instructional technology and the measurement of learning outcomes. Some questions, *American Psychologist*, 1963, 18, 519-521.
- GOLDSTEIN, H. et WOOD, R., Five decades of item response modelling, *Brit. J. Math. Stat. Psychol.*, 1989, 42, 139-167.
- GONOD, F., Problématique de la maîtrise sociale de la technologie, *Analyse des systèmes*, 1990, 16, 3, 3-43.
- GORMAN, T.P., PURVES, A., DEGENHART, R.E., *The IEA study of written composition, Vol.I : The international writing tasks and scoring scales*, Oxford, Pergamon, 1988.
- GRISAY, A., *Rendement de l'enseignement de la langue maternelle en Belgique francophone*, Bruxelles, MENC, Organisation des Etudes, Coll. "Recherche en éducation", 1974.
- GRISAY, A., L'enseignement belge forme-t-il trop peu de bons élèves ?, *Revue de la Direction générale de l'Organisation des Etudes*, 1974, 9, 3-12.
- GRISAY, A., La sélection peut-elle aider à combattre l'ennui scolaire ? Analyse de quelques résultats de la recherche de l'IEA sur le rendement scolaire en Belgique francophone, *Scientia Paedagogica Experimentalis*, 1976, 13, 2, 241-261.
- GRISAY, A., *Du mythe de la "bonne école" à la réalité (fuyante) de "l'école efficace"*, Liège, Service de pédagogie expérimentale de l'Université de Liège, 1988.
- GRISAY, A. et HENRY, G., La recherche IEA en sciences et en langue maternelle, *Revue de la Direction générale de l'Organisation des Etudes* (Bruxelles), 1973, 8, 35-39.
- GUTHRIE, J.W., *The evolving political economy of education and the implication for educational evaluation*, (texte d'une conférence), Leicester University, 1989.
- HABY, R., in't VELD, R.J. et TSCHOUMY, J.A., *Examens des politiques nationales d'éducation - Belgique*, Paris, OCDE, 1993.
- HARGREAVES, D.H., Assessment and Performance Indicators : the English Experience. In T. WYATT et A. RUBY, Eds, 1988, 125-134.
- HENRY, G., Le rendement de l'enseignement des mathématiques dans douze pays, *Information. Bulletin de l'Organisation des Etudes* (Bruxelles), 1970, 1, 3-16.

- HENRY, G., Un aspect de l'inégalité socio-culturelle entre sexes. Etude différentielle des résultats de l'IEA en sciences, *Revue de la Direction générale de l'Organisation des Etudes*, 1974, 3, 3-9.
- HENRY, G., Une étude du rendement de l'enseignement des sciences dans dix-neuf pays, *Actes du congrès des professeurs de sciences* (Bruxelles), 1974, 31, 1-31.
- HENRY, G., *Rendement de l'enseignement des sciences en Belgique francophone*, Bruxelles, MENC, Coll. "Recherche en éducation" 8, 1975.
- HENRY, G., Une étude des effets de la sélection au niveau de l'enseignement secondaire supérieur, *Revue de la Direction générale de l'Organisation des Etudes*, 1975, 1, 3-13.
- HERRIOTT, R.E., Some Observations on the Condition of Education, *Social Indicators Research*, 1979, 6, 181-194.
- HUTMACHER, W., *Quand la réalité résiste à la lutte contre l'échec scolaire*, Genève, Service de recherche sociologique, Cahier 36, 1993.
- IEA, *Science achievement in 17 countries*, Oxford, Pergamon, 1988.
- IZARD, J., *Assessing learning achievement*, Paris, UNESCO, Coll. Educational Studies and Documents 60, 1992.
- JOHNSON, E.G., The design of the National Assessment of Educational Progress, *Journal of Educational Measurement*, 1992, 29, 95-100.
- KAAGAN, S. et SMITH, M.S., Indicators of educational quality, *Educational Leadership*, Octobre 1985, 21-24.
- KEEVES, J., Scaling achievement test scores. In J. KEEVES, Ed., *Methodology and measurement in international educational surveys*, La Haye, IEA, 1992.
- KEEVES, J., *Learning science in a changing world. Cross-national studies of science achievement : 1970-1984*, La Haye, IEA, 1992.
- KEEVES, J. and ROSIER, M., *The IEA study of science I*, Oxford, Pergamon, 1993.
- KOGAN, M., Monitoring, Control and Governance of School Systems, in P. LADERIERE, Ed., *Governance and Accountability in Education*, Paris, OECD, 1993.
- KONTTINEN, R., *Research into the role of final examinations in secondary education*, Strasbourg, Conseil de l'Europe, DECS / Rech. 1993.
- KUHN, T.S., *The structure of scientific revolution*, Chicago, University Press, 1970, 2e éd.
- LAPOINTE, A.E., MEAD, N.A. et PHILLIPS, G.W., *A World of Differences. An International Assessment of Mathematics and Science*, Princeton, ETS, 1989.
- LECLERCQ, D., Computerized tailored testing, *European Journal of Education*, 1980, 5, 3.
- LE GUEN, M., *Measuring to succeed. National evaluation in France*, Paris, OCDE-CERI/INES, 1991, 9.
- LEWIS, E.G. and MASSAD, C.E., *The teaching of English as a foreign language in ten countries*, Stockholm, Almquist and Wiksel, 1975.
- LIENSOL, B. et MEURET, D., Les performances des lycées pour la préparation au baccalauréat, *Education et Formations*, 1987, 11, avril-juin, 25-36.

- LINN, R.L., BAKER, E.L. et DUBAR, S.B., Complex, performance-based assessment : expectations and validation criteria, *Evaluation Comment*, Hiver 1991-1992, 3-9.
- MADAUS, G. F., The influence of testing on the curriculum. In L. TANNER, Ed., *Critical issues in curriculum*, Chicago, University of Chicago Press, 1988.
- MADAUS, G. F. et KELLAGHAN, T., Curriculum evaluation and assessment. In P., W., JACKSON, Ed., *Handbook of research on curriculum*, New York, MacMillan, 1992, 119-154.
- MAGY, J., *L'enseignement et la formation en Communauté française. Produire et gérer la qualité*, Bruxelles, CEPESS, Coll. Monographies, 1992.
- MANTEL, N et HAENSZEL, W., Statistical aspects of the retrospective study of disease, *Journal of the National Cancer Institute*, 1959, 22,719-748.
- MARKLUND, S., Assessment of School Achievement in Compulsory and Post-Compulsory Schools in Sweden. In P. WESTON, Ed., *Assessment of pupil achievement : motivation and school success*, Strasbourg, Conseil de l'Europe, 1991, 31-44.
- MASSOZ, D. et HENRY, G., *Connaissances et attitudes socio-politiques d'élèves de l'enseignement secondaire*, Recherche de l'IEA, Bruxelles, MENC, Direction générale l'Organisation des Etudes, 1980.
- MASTERS, G.N., Rasch measurement and IEA studies, *IEA Bulletin*, 1993, 2, 2.
- MATALON, B., *L'analyse hiérarchique*, Paris, Gauthier-Villars, 1975.
- McKNIGHT, C.C. et al., *The underachieving curriculum - Assessing U.S. School Mathematics from an International Perspective*, Champaign, Ill., Stupp Publishing, 1987.
- MENESE, J., L'analyse modulaire des systèmes, *Hommes et techniques*, 1972.
- MINISTERE DE L'EDUCATION NATIONALE, DIRECTION DE L'EVALUATION ET DE LA PROSPECTIVE, *L'état de l'école*, Paris, DEP, 1992, 2.
- MISLEVY, R. J., Foundations of a new test theory. In N. FREDENKSEN et I. BEJAR (Eds.), *Test theory for a new generation of tests*, Lawrence Erlbaum, 1990.
- MULLIS, I.V.S. et JENKINS, L.B., *The science report card. Trends and achievement based on the 1986 National Assessment*, NAEP, Princeton, 1988.
- MULLIS, I.V.S. et al., *The science report card 1986*, Princeton, ETS-NAEP, 1988.
- NATIONAL CURRICULUM COUNCIL, *National Curriculum Council*, York, 1991.
- NATIONAL EDUCATION GOALS PANEL, *The National Education Goals Report*, Washington, 1992.
- NEAVE, G., *Education and the European Community*, Paris, Institut européen d'éducation et de politique sociale, 1983.,
- NGUYEN, N. T, LOBET-MARIS, C., BERLEUR, J. et KUSTERS, B., Methodological issues in information technology assessment, *The International Journal of the Management of Technology*, à paraître, 1993.
- NUTTALL, D., Choosing indicators, In OCDE-CERI, 1991.

- NUTTALL, D., School evaluation in the UK, in *Université Libre de Bruxelles, L'enseignement en Europe. L'enseignement en Belgique*, Bruxelles, ULB, 1993, 43-47.
- OAKES, J., *Educational indicators : a guide for policy makers*, New Brunswick, N.J., Center for policy research in education, Rutgers University, 1986.
- OCDE, *Regards sur l'éducation. Les indicateurs de l'OCDE*, Paris, OCDE, 1992.
- OCDE, *Une éducation et une formation de qualité pour tous*, Paris, OCDE, 1992.
- OCDE, *The OECD International Education Indicators : A framework for analysis*, Paris, OCDE, 1992.
- OCDE-CERI, *Manuel sur les indicateurs de l'enseignement*, Paris, OCDE, 1991 (Doc. CERI/INES (91) 17).
- OCDE-CERI, *Curriculum reform : Policies, Practices and Plans for the Future*, Paris, OCDE, 1993 (Doc. CERI/ED (93)12).
- OLMSTED, P. and WEIKART, D., *Profiles of child care and education in fourteen countries*, Ypsilanti, High/Scope Press, 1989.
- PALMER, R.F., *National Testing in Canada*, Strasbourg, Conseil de l'Europe, Rech. 1993,39.
- PELGRUM, H. and PLOMP, T., *The use of computers in education world wide*, Oxford, Pergamon, 1991.
- PETRELLA, R., Des politiques malades de leur culture. L'homme et l'outil : les valeurs qui basculent, *Le Monde diplomatique*, 20 septembre 1987.
- PIJL, S.J., Practices in Monitoring Student Progress, *Internationale Zeitschrift für Erziehungswissenschaft*, 1992, 38, 117-131.
- PORTER, A., Indicators : objective data or political tools?, *Phi, Delta, Kappan*, 1988, 3, 503-508.
- POSTLETHWAITE, T.N., *The Monitoring of Education and the Responsibility of Governments*, Bruxelles, International Academy of Education, 1993 (document provisoire).
- POSTLETHWAITE, T. N. and WILEY, D., Eds., *The IEA study in science II : Science achievement in 23 countries*, Oxford, Pergamon, 1993.
- PURVES, A., *Literature education in 10 countries*, Sockholm, Almquist and Wiksel, 1973.
- RASCH, G., *Probabilistic models for some intelligence and attainment tests*, Copenhagen, Danish Institute for Educational Research, 1960.
- RICHARDS, C.E., Indicators and three types of educational monitoring systems. *Phi Delta Kappan*, 1988, 3, 495-498.
- ROGERS, G. *Keeping the school under review. A method of self assessment for schools devised by the ILEA inspectorate*, Londres, ILEA, 1977.
- ROSS, K.N. et POSLETHWAITE, T.N., *Indicators of the quality of education: a summary of a national study of primary schools in Zimbabwe*, Paris, International Institute of Educational Planning (UNESCO), 1993.
- RUBY, A., The International Indicator Project, In T.J. WYATT et A. RUBY, Eds, *Indicators in Education*, Sidney, Australian Conference of Directors-General of Education, 1988.

- RUBY, A., Indicators, reporting and rationality : understanding the phenomena, in OCDE/CERI, *International educational indicators project. A reader on conceptual issues*, Paris, OCDE, 1991 (document de travail).
- ODDEN, A., Making sense of education indicators : the missing ingredient. In T. WYATT et A. RUBY, Eds, 1990, 33-50.
- RUBY, A. et WYATT, T.J., *Using indicators for school improvement*, Sydney, Australian Conference of Directors-General of Education, 1988.
- SCHEERENS, J., *Process Indicators of School Functioning*, Paris, OCDE-CERI, 1989.
- SCHEERENS, J., Caractéristiques des établissements d'enseignement. Programme à moyen terme, 1992-1994. Note de travail reproduite dans le *Rapport annuel* du Ministère de l'éducation, de la recherche et de la formation de la Communauté française de Belgique (Direction générale de l'organisation des études), janvier 1993.
- SEDLACEK, W.E., Noncognitive Indicators of Student Success, *Journal of College Admission*, 1989, 125, 2-10.
- SELDEN, R., Missing data, *Phi Delta Kappan*, 1988, 3, 487-491.
- SIMON, H.A., *La science des systèmes, science de l'artificiel*, Paris, EPI, 1974.
- SMITH, M.S., Educational Indicators, *Phi, Delta, Kappan*, 1988, 3, 492-494.
- SKINNER, B.F., *Walden Two*, New York, Macmillan, 1948
- STEVICK, B.F., *Skinner's Walden Two*, New York, Seabury Press, 1968.
- STEWART, J. et al., *Assessment for better learning*, Wellington, Ministère de l'Éducation, 1989, p. 19.
- STEWART, J. et al., *Tomorrow's standards. The report of the ministerial working party on assessment for better learning*, Wellington, Ministère de l'éducation, 1990.
- STIGGINS, R.J., Design and development of performance assessment, *Education Measurement : Issues and Practice*, 1987, 6, 3, 33-42.
- TALMAGE, H., Evaluation of programs, In H.E. MITZEL, Ed., *Encyclopedia of Educational Research*, New York, Free Press, 1982.
- THORNDIKE, R.L., *Reading comprehension education in 15 countries*, Stockholm, Almquist and Wiksel, 1973.
- TORNEY, J., OPPEHEIM, A.N. and FARNEN, R.F., *Civic education in ten countries*, Stockholm, Almquist and Wiksel, 1976.
- U.S. DEPARTMENT OF EDUCATION, National center for education statistics (Research and Development Reports), *International Mathematics and Science Assessments : What Have We Learned ?*, Washington, 1992.
- VAN HERPEN, M., Conceptual models in use for education indicators, In N. BOTTANI et H. WALBERG, 1992, pp. 25-52.
- VIGOTSKY, L.S., *Mind in society : the development of higher psychological processes*, Cambridge, Mass., Harvard University Press, 1978.
- VINCK, D., L'évaluation sociétale des choix technologiques, *Courrier du CETHES*, 1988, 5, 5-16.

- WALBERG, H., *The implications of cognitive psychology for measuring school achievement*, Chicago, University of Illinois, 1991, ronéotypé, pp. 19-22.
- WEDMAN, I., *Pupil assessment with a focus on assessment at secondary education*, Strasbourg, Conseil de l'Europe, DECS/Rech., 1993.
- WYATT, T.J., *What can be learned from the social indicators movement ?*, Sydney, Australian Conference of Directors-General of Education, 1988.
- WYATT, T., Indicators. Trends in the educational literature, in OECD-CERI, *International Educational Indicators Project. A reader on conceptual issues*, Paris, OCDE, 1991.
- WYATT, T. et RUBY, A., Eds, *Indicators in Education*, Sydney, Australian Conference of Directors-General of Education, 1988.
- WYATT, T. et RUBY, A., Eds, *Education indicators for quality, accountability and better practice*, Sidney, Australian Conference of Directors-General of Education, 1990.
- X, THE WHITE PAPER, *Choice and Diversity : A New Framework for Schools*, *School Update*, Autumn 1992, 2-3. (Londres, Department of Education).

Table des matières

Avant-propos	5
Introduction	7
Chapitre 1 – Nature du pilotage	13
Types de pilotage	15
Le pilotage administratif ou monitoring de conformité	15
Le pilotage formatif ou pilotage diagnostique	15
Le pilotage du rendement scolaire	16
Modèle technologique - Modèle humaniste	17
Sur le modèle technologique	17
Sur le modèle humaniste	19
Attitudes technocratiques - Attitudes démocratiques	20
Macropilotage - Micropilotage	21
Le macropilotage	21
Le micropilotage	22
Une complémentarité	22
Sur la valeur ajoutée	23
Cohérence, qualité, équité	27
Récapitulation des questions	29
Définir les buts de l'éducation et s'accorder à leur propos	30
Développer les services de statistiques	31
Piloter le pilotage...	32
Chapitre 2 – Les indicateurs	33
La notion d'indicateurs	35
Nécessité d'un modèle	39
Nombre d'indicateurs souhaitable ?	41
Qualités souhaitées	42
Directives pour une utilisation constructive	42
Synthèse des problèmes généraux	44

Chapitre 3 – Exemples de réalisations	47
I - INDICATEURS NATIONAUX	49
Etats-Unis	49
La "National Assessment of Educational Progress" (NAEP)	49
<i>Les rapports de recherche</i>	51
<i>Une masse de publications</i>	61
<i>The National Education Goals Reports</i>	61
<i>Autres évaluations nationales importantes : des études longitudinales</i>	62
<i>Le Projet Talent</i>	63
<i>National Longitudinal Study, 1972 : entrée dans l'enseignement supérieur et dans la vie active</i>	63
<i>National Educational Longitudinal Study, 1988 : facteurs de la réussite scolaire</i>	64
<i>Des systèmes d'indicateurs propres aux Etats</i>	65
Angleterre	66
L'Unité d'Evaluation du Rendement Scolaire	66
Le Curriculum National et son évaluation: un pilotage systématique	67
Des indicateurs de performance	72
France : Un développement exemplaire	73
L'évaluation des acquis pendant la scolarité	74
L'évaluation de la politique éducative	77
La production et la diffusion d'outils d'évaluation	77
Quelques précisions sur le testage à grande échelle	77
Un remarquable ensemble de publications	78
Canada	80
Au Québec : Les indicateurs de l'enseignement primaire et secondaire	80
Les indicateurs nationaux du Canada	81
Suède	83
Un système précurseur : une harmonisation libérale de la notation	83
<i>Le début</i>	83
<i>Aujourd'hui</i>	84
Evaluation nationale	85
Australie	86
Le pilotage des écoles et des systèmes scolaires dans les Etats.	87
Le projet national	88
Autres pays	89

47	II - Indicateurs internationaux	89
49	L'Association Internationale pour l'Évaluation du Rendement Scolaire (IEA)	90
49	Les débuts de l'IEA : Une étude de faisabilité	91
49	<i>Enquête sur le rendement de l'enseignement des mathématiques (1961-1967)</i>	91
51	<i>Le survey de six disciplines (1966-1973)</i>	95
61	<i>La 2e enquête sur l'enseignement des mathématiques (1980-1986)</i>	98
61	<i>La 2e enquête sur les sciences (1984)</i>	98
62	<i>Etude sur l'utilisation pédagogique des ordinateurs (1987-1990)</i>	100
63	<i>Etude sur l'éducation préscolaire : la qualité de vie des enfants de quatre ans (1986-1996)</i>	100
63	<i>Recherche sur la compréhension de la lecture (1990-1993)</i>	101
64	<i>La 3e étude sur l'enseignement des mathématiques et des sciences (1991-1995)</i>	105
65	Rôle historique et apports méthodologiques de l'IEA	105
66	<i>Une démarche inévitablement longue</i>	107
66	International Assessment of Educational Progress (IAEP)	108
67	L'enquête IAEP I (1987-1988)	108
72	L'enquête IAEP II (1990-1991)	110
73	Une critique fondamentale	110
74	Le projet de l'OCDE : International Indicators and Evaluation of Educational Systems (INES) - Indicateurs internationaux de l'enseignement	111
77	Une première expérience : les indicateurs sociaux	111
77	Le projet INES	113
78	<i>Problèmes particuliers relatifs aux attitudes, aux représentations, aux attentes</i>	121
80	<i>L'avenir</i>	122
80	Chapitre 4 – Micropilotage	125
81	<i>Généralités</i>	126
83	Des réalisations	128
83	En Angleterre : un exemple de checklists d'auto-évaluation	128
84	En France : la Proposition de Peretti	130
85	Au Danemark	132
86	En Nouvelle-Zélande	132
87	Le recours à des experts extérieurs	133
88	Conclusion	134
89		

Chapitre 5 – Le pour et le contre	137
Les bénéfiques escomptés	139
Une rupture d'isolement	139
Les dangers	140
A condition de s'en servir...	142
... et de raison garder	145
Annexe : Aspects techniques	147
Introduction	149
Sur l'évaluation sociétale des choix technologiques	150
Sur l'évaluation de programmes	153
Psychologie cognitive et perspectives nouvelles	155
Implications pour l'amélioration des tests de performances scolaires	156
En conclusion	157
Sur la validité et la fidélité des mesures	157
La généralisabilité	158
L'estimation de l'occasion d'apprendre	160
Epreuves de performance ou épreuves de récitation ?	160
<i>Nécessité de procédures de modération</i>	162
Sur l'administration des épreuves	162
L'échantillonnage multimatriciel	162
La mise en parallèle des tests	163
Sur l'interprétation des résultats	163
La notion de standard	163
La détermination de la compétence minimale	164
<i>Fixation de la note de réussite</i>	165
L'analyse du pourcentage de réussite	169
La probabilité de réponse correcte	170
<i>Quelques précisions sur la théorie des traits latents</i>	171
L'ancrage des échelles	173
<i>Exemples d'échelles de compétence</i>	174
<i>Critique de la théorie des réponses aux items</i> <i>(Goldstein et Wood, 1989)</i>	176
Le fonctionnement différentiel des items	176
Conclusion	177
Principales abréviations	179
Bibliographie	181
Table des matières	189

Série
BON
CAR
COO
DE

DE
de L
DEL

DEP
FOR
FOU
GAC

GAC
JON
LAR
NOE
SPIE

Série
BOU

CAR
CAR
DUR
GAR
HUG
POU
POU

Série
ANC
DE K
FIGA
HUB

Série
BON
MINI
MOR

PÉDAGOGIES EN DÉVELOPPEMENT

Direction

De Ketele Jean-Marie, Belgique

Comité scientifique international

Ben-Omar Boubker, Maroc; Chevrolet Daniel, France;

Girard André, Québec; Tapia Claude, France;

Tahiri Claudine, Côte d'Ivoire et Roosen Antoine, Belgique.

Série 1. *Problématiques et Recherches*

BONNET FR., DUPONT P., HUGET G., *L'école et le management.*

CARDINET J., *Pour apprécier le travail des élèves.*

COOMBS PH., *La crise mondiale de l'éducation.*

DE CORTE E. et al., *Les fondements de l'action didactique*, 2e édition

traduit et adapté du néerlandais par VAN CUTSEM V.

DE KETELE J.-M., (Éd.), *L'Évaluation, approche descriptive ou prescriptive ?*

de LANDSHEERE G., *Le pilotage des systèmes d'éducation.*

DELHAYE G., POURTOIS J.-P., STURBOIS G., *Les acteurs de l'avenir.*

Les défis d'ouvriers, de techniciens et d'ingénieurs au sortir de l'école.

DEPOVER CH., *L'ordinateur, média d'enseignement. Un cadre conceptuel.*

FORQUIN J.-C., *École et culture. Le point de vue des sociologues britanniques.*

FOUREZ G., *Éduquer-Écoles, éthiques, sociétés.*

GAGNE G., LAZURE R., SPRENGER-CHAROLLES L., ROPE F., *Recherches en didactique et acquisition*

du français langue maternelle. Tome 1 Cadre conceptuel. Tome 2 Répertoire bibliographique.

GAGNE G., PAGE M., TARRAB E., *Didactique des langues maternelles.*

JONNAERT PH., *Conflits de savoirs et didactique.*

LAROCHELLE M., DESAUTELS I., *Autour de l'idée de science. Itinéraires cognitifs d'étudiants.*

NOEL B., *La métacognition.*

SPIES-BONG G., *Pour une pédagogie en communauté de vie*

traduction commentée de PETERSEN P., *Le petit plan d'Iéna.*

Série 2. *Recueils*

BOUCHER A.-M., DUPLANTIE M. ET LEBLANC R., *Pédagogie de la communication dans l'enseignement d'une langue étrangère.*

CARDINET J., *Évaluation scolaire et mesure.* 2e édition.

CARDINET J., *Évaluation scolaire et pratique.* 2e édition.

DURNING P. et POURTOIS J.-P., *Education et famille.*

GARNIER C., BEDNARZ N., ULANOVSKAYA I., *Après Vigotsky et Piaget. Perspectives sociale et constructiviste.*

HUGON M.-A. ET SEIBEL C. (Éds), *Recherches impliquées, Recherches-action: le cas de l'éducation.*

POURTOIS J.-P. (Éd.), *Les thématiques en éducation familiale.*

POURTOIS J.-P. (Éd.), *Innovation en éducation familiale.*

Série 3. *Méthodologie de la Recherche*

ANCIAUX A., *Évaluation de la socialité. Manuel d'anthropologie appliquée au travail social.*

DE KETELE J.-M., ROEGIERS X., *Méthodologie du recueil d'information.*

FIGARI G., *Évaluer: quel référentiel ?*

HUBERMAN A.-M., MILES M.-B., *Analyse des données qualitatives.*

Série 4. *Traité*

BONHIVERS B., DE KETELE J.-M., *Pratique de la statistique.*

MINDER M., *Didactique fonctionnelle. Objectifs, stratégies, évaluation.*

MORIN L. BRUNET L., *Philosophie de l'éducation. 1. Les sciences de l'éducation.*

Série 5. Nouvelles pratiques de formation

ALLAL L., *Vers une pratique de l'évaluation formative.*

BRINCKERHOFF R.F., *Sciences, technologies et société au quotidien.*

Avec la collaboration et une préface de G. Fourez.

CALANDE G., DE BUEGER-VANDER BORGHT C. et al., *Plaisirs des sciences.*

Didactique des sciences et autonomie dans l'apprentissage.

CHARLIER E., *Planifier un cours. C'est prendre des décisions.*

DE KETELE J.-M., CHASTRETTE M., CROS D., METTELIN P., THOMAS J., *Guide du formateur.*

DONNAY J. ET CHARLIER E., *Comprendre des situations de formation.*

MERSCH-VAN TURENHOUDT S., *Gérer une pédagogie différenciée.*

MORISSETTE D. ET GINGRAS M., *Enseigner des attitudes ? Planifier, intervenir, organiser.*

STRAUVEN C., *Construire une formation.*

WAGNER M.-C., *Pratique du micro-enseignement.*

ZAY D., *Enseignants et partenaires de l'école. Démarches et instruments pour travailler ensemble.*

Série 6. Pratiques méthodologiques

ABRECHT R., *L'évaluation formative, une analyse critique.*

BASTIN G., *Bien choisir sa profession.*

COPPE M., SCHOONBROODT C., *Guide pratique d'éducation pour la santé.*

Réflexion, expérimentation et 50 fiches à l'usage des formateurs.

DE BRABANDERE L., *Il était une fois la multiplication. Ou découvrir le plaisir des mathématiques.*

GERARD F.M., ROEGIERS X., *Concevoir et évaluer des manuels scolaires.*

GIASSON J., *La compréhension en lecture.*

JONNAERT P., *De l'intention au projet. Concevoir un projet de formation*

LEWIS J., *L'école, prix d'excellence. Avec les meilleures entreprises pour modèle.*

MABILLE A., *Profils de profs. Portraits et styles d'enseignants en sciences.*

ROMAINVILLE M., *Savoir parler de ses méthodes. Métacognition et performance à l'université.*

ROOSEN A., BASTIN G., *L'école malade de l'échec.*

VAN HOUT G., *Et que le nombre soit!*

Le pilotage des systèmes d'éducation

Piloter un avion, c'est le conduire, dans les meilleures conditions possibles, à la destination choisie par les passagers, tout en faisant un usage optimum du matériel et des informations données par les instruments de bord; des signaux se déclenchent en cas de danger.

Piloter un **système éducatif**, un établissement particulier, voire un groupe-classe procède du même souci d'efficacité et de sécurité, et n'est pas incompatible avec une **option humaniste**. C'est ce que comprennent un nombre croissant de pays plongés dans la crise culturelle contemporaine, dont la nouvelle révolution industrielle n'est que l'un des aspects.

L'introduction du **pilotage** interpelle le monde de l'éducation, car il soulève des questions de fond et pose des problèmes techniques complexes.

Après une réflexion générale sur la nature de cette technologie et sur ses orientations majeures, l'auteur explique son fonctionnement et indique, plus particulièrement, ce que sont les **indicateurs** et comment on les utilise.

Les principales réalisations nationales et internationales en matière de

macropilotage et de **micropilotage** sont décrites de façon précise et leurs avantages ou leurs inconvénients éventuels sont examinés.

L'ouvrage s'achève par une annexe traitant d'aspects techniques particulièrement importants.

Gilbert de Landsheere

est Professeur émérite de l'Université de Liège où il a dirigé le Laboratoire de pédagogie expérimentale. Il est spécialiste de la construction des curriculum d'enseignement et de l'évaluation dans leurs divers aspects. Il est auteur ou co-auteur de 17 livres, traduits en 9 langues, et de 250 articles.

Il exerce des fonctions dans plusieurs organisations internationales gouvernementales, notamment au Bureau international d'éducation (BIE-Unesco) à Genève et préside l'Académie mondiale de l'éducation. Il est membre fondateur de l'Academia Europaea.

PILSYS B711
ISSN 0777-5245

ISBN 2-8041-1876-2



PÉDAGOGIES EN DÉVELOPPEMENT

Collection dirigée par Jean-Marie De Ketele, professeur au Laboratoire de Pédagogie Expérimentale de l'Université de Louvain-la-Neuve.

- La collection comporte six séries
1. Problématiques et recherches
 2. Recueils
 3. Méthodologie de la recherche
 4. Traites
 5. Nouvelles pratiques de formation
 6. Pratiques méthodologiques

Série 1 PROBLÉMATIQUES ET RECHERCHES

Cette série accueille des ouvrages qui, sous forme d'une synthèse cohérente, présentent une problématique et/ou pratique dans un ou plusieurs domaines des Sciences de l'Éducation. Elle réserve aussi une place aux recherches inédites importantes. Thèses doctorales remaniées dans leur forme; ensemble coordonné de recherches dans un même domaine. Le comité scientifique accorde un intérêt particulier aux bonnes revues critiques de la littérature scientifique dans les domaines d'actualité.