

Forces et faiblesses de l'évaluation du système éducatif en France

Claude PAIR
ancien recteur d'académie

Octobre 2001

Rapport établi à la demande du
Haut Conseil de l'évaluation de l'école

<u>INTRODUCTION</u>	3
<u>1 - ÉVALUER L'ÉVALUATION ?</u>	4
1.1 <u>LE CONTEXTE</u>	4
1.2 <u>POURQUOI ÉVALUER ?</u>	4
1.2.1 <u>Rendre compte</u>	4
1.2.2 <u>Être responsable</u>	5
1.2.3 <u>Aider à progresser</u>	6
1.2.4 <u>Entrer dans une régulation</u>	6
1.3 <u>DANS QUELS DOMAINES ÉVALUER ?</u>	7
1.4 <u>LES QUESTIONS À POSER</u>	7
1.5 <u>ÉVALUER LES ACTIONS D'ÉVALUATION</u>	9
<u>2 - L'ÉVALUATION DANS TOUS SES DOMAINES</u>	10
2.1 <u>LES ÉLÈVES ET LES ÉTUDIANTS</u>	10
2.1.1 <u>Une évaluation intégrée à l'acte éducatif</u>	10
2.1.2 <u>Les examens et autres modes de délivrance des diplômes</u>	12
2.1.3 <u>Les évaluations de masse</u>	14
2.1.4 <u>En conclusion</u>	17
2.2 <u>LES PERSONNELS</u>	18
2.3 <u>LES UNITÉS</u>	19
2.3.1 <u>Les établissements d'enseignement supérieur : le CNE</u>	20
2.3.2 <u>Les établissements scolaires : indicateurs et évaluation</u>	22
2.3.3 <u>Les académies</u>	25
2.3.4 <u>En conclusion</u>	27
2.4 <u>LE SYSTÈME ÉDUCATIF NATIONAL</u>	27
2.4.1 <u>Évaluations thématiques</u>	27
2.4.2 <u>Le système éducatif dans son ensemble</u>	30
2.5 <u>EN RÉSUMÉ</u>	31
<u>3 - L'ÉVALUATION, AILLEURS</u>	36
3.1 <u>L'ÉVALUATION EN ÉDUCATION : UN CAS PARTICULIER ?</u>	36
3.1.1 <u>L'évaluation des politiques et organismes publics</u>	36
3.1.2 <u>Pour l'éducation, un paysage varié</u>	36
3.2 <u>TROIS ÉTUDES DE CAS</u>	38
3.2.1 <u>En Angleterre</u>	38
3.2.2 <u>En Espagne</u>	39
3.2.3 <u>En Suède</u>	40
3.3 <u>SIMILITUDES ET DIFFÉRENCES</u>	41
<u>4 - FORCES, FAIBLESSES ET MARGE DE PROGRÈS</u>	43
4.1 <u>COMPARER DANS LE TEMPS ET DANS L'ESPACE</u>	43
4.2 <u>SE RÉFÉRER AUX BUTS POURSUIVIS</u>	44
4.3 <u>DIVERS MODES DE PILOTAGE</u>	45
4.4 <u>L'ÉVALUATION DANS UN PILOTAGE CONCERTÉ</u>	46
<u>BIBLIOGRAPHIE</u>	49
<u>LISTE DES PERSONNES CONSULTÉES</u>	53
<u>LISTE DE SIGLES</u>	56

Introduction

Le Haut Conseil de l'évaluation de l'école est chargé de trois missions : expertise, synthèse, proposition. « Il n'a pas pour mission d'évaluer l'école mais de faire l'état de l'évaluation de celle-ci » (Jack Lang, présentation du Haut Conseil, 15 novembre 2000). Pour orienter son action, il doit donc naturellement, au début de son travail, s'interroger sur les forces et les faiblesses du dispositif d'évaluation du système éducatif en France.

Le présent rapport vise à servir de support à une discussion du Haut Conseil. Son but n'est donc pas de donner un avis complet sur les forces et faiblesses, mais de présenter une grille d'analyse et un panorama faisant apparaître des sujets susceptibles d'être explorés de manière plus approfondie. Comme il sera rendu public, il pourra aussi contribuer à la réflexion des divers acteurs de l'Éducation.

Le chapitre 1 décrit rapidement le contexte dans lequel se place aujourd'hui l'évaluation de l'École, puis répond à trois interrogations générales : pourquoi évaluer, dans quels domaines, quelles questions se poser sur les évaluations. Le chapitre 2 examine pour chaque domaine les questions posées. Le chapitre 3 replace la situation française dans le contexte international, notamment européen. Enfin, le chapitre 4 cherche à cerner les forces et les faiblesses transversalement aux domaines d'évaluation.

Les forces et faiblesses de l'évaluation ne peuvent être décrites en ne considérant que l'évaluation elle-même. Elles sont liées à la manière dont celle-ci est reçue et utilisée par ses destinataires et par les acteurs du système éducatif. Pour écrire ce rapport, je n'ai donc pas seulement exploré une documentation abondante, reprise pour l'essentiel en bibliographie. J'ai aussi consulté un certain nombre de personnes, à l'intérieur du Haut Conseil et en dehors, dont la liste figure en annexe. Il ne s'agissait pas d'effectuer une enquête systématique auprès d'un échantillon « représentatif » - le temps était beaucoup trop court pour cela - mais plutôt de recueillir la réflexion de personnes engagées dans une activité d'évaluation ou particulièrement intéressées par ce domaine. Il existe une bonne convergence entre les avis recueillis. La synthèse est mienne, j'en suis le seul responsable. Mais je tiens à souligner la richesse et la variété de ces entretiens, la stimulation intellectuelle qu'ils ont constituée pour moi. Que l'évaluation suscite une telle implication chez mes interlocuteurs est déjà la manifestation d'une force. Je les remercie vivement. Je remercie aussi le secrétariat du Haut Conseil, et en particulier Jean-Claude Émin et Anne-Marie Bourhis.

Ce rapport est également l'occasion de rendre hommage à ceux qui ont créé et établi l'évaluation de l'École en France, en particulier Claude Seibel, Jean-Pierre Boisivon et Claude Thélot.

1 - Évaluer l'évaluation ?

1.1 Le contexte

L'évaluation du système éducatif fait l'objet du titre V de la loi d'orientation du 10 juillet 1989¹ : « L'inspection générale de l'Éducation nationale et l'inspection générale de l'administration de l'Éducation nationale procèdent, en liaison avec les services administratifs compétents, à des évaluations départementales, académiques, régionales et nationales qui sont transmises aux présidents et aux rapporteurs des commissions chargées des affaires culturelles du Parlement. Les évaluations prennent en compte les expériences pédagogiques afin d'en faire connaître les pratiques innovantes. (...) Le Comité national d'évaluation des établissements à caractère scientifique, culturel et professionnel constitue une autorité administrative indépendante. » Le rapport annexé à la loi précise : « L'évaluation s'applique à tout le système éducatif : les élèves, les étudiants, les établissements, les personnels, les services extérieurs, l'administration centrale. Loin de conduire à une mise en concurrence des établissements et de leurs enseignants, elle contribue à l'amélioration du système éducatif en vérifiant la mise en œuvre des objectifs nationaux, en les adaptant aux différents publics auxquels ils s'adressent et en opérant une régulation permanente du système éducatif. »

Si l'on se reporte à la situation des années 80, on peut trouver des raisons à cette apparition de l'évaluation dans la loi : l'évolution de l'économie (crise, recherche d'efficacité et de qualité), de la société (consumérisme, mais aussi demande de transparence et de participation), de l'organisation politico-administrative (décentralisation), des idées (place des sciences sociales, systémique) et de l'École elle-même (développement de la scolarisation, accueil de nouveaux publics, préoccupation de l'échec scolaire)... « L'évaluation est donc engendrée par une situation complexe dont elle essaie de prendre la mesure » (Vogler, 1996). Le mot « évaluation », qui n'était guère employé autrefois, est ainsi devenu un terme à la mode.

Cependant, il faut être attentif aux résistances. Pour J.-M. Barbier (1990), « tout se passe comme si un espace idéologique s'était construit autour de l'idée d'évaluation, structuré par deux pôles : un pôle négatif organisé autour des notions de répression, de sélection, de sanction, de contrôle, et un pôle positif organisé autour des notions de progrès, de changement, d'adaptation, de rationalisation ». Peut-on dire que le second pôle est celui des évaluateurs et des décideurs qui commandent les évaluations, alors que le premier concerne plutôt les évalués ? Dans une enquête auprès d'écoles de l'Union européenne (Mac Beath et al., 1999), 8% seulement répondent que l'évaluation externe est appréciée par la plus grande partie du personnel...

1.2 Pourquoi évaluer ?

1.2.1 *Rendre compte*

Dans sa présentation au Haut Conseil, le Ministre Jack Lang indique que cette instance « doit rendre l'information sur l'école la plus indépendante et transparente possible et favoriser ainsi la qualité du débat public sur l'éducation ». On peut comprendre que le but sous-jacent pour l'évaluation est de rendre compte. Effectivement, dans une démocratie, un service public comme celui de l'Éducation nationale doit rendre compte de son fonctionnement et de la manière dont il utilise les ressources qui lui sont affectées².

1 Journal officiel du 14 juillet 1989.

2 « L'évaluation des politiques publiques (...) répond en effet à une exigence de démocratie en ce qu'elle permet de donner aux citoyens des informations et des appréciations objectives qui leur permettent de mieux fonder leur jugement sur la marche des

Les destinataires de ce compte rendu sont multiples :

- les familles : rendre compte, par exemple, des progrès des élèves mais aussi du fonctionnement des écoles et des établissements, sous toutes ses formes : matériel, organisationnel, pédagogique ;
- certaines composantes du système lui-même : c'est ainsi que les résultats des examens et de l'orientation sont utilisés pour l'admission des élèves dans des établissements où ils pourront poursuivre leurs études ;
- les employeurs, pour qui les diplômes obtenus par les élèves et les étudiants jouent un rôle de signal, et les syndicats, pour qui la question de la reconnaissance des qualifications est importante ;
- la société, sur l'efficacité du système global et de ses unités fonctionnelles, par exemple les établissements et les académies, ainsi que sur certains aspects du fonctionnement du système, en particulier pédagogiques (méthodes, expériences, innovations, réformes...) ; la presse est alors souvent l'intermédiaire ;
- l'État et les autres pouvoirs publics qui sont les principaux financeurs, notamment à travers des instances élues : la loi prévoit une transmission des évaluations aux présidents et aux rapporteurs des commissions compétentes du Parlement.

1.2.2 Être responsable

L'exigence de rendre compte rencontre celle de responsabilité, que les usagers du service public attendent de chacun de ses agents, en particulier de ceux qui occupent une position de direction d'une des unités du système : écoles, établissements scolaires et universitaires, services départementaux, académies. Pour désigner cette mise en responsabilité, les anglo-saxons emploient d'ailleurs le terme *accountability* (littéralement, capacité de rendre des comptes) et ils l'appliquent en particulier aux établissements.

Cependant, en matière éducative, la notion de responsabilité n'est pas simple à cerner et à mettre en œuvre. La responsabilité d'un acteur implique un objectif à atteindre, la capacité de décider et mener des actions et, si le résultat n'est pas conforme à l'objectif, une correction ou une réparation, éventuellement une sanction. Mais en éducation, les résultats ne sont pas faciles à définir. Ils sont souvent déterminés par le système lui-même : par exemple, les orientations à la sortie d'un collège dépendent des décisions qui y sont prises, autant que des connaissances réellement acquises par ses élèves. Dans le cas d'un élève, on peut envisager un niveau minimum, par exemple celui qui est nécessaire pour obtenir un diplôme, et l'indicateur est alors la réussite à l'examen ; mais peut-on se contenter du minimum, ne doit-on pas tendre à ce que l'élève atteigne le maximum de ses capacités... et ce maximum est-il jamais atteint ? Surtout, le résultat dépend à la fois de l'école et de l'élève. Qui donc est alors responsable ? Sans doute pas l'élève seul, comme cela a pu être le cas dans les conceptions d'autrefois où celui qui échouait était considéré soit comme paresseux soit comme incapable ; mais pas non plus l'enseignant ou l'établissement seul. Une solution qui a longtemps prévalu est l'obligation de moyens : la responsabilité des enseignants et des établissements portait sur l'application des règles, pas sur les résultats obtenus. À l'évidence, on demande aujourd'hui une obligation de résultat, mais elle se heurte au fait que, les élèves n'étant pas les mêmes partout, on ne peut obtenir partout les mêmes résultats.

Que retenir ? Être responsable, c'est répondre, donc reconnaître ses partenaires, être pour eux un interlocuteur crédible et, en concertation avec eux, fixer des objectifs ; c'est ensuite apporter des corrections en fonction des résultats obtenus.

services publics » (Rapport de présentation du décret du 22 janvier 1990 relatif à l'évaluation des politiques publiques).
« L'évaluation d'une politique publique (...) a pour objet d'apprécier (...) l'efficacité de cette politique en comparant ses résultats aux objectifs assignés et aux moyens mis en œuvre » (décret du 18 novembre 1998).

Dans un premier temps, l'exercice de la responsabilité demande donc un débat objectif, d'égal à égal : l'évaluation fournit les connaissances sur lesquelles reposera ce débat, elle apporte « un élément de rationalité du dialogue », selon l'expression de B. Saint-Girons (Conférence des présidents d'université, 1998) ; voilà le sens du compte rendu.

Mais la responsabilité pose aussi la question des conséquences à tirer de l'évaluation. On peut penser à des sanctions, positives ou négatives. Mais sur quelle base appliquer une sanction puisque les résultats sont difficiles à définir ? Et à qui l'appliquer puisqu'ils ne dépendent pas d'une seule personne ? Surtout, la sanction, si elle peut éveiller une motivation, n'est pas efficace en elle-même. Il faut donc envisager des remèdes aux difficultés constatées, de manière à rendre possible un progrès, de « contribuer à l'amélioration du système éducatif », comme le dit le rapport annexé à la loi d'orientation.

1.2.3 Aider à progresser

Rendre compte n'est donc pas la seule raison pour évaluer, et à mon avis pas la principale. Par exemple, évaluer les élèves sert, à certains moments, à délivrer un diplôme qui sera un signal à l'extérieur de la classe ; mais, chaque jour, dans la classe, cela les aide à progresser. Les pédagogues distinguent d'ailleurs entre l'évaluation sommative, dans le premier cas – rendre compte d'acquis – et l'évaluation formative, dans le second – aider l'élève à progresser dans sa formation (Peretti, 1980). De même, à quoi sert d'évaluer un établissement ? Les sanctions sont difficiles à prendre, même à ce niveau collectif : peut-on le fermer s'il obtient des mauvais résultats ? On pense parfois à une conséquence sur les moyens attribués ; mais lorsqu'un établissement a des résultats insuffisants, faut-il diminuer ses moyens, au risque de l'enfoncer encore, ou lui en donner davantage, au risque de récompenser un fonctionnement défectueux ? Pour savoir quoi faire, il est nécessaire de dépasser le constat, ou le contrôle, des résultats pour comprendre les raisons d'efficacité ou d'inefficacité. Et l'objectif, c'est alors d'aider l'établissement à progresser. Il en est encore de même pour les personnels. On dit souvent qu'on les évalue pour les noter et ainsi décider de leurs promotions : évaluation sommative. Ce n'est certes pas méprisable, mais il y a quelque hypocrisie à considérer que des notes aussi encadrées par les habitudes et les règles sont les résultats d'une évaluation. En fait, si le but est que le système éducatif progresse, ce ne sera que par le progrès de ses personnels, et le premier objectif de leur évaluation devrait être, non pas la sanction constituée par l'avancement de carrière, mais un progrès dans leur professionnalité.

1.2.4 Entrer dans une régulation

L'évaluation a donc deux buts : rendre compte et aider à progresser. C'est aussi ce qu'écrit Claude Thélot (dans Champy et Étévé, 2000) lorsqu'il distingue deux fonctions pour l'évaluation :

- « une fonction externe : rendre transparent le système et rendre compte de sa réalité aux différents partenaires, usagers et responsables (collectivités locales, familles, élèves, hommes politiques, etc.), susciter un débat public pertinent, favoriser une saine émulation entre unités d'enseignement ;
- une fonction interne : aider les acteurs à réfléchir à leur action au vu de l'état du système dans lequel ils travaillent et par là améliorer cette action même ».

Les deux buts ne sont pas en contradiction : rendre compte peut en effet aider à progresser³. Mais ce n'est pas nécessairement le cas, cela peut par exemple conduire à une concurrence stérile.

³ Le rapport de présentation du décret relatif à l'évaluation des politiques publiques, cité ci-dessus, continue par : « Elle constitue aussi un élément de la modernisation de l'État, car une meilleure connaissance de l'efficacité réelle de l'action de l'administration exerce sur cette dernière une pression analogue à celle que le marché fait peser sur les entreprises ».

En réalité, l'évaluation ne suffit pas à provoquer un progrès. Le progrès demande des décisions. Pour y conduire, l'évaluation doit être intégrée dans un pilotage (Michel, 2001) où elle fait partie d'un processus de régulation. En matière d'éducation, les décisions essentielles sont prises par les acteurs directs, et donc la régulation doit être fondée sur la concertation avec eux.

Nous reviendrons au chapitre 4 sur les composantes d'une telle régulation : l'impulsion, le suivi, l'animation, le conseil, la formation et l'évaluation (Pair, 1998). L'évaluation ne constitue pas à elle seule la régulation comme dans un pilotage mécanique, mais elle nourrit les autres composantes, elle est un relais entre l'action de terrain et la décision globale.

1.3 Dans quels domaines évaluer ?

Nous avons rencontré les domaines d'évaluation au cours de la discussion qui précède ; ils figurent dans la loi d'orientation et son rapport annexé :

- Les élèves et les étudiants : en classe ; en conseil de classe (pronostic en vue de l'orientation, notamment) ; par des examens ou d'autres modes de délivrance des diplômes.
- Les personnels, lors de leur recrutement, puis pour les conseiller et pour leur évolution de carrière.
- Les unités : écoles, établissements scolaires et universitaires, services départementaux, académies, administration centrale.
- Le système éducatif national, dans divers aspects du fonctionnement – pratiques éducatives, expériences, innovations, politiques, réformes... - et globalement, en particulier sur l'efficacité (qui rapporte les résultats au fonctionnement), l'efficience (rapportant le fonctionnement au coût), le rendement (rapportant les résultats au coût) (Thélot, 1993)⁴.

Les deux premiers domaines sont traditionnels (Vogler, 1996) : ils concernent ce que l'on peut nommer la « micro-évaluation ». Le dernier s'est développé plus récemment, même si les inspecteurs généraux de l'instruction publique tiraient déjà des synthèses de ce qu'ils constataient sur le territoire : il s'agit de « macro-évaluation ». L'évaluation des unités demeure jusqu'à présent lacunaire, contrairement à ce qui se passe dans beaucoup de pays étrangers ; la loi d'orientation prévoit pourtant que le projet d'établissement fait l'objet d'une évaluation. Certains parlent à ce niveau de « méso-évaluation ».

On peut se demander si ces domaines, en particulier la micro-évaluation d'une part, la méso- et la macro-évaluation d'autre part, ne sont pas trop différents pour faire l'objet d'un discours commun. Pour J.-L. Derouet, il en résulte que « l'évaluation reste une question sociale qui n'est pas complètement constituée sur le plan scientifique »⁵. J. Heywood (2000) adopte en revanche un point de vue pragmatique : « qu'il s'agisse de l'évaluation de l'apprentissage d'un étudiant ou de l'efficacité des institutions ou des enseignants dans la réalisation cet apprentissage, le processus d'évaluation est le même ». Il nous semble effectivement que les mêmes questions se posent dans tous les cas.

1.4 Les questions à poser

Pour examiner les forces et faiblesses de telle ou telle action d'évaluation, on peut poser les questions suivantes :

- *Sur quoi porte l'évaluation ?* On parle de *référé* pour désigner l'objet évalué, appréhendé à travers des informations ou des données, et des indicateurs : par exemple, dans un examen, on peut distinguer entre les informations constituées par les copies et les activités dont elles sont la trace matérielle, qui sont des indicateurs permettant une inférence sur un

4 D'après le Conseil scientifique de l'évaluation (1996), une évaluation doit permettre d'apprécier si une politique est cohérente, efficace, efficiente, effective (dans ses conséquences pour la société), pertinente (en adéquation avec les problèmes qu'elle veut résoudre).

5 Communication personnelle.

niveau de connaissance ou de compétence. Le référé est évalué par rapport à un *réfèrent* : objectifs avec des normes ou des critères ; dans l'exemple d'un examen, les objectifs peuvent être définis en termes de compétences et les critères indiquer des échelles de performance⁶. En outre, pour savoir ce que l'on évalue réellement, la question se pose du caractère représentatif des informations et indicateurs vis-à-vis de l'objet à évaluer, et des normes et critères vis-à-vis des objectifs. Par exemple, évalue-t-on vraiment un niveau de connaissance ou seulement des copies par rapport à un barème ? La représentativité varie en sens inverse de l'ambition, elle n'est parfaite que pour les référés et référents les plus rudimentaires.

- *Quel type de résultat obtient-on ?* On peut faire une différence entre l'évaluation et le contrôle. Le contrôle a comme résultat des informations sur le fonctionnement concret : les « palmarès » des établissements sont de cet ordre. Le résultat de l'évaluation comporte un *jugement de valeur* : dans l'exemple de l'examen, il porte sur les capacités inférées du candidat, jugées suffisantes ou non relativement aux objectifs ; dans l'évaluation d'un établissement, il peut s'exprimer en points forts, points faibles, recommandations. L'évaluation exige donc une mise en relation d'informations (qui peuvent être fournies par un contrôle préalable). En ce sens, elle produit des connaissances qu'il est possible de partager. Si le résultat est trop simple, trop unidimensionnel, l'évaluation est réductrice ; s'il est trop riche, trop complexe, il est difficile à appréhender et à utiliser.
- *Qui évalue ?* Autrement dit, qui a le droit et le pouvoir de porter un jugement de valeur ? Pour les élèves et les étudiants, ce sont des enseignants : ceux de chaque classe ou groupe, individuellement ou en équipe, et les jurys de délivrance des diplômes. Pour les personnels, ce sont les chefs d'établissement et, en ce qui concerne les enseignants des premier et second degrés, les corps d'inspection. Certaines académies ont mis en place des évaluations d'établissements du second degré et le Comité national d'évaluation évalue les établissements d'enseignement supérieur ; les inspections générales ont commencé des évaluations d'académies. Les inspections générales, la direction de la programmation et du développement (DPD) et certaines équipes de recherche procèdent à des évaluations de certains aspects du fonctionnement du système éducatif. La DPD rassemble des informations pour conduire à une évaluation du système : par exemple, une publication comme « l'état de l'école » donne au fil des années une image des coûts, des activités, des résultats.
- *Comment procède-t-on ?* Dans l'idéal, les composantes de l'acte d'évaluation, notamment le référé et le réfèrent, sont explicites, comme le jugement de valeur qui est ensuite utilisé. Il devrait en être ainsi pour toute évaluation instituée, comme celles que nous étudions dans ce rapport. Des formes dégradées sont l'évaluation spontanée où le référé et le réfèrent ne sont pas tous deux explicités (par exemple l'évaluation par les familles de la qualité d'un établissement) et l'évaluation totalement implicite où le jugement de valeur ne se manifeste que par ses effets. D'autres questions portent sur les modalités du processus, en particulier son échelonnement dans le temps (évaluation ponctuelle ou plus ou moins continue).
- *Quels sont les destinataires et quel usage est fait des résultats ?* Connaître, informer, partager, agir, décider, modifier des rapports entre personnes ou d'une personne à elle-même (estime de soi)... Par exemple, le succès à un examen peut permettre au lauréat d'entreprendre des études ultérieures ou d'accéder à un emploi ; l'évaluation d'un établissement peut transformer les rapports à son intérieur ou avec les parents ; l'évaluation d'une expérience pédagogique peut conduire à l'infléchir ou à l'étendre.
- *En conclusion, quelles sont les forces et les faiblesses de l'évaluation, notamment par rapport aux deux buts envisagés plus haut : rendre compte et aider à progresser ?*

⁶ Dans cette partie, nous utilisons la terminologie de (Barbier, 1990).

1.5 Évaluer les actions d'évaluation

Puisqu'il nous est demandé d'étudier les forces et faiblesses du dispositif d'évaluation du système éducatif, nous analyserons ces questions pour chacun des domaines envisagés ; en outre, nous indiquerons quelques points qui pourraient être approfondis.

On peut dire qu'il s'agit de l'amorce d'une évaluation de l'évaluation de l'École en France. Pour chaque domaine, le référent est constitué des activités d'évaluation, avec des informations écrites (rapports, articles...) et orales (entretiens) sur ces actions, avec des indicateurs sur les résultats et leurs usages. Le référent est lié aux deux buts de rendre compte et d'aider à progresser, avec des critères pour chaque domaine : par exemple, pour rendre compte, un niveau de connaissance des résultats par les destinataires, une pertinence pour une discussion ou des décisions... ; pour aider au progrès, la participation à une régulation fondée sur la concertation.

Mais le jugement de valeur est aussi éclairé par une comparaison avec d'autres pays. Comme l'écrit Fritz Zorn, « les choses que l'on ne compare pas à d'autres sont toujours sans valeur en soi, elles demeurent solitaires et incompréhensibles dans un espace froid, irréel »⁷. Nous donnerons donc des éléments sur quelques pays européens. D'ailleurs, la prise de conscience croissante de l'exigence d'évaluation tient aussi à une évolution internationale, relayée par les organisations intergouvernementales.

Tout cela permettra, en conclusion, de tenter une synthèse dégageant les forces et les faiblesses du système d'évaluation de l'École en France.

⁷ cité par G. Pelletier (1998).

2 - L'évaluation dans tous ses domaines

2.1 Les élèves et les étudiants

La plus grande partie des activités d'évaluation concerne les élèves, parce que la délivrance des diplômes est un des produits du système éducatif, celui qui est le plus visible, et aussi parce que l'évaluation est un des moyens de l'éducation et de la formation. D'après une étude de l'inspection générale portant sur l'enseignement secondaire (IGEN, 1998), l'évaluation en classe occupe un tiers du temps scolaire ; quant aux examens, une partie importante de l'année leur est consacrée dans les lycées, particulièrement longue dans l'enseignement professionnel. Certains types d'évaluation sont traditionnels : la notation des élèves par les maîtres et les examens permettant la délivrance des diplômes. Une forme plus nouvelle est celle des « évaluations de masse » à certains niveaux de l'enseignement scolaire, mises en place systématiquement à partir de 1989 sous la forme d'épreuves standardisées.

2.1.1 Une évaluation intégrée à l'acte éducatif

L'évaluation des élèves par leurs professeurs est aussi vieille que l'enseignement, même si l'usage du terme est récent. Elle est une des composantes de la formation. L'expression d'évaluation formative est apparue autour de 1970⁸. Dans les établissements secondaires, en dehors de la classe proprement dite mais dans son prolongement immédiat, le conseil de classe confronte et synthétise les évaluations des enseignants appartenant à ce que l'on nomme de plus en plus une « équipe pédagogique » ; à ce propos, Baluteau (1993) parle d'une « évaluation très composite ». Le conseil de classe fait des propositions pour l'orientation des élèves. Cette orientation ne fait plus appel à des tests comme autrefois, mais davantage à l'auto-évaluation et aux choix de l'élève, alors que pour les adultes sont apparus les bilans de compétence.

La loi d'orientation de 1989 est sans doute le premier texte législatif qui emploie le mot d'évaluation à ce propos. Elle indique que « la scolarité est organisée en cycles pour lesquels sont définis des objectifs et des programmes nationaux de formation comportant une progression annuelle ainsi que des critères d'évaluation » (article 4), puis que « les enseignants apportent une aide au travail personnel des élèves et en assurent le suivi, ils procèdent à leur évaluation » (article 14) ; pour l'orientation, elle mentionne « la proposition du conseil de classe » (article 8).

L'évaluation, liée au processus d'enseignement-apprentissage, est continue avec des temps forts. Il s'agit largement d'une évaluation « spontanée », au sens où il est rare qu'elle porte sur un référent et un référé tous deux explicites avec un lien clair du second au premier. Le référent doit être cherché dans les programmes : ils ont évolué dans le sens de mieux préciser objectifs et compétences attendues, comme le demande la loi, mais de manière extrêmement variable selon les disciplines ; pour l'école primaire, les livrets scolaires, institués par le décret du 6 septembre 1990 et diffusés sous diverses formes, explicitent un référent très détaillé sous forme de compétences ; au collège, le Ministre a récemment demandé la mise au point de « cahiers d'exigences »⁹. Le référé est constitué par les compétences acquises par l'élève : savoirs, savoir-faire, aptitudes. Du côté des informations qui le manifestent, les anciennes compositions mensuelles ou trimestrielles ont disparu mais ont été remplacées par de multiples devoirs en temps limité ou autres examens blancs ; s'y ajoutent d'autres devoirs, exercices et productions orales, avec des modalités et une fréquence extrêmement variable selon les classes. À l'école

8 A. de Peretti (1980) fait remonter la notion vers 1967 et l'attribue à Michael Scriven. Il indique qu'elle a ensuite été reprise par Bloom et diffusée par les travaux de G. de Lansheere.

9 *Orientations pour l'avenir du collège*, 5 avril 2001.

primaire, le livret scolaire a induit l'utilisation d'exercices d'évaluation. Récemment, une clarification est apparue pour les lycées, avec l'interdiction de mêler évaluation des connaissances et sanction disciplinaire.

Les résultats des évaluations sont variés, mais les notes continuent à occuper une position tout à fait dominante pour les destinataires que sont les élèves et leurs parents, ainsi que pour les enseignants. Dans l'usage qui est fait des résultats, elles prennent le pas, malgré leur caractère « unidimensionnel » ou à cause de lui, sur les autres jugements de valeur plus riches que sont les corrections, appréciations et listes de compétences acquises. En particulier, elles jouent un rôle important en conseil de classe. Les propositions d'orientation sont alors largement fondées sur des moyennes, au détriment d'outils plus fins de pronostic et avec la question que posent les énormes disparités dans la notation. Les classements ont officiellement disparu, mais quelques sondages semblent montrer qu'à l'école élémentaire en tout cas, les élèves et leurs parents continuent de savoir qui occupe les premières places.

On pourrait en tirer la conclusion que ces évaluations sont plus sommatives que formatives. Il n'en demeure pas moins que les jugements de valeur, fréquents et multiples, ont des conséquences importantes, positives et négatives, sur les études des élèves, voire sur leur développement personnel. N'a-t-on pas attribué la moindre propension des jeunes filles à faire des études scientifiques à une plus grande sensibilité aux jugements négatifs de leurs professeurs ? En fait, les usages sont variés, d'une contribution à la motivation et à la formation jusqu'à l'orientation, en passant par le compte-rendu à travers les bulletins et livrets scolaires. Ils ne sont pas toujours très conscients et on ne distingue guère entre les usages diagnostiques, formatifs, sommatifs, probatoires, pronostiques.

On ne peut nier que ces évaluations jouent un rôle effectif tant dans le compte rendu aux élèves eux-mêmes et à leurs parents que dans la progression des élèves et des enseignements. Sur le premier point, elles sont très attendues des familles, malgré des contestations sur les différences entre enseignants et des difficultés de compréhension des évaluations par compétence ; sur le second, leur force est qu'elles entrent dans un processus de régulation (pédagogique) au sens où nous l'avons envisagé au chapitre 1 : c'est d'ailleurs ce que demande l'article 14 de la loi d'orientation. Mais la qualité de cette intégration et le caractère concerté avec l'élève, pour que l'évaluation puisse prétendre à l'épithète « formative », restent très inégaux selon les enseignants et les niveaux d'enseignement. L'enseignement du premier degré semble être celui qui a le plus progressé, même s'il y reste beaucoup à faire, grâce aux outils mis en œuvre : livret d'évaluation organisé en compétences et cahiers d'exercices permettant d'évaluer celles-ci. Dans le second degré, l'enseignement professionnel a plus progressé que l'enseignement général, grâce à l'influence du contrôle continu¹⁰ puis du contrôle en cours de formation, ainsi que de l'évaluation des acquis professionnels ; au collège, les nouvelles formes de bulletin introduisent l'appréciation de compétences, mais ils ne semblent pas utilisés partout. L'enseignement supérieur est beaucoup moins concerné que les enseignements scolaires : le plus souvent, le processus de formation ne fait en rien appel à l'évaluation.

En fait, dès que l'évaluation est intégrée à l'action pédagogique, elle entre dans la liberté pédagogique de l'enseignant. Il existe certes de nombreuses recommandations écrites, mais leur caractère contraignant est incertain et leur mise en œuvre douteuse : il ne s'agit que d'incitations voire d'exhortations. Ainsi, l'inspection générale, dans un rapport rédigé sur ce sujet (IGEN, 1998), relève une grande diversité selon les disciplines, les établissements, les enseignants. Mais elle constate que les pratiques traditionnelles sont largement dominantes, que l'écrit demeure privilégié, que l'évaluation est parfois « chronophage » et que la note, rassurante et confortable pour l'institution, est trop souvent perçue par les élèves comme une sanction. Elle recommande de donner à l'évaluation un cadre réglementaire et d'accompagner

¹⁰ Le contrôle continu, pratiqué dans certains lycées professionnels entre 1972 et le début des années 90, a provoqué une intense réflexion pédagogique.

sa mise en œuvre, en particulier en organisant le travail en équipe et en formant mieux les enseignants. Or, on sait mal comment les questions concernant l'évaluation figurent dans la formation.

En résumé, la variété et les ambiguïtés qui règnent font qu'il est difficile de porter un jugement d'ensemble. Comme nous l'a dit un des membres du Haut Conseil, « c'est un point obscur du système ». Quelques signes montrent un progrès, mais il est lent et inégal. Nous avons dit qu'une force est l'intégration de ces évaluations dans la régulation pédagogique interne à la classe. La faiblesse tient à une insuffisante intégration des questions concernant l'évaluation en classe à la régulation du travail des enseignants. Le progrès est lié au renouvellement des demandes faites aux professeurs et aux outils qui leur sont proposés pour y répondre : c'est ce qui s'est passé à l'école primaire ; dans le second degré, les nouvelles pratiques liées à la pédagogie de projet, par exemple les travaux personnels encadrés, posent des problèmes d'évaluation, mais elles peuvent être facteurs de progrès. Encore faut-il que le réexamen des modalités traditionnelles bénéficie d'une impulsion, d'une animation et d'un suivi de la part des corps d'inspection, et entre dans la formation des enseignants et dans leur propre évaluation.

Des études possibles pourraient porter :

- en ce qui concerne l'objectif de rendre compte, sur la compréhension par les familles des livrets dans le premier degré et des bulletins dans le second, en particulier les changements qu'ont pu apporter les nouvelles formes de bulletin au collège ;
- pour l'objectif d'un progrès dans l'intégration au processus pédagogique, sur l'usage des exercices d'évaluation à l'école primaire, d'une part, sur la place de l'évaluation des élèves en classe dans la formation initiale et continue des enseignants et dans les pratiques d'inspection, d'autre part.

Il serait également utile de faire le point sur la liaison entre évaluation et orientation, à partir de travaux comme ceux de Marie Duru-Bellat et Alain Mingat.

2.1.2 Les examens et autres modes de délivrance des diplômes

Les examens sont aussi un mode traditionnel d'évaluation. Leur but premier est de rendre compte des compétences (savoirs, savoir-faire) des élèves et des étudiants à certaines étapes, en particulier à la fin d'une formation. On parle à leur propos d'évaluation sommative.

En fin de collège, le brevet présente l'originalité de ne plus servir de diplôme de fin d'études et de ne pas avoir de conséquence sur l'accès aux études ultérieures. Cette absence d'usage réel crée un malaise (Salines et Vrignaud, 2001). Les récentes déclarations du Ministre devraient transformer la situation : « le brevet d'études fondamentales (...) sera un élément déterminant au moment de la décision du conseil de classe pour la poursuite d'études en classe de seconde générale, technologique et professionnelle » ; mais comment cela sera-t-il possible sans remettre en cause un processus d'orientation concerté ?

Au niveau du lycée se côtoient le « monument national » (Legrand, 1995) que constituent les baccalauréats, dont le principal usage est de donner le droit d'entrer dans l'enseignement supérieur, et des diplômes qui attestent une compétence professionnelle (CAP, BEP...) ; le baccalauréat professionnel est à l'intersection des deux catégories. On peut ajouter le BTS, diplôme professionnel de niveau supérieur mais préparé dans les lycées. Il faut noter que les diplômes professionnels ne sont pas seulement préparés dans le système scolaire mais aussi en apprentissage et en formation continue.

Tous ces diplômes se placent dans un cadre rigoureux, associant une réglementation nationale très précise et une organisation au niveau académique ou départemental, même si le brevet des collèges fait une place notable à un contrôle continu organisé par chaque établissement, et les diplômes professionnels au contrôle en cours de formation (IGEN, 2000), d'une part, à la validation des acquis d'autre part.

Cette situation contraste avec celle qui a cours dans l'enseignement supérieur, à cause de la tradition d'autonomie des établissements de ce niveau. S'il existe un cadre réglementaire national, la délivrance des diplômes est de la compétence des établissements et les enseignants évaluent leurs propres étudiants. Dans certains cas, notamment les IUT et les grandes écoles, le diplôme est délivré sans examen terminal, par un jury qui tient compte de l'ensemble de la scolarité, y compris d'activités telles que stage et mémoire : il s'agit d'établissements pratiquant la sélection à l'entrée et où les taux de succès sont élevés.

Les modalités de délivrance des diplômes sont donc finalement très variées : plus ou moins de « contrôle continu », ce terme recouvrant des réalités fort diverses ; délivrance par les établissements ou indépendamment d'eux ; intervention ou non des professeurs des candidats ; présence éventuelle dans les jurys de personnes n'appartenant pas au système éducatif ; prédominance de la rédaction ou questionnaires à choix multiple ; épreuves globales ou unités capitalisables...

Le référent est constitué des compétences du candidat. Elles sont évaluées relativement à un référent qui est un programme, les critères et échelles de performance n'étant en général guère explicites. La situation est meilleure pour les diplômes professionnels où sont définis un « référentiel d'activités professionnelles » et un « référentiel de certification » précisant les compétences attendues. En réalité, on n'évalue que les compétences manifestées par les épreuves, et souvent leurs modalités ne permettent pas la prise en compte de certaines compétences fondamentales, par exemple expérimentales ou de communication orale. Le référent réel est donc restreint et sa couverture, forcément incomplète par rapport au référent théorique, n'est guère examinée. Il en est de même pour la justesse et la fidélité de la mesure. Les travaux sur la docimologie (Landsheere, 1980) sont ignorés. Faiblesses certaines...

Et pourtant l'acceptation sociale est bonne : d'une manière générale, les résultats ne font l'objet que de peu de contestations formelles. L'honnêteté de l'évaluation n'est pas réellement suspectée. L'impression prévaut que, malgré l'imperfection de l'évaluation, les précautions prises (par exemple l'harmonisation entre jurys ou la consultation du livret scolaire) et le hasard des compensations font que les résultats reflètent convenablement la valeur des candidats. Le sentiment dominant est qu'ouvrir la boîte de Pandore présenterait plus de dangers que d'avantages... et qu'il s'agit du pire système à l'exception de tous les autres. Curieusement, ce sentiment n'est pas ébranlé par la grande diversité des modalités ni par la comparaison avec l'étranger (Legrand, 1995). Tout se passe comme si la société donnait délégation au système éducatif, et notamment aux enseignants en tant qu'experts, de délivrer des titres qui influenceront sur la position qu'occuperont les individus. Il s'agit d'une fonction « politique » qui dépasse l'évaluation.

La force du système est donc grande. En outre, même si leur premier but est de rendre compte des compétences, les modes de délivrance des diplômes ont une influence sur la formation elle-même, pour l'apprentissage qu'ils motivent comme pour l'enseignement qu'ils orientent. Influence pas nécessairement positive, comme en témoigne le mot « bachotage ». Les résultats servent également à rendre compte du fonctionnement des établissements, à travers des indicateurs et des « palmarès » : nous y reviendrons plus loin ; sur ce plan, on voit la supériorité d'un mode de délivrance qui ne fasse pas intervenir les établissements.

Finalement, il est un peu décourageant de s'attaquer à ce monument multiforme et il n'est pas certain que de nouvelles études puissent réellement conduire à des progrès décisifs. On pourrait, pour tel ou tel examen, par exemple le baccalauréat, mettre à jour pour certaines disciplines les compétences effectivement évaluées par les sujets des dernières années. Ou étudier si l'acceptation sociale est aussi grande que nous l'avons affirmé, notamment dans le cas de l'enseignement supérieur. Ou encore si les diplômes de même intitulé sont réellement comparables d'une université à l'autre...

Cependant cette stabilité et cette acceptation générale pourraient être ébranlées par l'évolution économique et sociale. C'est du côté des diplômés professionnels que le risque est le plus grand, en liaison notamment avec une internationalisation du marché de l'emploi et avec la formation tout au long de la vie. Il faut reconnaître que les modalités de délivrance des diplômes sont en général définies en fonction de la formation initiale à temps plein ; et que l'alternative qu'est la validation des acquis professionnels peine à se mettre en place. Cela pourrait conduire à des évaluations qui soient indépendantes du système éducatif, plus claires et « professionnelles » dans leurs modalités, pouvant faire apparaître un marché de la validation des compétences.

2.1.3 Les évaluations de masse

Les évaluations de masse apparaissent comme la nouveauté la plus marquante en termes de notoriété : lors des entretiens à contenu ouvert, c'est le plus souvent le premier point abordé, celui qu'évoque spontanément le mot « évaluation ». Elles ont été mises en place depuis la rentrée 1989 au début du cycle des approfondissements (CE2) et du collège (sixième), en français et mathématiques, et depuis 1992 au début du lycée dans quatre disciplines : français, mathématiques, histoire-géographie, première langue vivante, pour les secondes générales et technologiques ; français, mathématiques, sciences et techniques industrielles, économie-gestion, pour les secondes professionnelles.

Des efforts importants ont été faits pour définir un référent sous formes de compétences représentatives des programmes et pour y adapter les épreuves (Vugdalic, 1996). Celles-ci se présentent sous la forme d'exercices regroupant des items : pour la sixième en 2000, par exemple, 21 exercices regroupant 87 items. Les résultats sont, pour chaque élève, ses réussites aux items, qu'il est possible de regrouper de diverses manières.

Pour les usages, l'objectif affiché consiste à « doter les enseignants d'un outil pédagogique susceptible de les guider pour définir leurs stratégies d'apprentissage et les actions de remédiation pour leurs élèves, (...) de mieux déceler, en début d'année scolaire, les acquis et les faiblesses des élèves » (Andrieux et Colmant, 2000). L'usage proposé est donc orienté vers les enseignants pour les inviter à prendre en compte individuellement chacun de leurs élèves. En outre, depuis le début de l'opération, il est prévu une restitution aux parents. Les deux objectifs de rendre compte et progresser sont donc bien présents au niveau de la classe.

En outre, il existe en beaucoup d'endroits des traitements statistiques des résultats au niveau local (par exemple un collège, une ZEP, une circonscription, un département, une académie), pour exercer une régulation et un pilotage : analyse des difficultés, impulsion de pratiques pédagogiques, voire gestion des moyens ou définition de secteurs de recrutement. Les résultats des évaluations constituent aussi un matériau pour des recherches didactiques (Rémond, 1996). Au plan national, des synthèses sont faites à partir d'un échantillon représentatif et des moyennes académiques sont publiées, corrigées de l'influence de la composition sociale de la population¹¹ : on peut alors parler d'évaluation-bilan. Un pas n'a pourtant pas été franchi, qui répondrait à une préoccupation fondamentale dans le compte rendu sur l'efficacité du système éducatif, celle de l'évolution des performances au cours du temps : « les protocoles d'évaluation (...) n'ont pas été conçus (...) de façon telle que les comparaisons d'une année à l'autre (...) puissent apparaître pertinentes et méthodologiquement fondées » (Andrieux et Colmant, 2000). Une étude indépendante a été menée pour comparer les performances en lecture en fin d'école élémentaire à dix ans d'intervalle: elle montrait une stabilité des résultats (Baktavatsalou et Pons, 1998)¹².

11 Voir (Andrieux et Colmant, 2000 ; Andrieux et al., 2000) et les divers numéros de *Géographie de l'École*, Ministère de l'Éducation nationale, Paris.

12 Voir aussi *Les Dossiers d'Éducation et Formations* 100, 1997.

Il n'existe pas d'accord sur la compatibilité des divers objectifs des évaluations de masse, les représentants des organisations syndicales d'enseignants étant les plus réservés devant les usages autres que purement pédagogiques. Les arguments de part et d'autre peuvent se résumer ainsi : on dispose d'une information riche qu'il serait dommage de ne pas utiliser ; les évaluations n'ont été conçues que comme un outil pédagogique et les employer, par exemple, pour comparer des performances conduirait à un bachotage réducteur, à une « normalisation », à des résistances.

En revanche, les avis recueillis à propos de la mise en œuvre des évaluations sont très convergents. Il faut distinguer les niveaux CE2 et sixième de celui de seconde. Dans le premier cas, l'opération est acceptée et effectuée quasiment partout ; les résultats sont restitués aux parents. Comme cela nous a été dit, « elle est entrée dans le paysage » ; il semble même que, quelles que soient les réticences qui l'ont accompagnée à son début, des protestations s'élèveraient s'il était décidé aujourd'hui d'y renoncer, comme cela aurait pu arriver dans un passé récent. Selon un sondage de la SOFRES pour le SNUIPP (janvier 1999), telle qu'elle est, l'évaluation CE2-sixième des élèves paraît tout à fait satisfaisante ou satisfaisante à 63 % des enseignants du premier degré, 28 % étant d'un avis contraire ; et il n'y a guère de différence selon la proximité syndicale ou politique, la situation géographique ou l'âge ; on peut seulement noter que les femmes sont moins convaincues que les hommes. Ces résultats recourent les réponses à des interrogations effectuées dans les premières années de l'opération (Chauvet, 1991 ; Férole, 1994). Mais en seconde la situation semble beaucoup moins favorable : d'après les avis recueillis, l'évaluation est loin d'être effectuée partout et elle est exploitée moins souvent encore, les résultats n'étant donc guère communiqués aux collègues dont viennent les élèves. Ici, le résultat ne concorde pas avec les déclarations faites par les enseignants au début de l'opération (Férole, 1994 ; Peretti, 1994) et il est plus pessimiste que celui qui émane en octobre 2000 du panel de chefs d'établissement organisé par la DPD. Pour certains interlocuteurs, on observerait un recul au fil des années. En tout cas, la situation paraît très variable d'un lycée à l'autre selon qu'il existe une politique d'établissement ou que les choix sont laissés à chaque professeur individuellement.

La différence entre les niveaux ne tient pas seulement à l'ancienneté de l'opération, même s'il semble que les efforts d'impulsion et de formation aient été plus importants en 1989 qu'en 1992. Mais, en seconde, le caractère plus prospectif de l'évaluation rend les choses plus délicates, sans que pour autant toutes les compétences travaillées au lycée soient testées. La pertinence est donc discutée. En outre, les enseignants de lycée sont peut-être moins réceptifs que ceux de l'école ou du collège au questionnement pédagogique créé par l'hétérogénéité des acquis des élèves. Enfin, il semble bien que, tant à l'administration centrale que dans beaucoup d'académies, on ait renoncé à ce niveau à tout effort d'impulsion et d'animation.

Cependant, même au CE2 et plus encore en sixième, il semble que l'objectif fondamental de mettre un outil à la disposition des enseignants pour mieux répondre aux besoins individuels des élèves ne soit pas véritablement atteint. Les « programmes personnalisés d'aide et de progrès », souhaités par Madame Royal, ne sont guère mis en place. Au mieux, les résultats sont utilisés en début d'année pour constituer des groupes de besoins ou de remise à niveau et les exercices constituent des références pour les enseignants. Certains observateurs pensent que l'ergonomie de l'outil est en cause et que les obstacles matériels sont un prétexte pour ne pas approfondir les résultats individuels. D'autres disent que les enseignants n'ont l'habitude que d'évaluations unidimensionnelles traduites par une note. Un professeur de français regrette « un excès d'analyse », considérant que les problèmes sont beaucoup plus globaux, et pas seulement d'ordre cognitif. Un IA-IPR parle d'une « étrangeté radicale des évaluations nationales par rapport au travail quotidien des maîtres, comme s'il n'y avait pas de rapport entre les deux » ; il ajoute que la prise en compte des indications apportées par les évaluations supposerait que l'enseignant « procède par succession de choix entre plusieurs solutions didactiques ou pédagogiques possibles » alors que « dans les faits, les pratiques sont univoques, les choix sont rares ». Il y aurait même une « usure » chez les enseignants qui

étaient les plus intéressés au début, due à une « routine » ; un professeur qualifie l'opération de « pensum ».

Les remarques précédentes sont fondées sur des informations trop parcellaires pour constituer une évaluation de ces évaluations, qui reste à faire. On peut cependant dire qu'une force est l'acceptation générale de l'opération, en tout cas aux niveaux CE2 et sixième ; cette acceptation a créé un mouvement qui, à l'école primaire, converge avec l'évaluation périodique reportée sur le livret scolaire. Le compte-rendu aux familles est fait à ces deux niveaux ; il l'est moins souvent en seconde, tant aux élèves qu'à leurs parents. La principale faiblesse porte sur l'atteinte de l'objectif d'intégration à la pédagogie, et donc de contribution au progrès.

Que faire sur ce point, outre de rendre les outils plus conviviaux ? On peut d'abord penser que « l'étrangeté » dont il était question plus haut vient du caractère trop ponctuel de ces opérations : une fois à l'école, une autre au collège, une dernière au lycée, et dans un nombre limité de disciplines ; il est irréaliste d'espérer développer ainsi une « culture de l'évaluation ». Une réponse réside dans l'existence d'une « banque d'outils d'aide à l'évaluation », mise à la disposition des enseignants depuis 1992 pour l'école, puis le collège, puis la classe de seconde ; ces outils ont été diffusés sur support papier et sont accessibles depuis peu par Internet ; mais ils restent encore mal connus : dans beaucoup d'établissements, on a oublié jusqu'à leur existence et l'accès par Internet semble rester confidentiel. Un appel doit être lancé aux académies pour les enrichir, ce qui devrait aussi favoriser leur utilisation. En outre, c'est dans le cadre de cette « banque » que sont prévues, à la rentrée 2001, de nouvelles évaluations au début de la dernière année d'école maternelle et du cours préparatoire. C'est aussi de cette manière qu'il est prévu de revoir l'opération en seconde et qu'est envisagée l'extension annoncée récemment par le Ministre au début de la classe de cinquième et à d'autres disciplines en sixième.

En réalité, on peut craindre que le problème soit plus profond. Une opération nationale d'évaluation n'a de chance de s'imposer réellement que si elle est intégrée dans un processus de régulation concertée, donc si elle apparaît elle-même comme concertée et si elle est liée aux autres composantes de la régulation. Or, l'écoute du terrain par l'académie et de l'académie au plan national n'apparaît pas systématiquement organisée. L'impulsion n'a pas été continue : par exemple, lorsque des mesures ont été décidées pour les collèges et les lycées en 1999, l'évaluation ne semble pas avoir été prise en compte pour les décisions et elle n'a pas non plus été reprise dans les mesures, laissant l'impression qu'elle ne figurait plus dans les priorités ; en revanche, elle est explicitement présente dans les orientations de 2001 sur l'avenir du collège. Si la situation semble mieux se présenter à l'école qu'au collège et surtout qu'au lycée, c'est parce qu'un certain nombre d'IEN l'ont intégrée dans leur activité d'animation, de suivi, de conseil. C'est sans doute moins vrai pour les IA-IPR ; comme l'écrit l'un d'entre eux, animateur académique pour ces évaluations : « les enseignants ont été laissés trop seuls pour l'exploitation des résultats qu'on a voulu trop exclusivement lier à l'individualisation de l'enseignement (...) l'encadrement pédagogique devrait se pencher sur les résultats (par champs de compétences, par items), en déduire des urgences, des priorités et combattre un certain nombre d'idées reçues ». Des syndicalistes ne disent pas autre chose lorsqu'ils constatent que « le dispositif d'aide et de suivi n'a pas fonctionné ». Quant à la présence de l'évaluation dans la formation des enseignants, elle paraît tout à fait marginale en formation initiale et n'avoir pas tenu dans la durée en formation continue.

Dire ainsi que ces évaluations doivent trouver leur place dans la régulation du système fournit une réponse à la question posée plus haut : elles ne peuvent se limiter à l'objectif initial d'être un outil pour chaque enseignant dans sa classe. Certes, il y a des dangers à une extension, mais il faut les affronter... et il vaut mieux le faire avant que, sous l'influence de l'évolution technique, l'évaluation des élèves soit devenue un marché. Il est même douteux qu'on puisse se limiter à des usages de nature purement pédagogique, sans envisager des questions comme celles de la définition des secteurs scolaires ou de la répartition des moyens. Après

tout, pourquoi des indicateurs de résultats au baccalauréat sont-ils publics sans qu'on puisse les relier publiquement aux évaluations à l'entrée du lycée ?

On ne peut donc que recommander que soit entreprise une étude approfondie des usages de ces évaluations de masse, qui puisse ensuite nourrir une discussion sur la manière de procéder à une relance.

2.1.4 En conclusion

Les trois grands types d'évaluation que nous venons d'évoquer apparaissent extrêmement différents dans leurs buts comme dans leurs modalités. Les évaluations de masse se font au début d'un cycle, elles ont un caractère de diagnostic. En classe, l'évaluation a un but formatif ; mais elle débouche aussi sur l'orientation qui demande un pronostic. Les diplômes résultent d'une évaluation finale, de caractère sommatif.

Socialement, les évaluations les plus visibles sont les examens dont le but est de rendre compte des compétences. Et pourtant, l'orientation constitue un enjeu plus fort. Comme l'écrit P. Legrand (1995), « le véritable baccalauréat, l'élève ne le passe pas au mois de juin de l'année terminale ; il l'a déjà passé, par morceaux, au fur et à mesure des barrages qui ont jalonné son parcours ». Peut-être est-ce une raison de l'acceptation sociale des examens, que nous avons soulignée. Effectivement, l'orientation, malgré une progressive libéralisation, paraît susciter davantage de rancœurs.

Pour le but d'aider les élèves à progresser, les trois types d'évaluation y contribuent, même si c'est l'évaluation en classe qui est ici la plus décisive. Une des difficultés est que ces trois types participent de conceptions disjointes. Par exemple, à l'interface entre le collège et le lycée, on trouve les trois sortes : l'orientation, le brevet, l'évaluation de début de seconde, qui ne sont pas articulées et dont personne ne sait trop si elles donnent des résultats comparables.

On peut penser que ce sont les évaluations de masse qui sont conçues de la manière la plus claire, parce qu'elles sont fondées sur une analyse des compétences mises en jeu, et qui sont les plus proches d'une véritable évaluation. Pourtant, ce sont les plus discutées et celles dont l'influence est la plus faible, parce qu'elles n'entrent pas vraiment dans le processus pédagogique. Le principal enjeu est aujourd'hui de savoir comment elles peuvent servir de noyau pour influencer les autres, vers une évaluation mieux analysée et plus formative, une meilleure comparabilité entre les classes, des examens plus fiables. Il existe quelques progrès sur ces points, qui paraissent plus nets à l'école primaire. Les confirmer et les accentuer passe par une intégration de l'évaluation des élèves dans une régulation concertée du système éducatif aux divers niveaux : établissement, département et académie, national.

L'évaluation peut avoir un autre but, non plus individuel mais global : rendre compte des résultats du système éducatif par un bilan des connaissances et compétences acquises par l'ensemble d'une classe d'âge, ou par l'ensemble des élèves à certains niveaux ; et si possible envisager son évolution dans le temps. Les examens ne se prêtent pas à cet usage. Il faut plutôt utiliser des épreuves standardisées appliquées à un échantillon représentatif de la population considérée. C'est ainsi que les évaluations de masse conduisent à des bilans annuels, mais elles ne sont pas construites pour que des conclusions puissent être valablement tirées dans la durée. La DEP a effectué des bilans des acquis des élèves dans divers domaines et à divers niveaux, et certaines d'entre elles présentent des comparaisons temporelles (Pons, 1996 ; Dessus et al., 1996 ; Flieller, 1997) ; mais les études de ce type se sont raréfiées aujourd'hui. Le ministère de la Défense fournissait le niveau général des conscrits à l'aide de tests effectués chaque année dans les centres de sélection ; dans le cadre des journées d'appel de préparation à la défense est maintenant passé un test de lecture (Jeantheau et Murat, 2000). La validité de ces « évaluations-bilans » pose des questions d'ordre psychométrique (Salines et Vrignaud, 2001) ; en particulier, la question du choix du référent est délicate. Faut-il se rapporter aux programmes scolaires, alors qu'il n'est pas certain qu'ils

reflètent bien les besoins sociaux et alors qu'ils sont fréquemment modifiés ? Ou, comme pour certains tests internationaux, tenter de définir ces besoins, mais se mettre d'accord sur eux est bien difficile et eux-mêmes varient dans le temps ? Ou encore se fonder sur des tests de capacités intellectuelles générales, mais peut-on leur attribuer un sens en eux-mêmes ou reflètent-ils l'état changeant d'une société ? D'ailleurs, on a constaté dans divers pays que les performances aux tests d'intelligence d'élèves du même âge s'améliorent au fil des années, sans que ce phénomène soit véritablement expliqué (Flieller, 2001).

2.2 Les personnels

L'évaluation du travail des personnels est un autre domaine traditionnel, auquel beaucoup d'efforts sont consacrés. Elle constitue une grande part du travail des corps d'inspection.

Nos interlocuteurs hésitent à s'exprimer à ce sujet. Pour ceux qui n'appartiennent pas à l'Éducation nationale, l'évaluation des personnels est une activité purement interne au système, qui ne les concerne pas, sauf peut-être dans le cas des personnes notoirement insuffisantes « dont il est impossible de se débarrasser ». Pour les représentants des personnels, il s'agit d'un domaine délicat. Comme l'écrit Alain Picquenot¹³ : « Il n'existe pas de réelle différence de traitement entre celui qui fait son travail dans l'excellence et celui qui se contente de faire le minimum. Paradoxalement, les enseignants qui regrettent cet état de fait ne semblent pas très friands d'une évaluation qui permettrait de distinguer les meilleurs : ils redoutent l'arbitraire du chef d'établissement ou de l'inspecteur, ils ne veulent pas être notés à la tête du client ». Un large accord s'exprime pour dire que les personnels ne font pas l'objet d'une évaluation, mais seulement d'une notation.

En évaluant les personnels, on peut viser deux buts : impulser des progrès individuels et collectifs, et en particulier favoriser l'évolution des fonctions qu'exerce la personne évaluée ; gérer l'avancement de carrière. En réalité, le second but a pris le pas sur le premier, au point de l'occulter dans l'esprit des personnels et de l'administration.

Pour les enseignants du supérieur, le Conseil national des universités traite de la qualification à l'entrée dans les corps et des promotions de classe à l'intérieur d'un corps : l'évaluation faite alors porte surtout sur la recherche, les pratiques étant cependant variables selon les disciplines. Mais la moitié des promotions de classe sont maintenant décidées localement, ce qui permet de faire intervenir les tâches d'enseignement et les fonctions assurées dans le cadre de l'établissement. De même, les personnels ingénieurs, techniciens, administratifs de l'enseignement supérieur et de la recherche (ITARF) sont évalués et promus par leur établissement. En outre, conformément à l'arrêté du 9 avril 1997, une évaluation des enseignements par les étudiants se met progressivement en place : elle n'a aucune conséquence sur la carrière des enseignants, mais elle favorise le progrès des contenus et méthodes d'enseignement (Conférence des présidents d'université, 1998)¹⁴. On peut donc dire que la situation s'est améliorée ces dernières années dans un secteur où elle était loin d'être satisfaisante.

Les choses sont différentes dans l'enseignement scolaire. Ici, il ne s'agit que de rendre compte pour gérer les promotions, et à un destinataire qui n'est pas une personne mais un système d'avancement automatisé : nous sommes dans un contexte bureaucratique au sens de M. Crozier, ce qui a l'avantage de minimiser les conflits mais pas de contribuer au progrès.

Le résultat utile est alors une note, même si elle est accompagnée d'un rapport ou d'une appréciation faisant largement appel à la langue de bois. En outre, la note n'a pas de signification en elle-même, mais seulement en référence à la note précédente (normalement, elle augmente au fil de la carrière... contrairement à ce qui se passe pour les élèves) et aux

13 *Le Courrier d'Éducation et Devenir*, avril 2001.

14 Sur ce sujet, voir aussi le chapitre 4 de (Heywood, 2000).

notes des collègues du même échelon : c'est pourquoi elle est soigneusement encadrée car on se méfie de la subjectivité du notateur. Pour J. Vogler (1996) « le degré d'acceptabilité du système d'évaluation du personnel se mesure à son caractère illusoire, voire à sa quasi-inutilité ».

La préoccupation de la représentativité du référé par rapport à ce qui est censé être évalué – la qualité du travail – et celle de son rapport à un référent disparaissent alors : le référé n'est guère représentatif dans le cas des inspections, qui sont rares et se limitent encore souvent à l'observation d'un cours, et il reste très implicite dans le cas des notations par les chefs d'établissement qui portent sur des attitudes plus que sur des compétences. En outre, il n'existe pas de véritable référent : des référentiels précisant les compétences attendues ont bien été publiés en 1994 et 1997 pour les enseignants des premier et second degrés, mais ils ont été présentés comme ne s'appliquant qu'à la formation ; pour les ATOS, il existe un outil, nommé OMEGA, qui analyse les divers métiers, mais il n'est guère utilisé pour l'évaluation. Les pratiques sont très différentes d'un évaluateur à l'autre. Les concours de recrutement souffrent de défauts analogues : se soucier davantage de rendre compte du « niveau » des candidats que de faire un pronostic sur leur adaptation au métier et leur capacité de progrès. Dans les IUFM, l'amélioration de l'évaluation des stagiaires est une préoccupation, mais il est difficile d'avoir sur elle une vue d'ensemble : il est significatif que cet aspect ne soit pratiquement pas abordé dans le rapport que le Comité national d'évaluation vient de consacrer à ces établissements (CNE, 2001 b).

Une conséquence est que dans tous les cas où l'on a besoin d'informations précises, soit pour des promotions de corps soit pour des personnels en difficulté soit pour des postes à exigences particulières, les évaluations effectuées ne suffisent pas. Il faut utiliser d'autres voies ne donnant pas toujours aux personnels et à l'administration les garanties requises. Quant aux personnels agrégés, certifiés ou IATOS exerçant dans l'enseignement supérieur, ils concourent pour leur avancement avec leurs collègues du second degré, alors que les pratiques de notation sont différentes.

Des travaux avaient été menés par l'inspection générale de l'Éducation nationale (Ferrier, 1996 ; Ovaert et Pietrzyk, 1996) mais ils paraissent abandonnés. Plus récemment, des propositions ont été faites (Pair, 1998 ; Monteil, 1999) pour faire contribuer l'évaluation au progrès des personnels en le replaçant dans une régulation, mais elles n'ont pas été suivies d'effet. Pourtant, l'existence de postes à exigences particulières devrait faire évoluer le dispositif. Pour les personnels de direction, qui ne sont plus évalués depuis 1993 (Blanchet, 1999), le protocole d'accord signé en novembre 2000 comprend un référentiel des fonctions et des modalités d'évaluation conciliant un but « d'accompagnement, d'aide et de soutien » et un but relatif à l'évolution de carrière : ici, la régulation est concrétisée par un diagnostic de l'établissement établi par son chef et partagé par sa hiérarchie, donnant lieu à une lettre de mission du recteur qui sera ensuite la base de l'évaluation. Il reste à mettre en œuvre ce protocole.

En conclusion, dans l'état actuel des choses, les efforts consacrés à ce domaine n'ont pas de réelle efficacité. L'impression domine chez nos interlocuteurs qu'il faudrait séparer évaluation et notation. Mais le problème renvoie à des questions statutaires avant de relever d'une appréciation sur l'évaluation.

2.3 Les unités

Contrairement à ce qui se passe pour l'évaluation individuelle (élèves, étudiants, personnels), il n'existe pas en France de tradition de l'évaluation des unités qui composent le système éducatif. C'est dans l'enseignement supérieur que les choses ont commencé, avec le Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel (CNE), créé en 1985. Dans l'enseignement scolaire ont été diffusés à partir de 1993 des « indicateurs de pilotage » pour les établissements secondaires (IPES), puis en 1998

pour les écoles ou plus exactement le cheminement des élèves de l'école au collège (INPEC) ; en outre, des indicateurs concernant l'éducation prioritaire sont en cours de mise en place. À partir de 1999, l'évaluation des académies a été entreprise par les Inspections générales.

2.3.1 Les établissements d'enseignement supérieur : le CNE

« La création du CNE en 1985 est une initiative de promotion des établissements. La vocation de ce comité est en effet d'évaluer les universités et non les disciplines, (...) dans un système universitaire tellement marqué par les logiques disciplinaires » (Musselin, 2000). En réalité, le CNE réalise aussi quelques rapports globaux sur l'enseignement de disciplines, mais c'est un aspect minoritaire de son activité : nous y reviendrons plus loin.

Le CNE est créé par la Loi, celle de 1984 sur l'enseignement supérieur. La loi d'orientation de 1989 sur l'éducation en fait une autorité administrative indépendante. Ses dix-sept membres sont nommés, pour une durée de quatre ans non renouvelable, par le Président de la République en Conseil des Ministres, après une procédure de proposition par de hauts organismes. Il dispose d'un budget propre. Rien n'a donc été épargné pour lui donner autorité et autonomie.

Deux décrets, du 21 février 1985 et du 7 décembre 1988, définissent son fonctionnement. Il s'agit d'abord de son champ d'investigation. « Le Comité assure, au cours d'une période de quatre ans, l'évaluation de l'ensemble des établissements publics à caractère scientifique, culturel et professionnel ». En fait, cette périodicité de quatre ans n'a pas été respectée jusqu'à présent. « Il peut également (...) procéder à l'évaluation d'autres établissements d'enseignement supérieur » dépendant du ministère de l'Éducation nationale ou d'autres départements ministériels : cette extension du champ retenu dans le titre du Comité a été effectivement mise en œuvre.

Les décrets précisent les résultats attendus et leurs destinataires. « Les analyses du Comité national d'évaluation sont consignées dans des rapports élaborés par établissement et par thème. Les rapports par établissement sont adressés au ministre chargé de la tutelle de ces établissements. Ils sont, en outre, adressés aux responsables de ces derniers. Les rapports par thème sont adressés au ministre de l'Éducation nationale et aux autres ministres concernés. Les activités du Comité font l'objet d'un rapport adressé annuellement au Président de la République. En outre, le Comité national d'évaluation dresse tous les quatre ans un bilan de synthèse sur l'état de l'enseignement supérieur. Ce bilan est adressé au Président de la République. Le rapport annuel et le bilan de synthèse sont rendus publics ». En fait, tous les rapports sont publics.

La loi précise l'objet de l'évaluation : « les réalisations dans l'accomplissement des missions » des établissements, c'est-à-dire « la formation initiale et continue ; la recherche scientifique et technologique ainsi que la valorisation de ses résultats ; la diffusion de la culture et l'information scientifique et technique ; la coopération internationale ». À la lecture des rapports, on constate que l'évaluation des établissements ne porte pas seulement sur un état, comme c'est essentiellement le cas lors d'évaluations individuelles (l'examen des élèves ou l'inspection d'un professeur, par exemple) mais plutôt sur un fonctionnement¹⁵. Il est alors moins facile de définir un référent. Celui-ci se trouve implicitement dans la confrontation des expériences des différents évaluateurs. D'ailleurs, le choix a été fait que les membres du Comité et les experts qui les assistent ne soient pas des spécialistes en évaluation : c'est conforme à la tradition universitaire d'évaluation « par les pairs ». Il existe cependant un guide de l'évaluation destiné aux établissements (Comité national d'évaluation, 2001 a) et un mémento pour les experts (Comité national d'évaluation, 2000 a). Le CNE vient de mettre en place un groupe technique de spécialistes de l'évaluation pour contribuer à l'amélioration de ces documents et plus

15 (Barbier, 1980) fait aussi une différence entre l'évaluation des personnes formées et celle des formations, parlant dans ce dernier cas de l'évaluation d'un processus. Voir encore (Bouvier, 1998).

généralement pour perfectionner la méthodologie. Les rapports sont plus descriptifs qu'interprétatifs, mais ils comportent des recommandations.

L'évaluation d'un établissement associe un aspect interne, d'auto-évaluation, et un aspect externe, par des experts extérieurs à l'établissement. Elle comprend les phases suivantes, qui s'étalent sur une année :

- la préparation, au cours de laquelle des informations sont collectées par les services du CNE auprès de directions ministérielles et d'autres organismes¹⁶, et des études préalables effectuées ;
- l'évaluation interne, faite par l'établissement sur la base du guide d'évaluation ;
- l'analyse des résultats de cette évaluation interne par une équipe du CNE, qui la rapproche de l'information rassemblée de manière à préparer les expertises ;
- la visite des experts dans l'établissement, dont la durée est en principe de trois jours ;
- la rédaction d'un projet de rapport, soumis au Comité en séance plénière ;
- la mise au point du rapport final, tenant compte des remarques faites par l'équipe de direction de l'établissement, et son approbation par le Comité ;
- la publication du rapport, sur papier et sur le serveur Internet du CNE, accompagné de la réponse du responsable de l'établissement.

Il s'agit donc bien d'un rapport de l'ensemble du Comité, et non pas celui d'une équipe d'experts, et il est largement accessible. Au-delà des destinataires prévus par les textes réglementaires, il est d'ailleurs diffusé à la communauté universitaire (en particulier aux présidents d'universités), aux Parlementaires, aux collectivités territoriales et organismes consulaires concernés et à la presse nationale et régionale¹⁷. Il est aussi placé sur le site Internet de CNE, qui est largement consulté¹⁸. Un document succinct reprend en quatre pages les grandes lignes de l'évaluation et les principales recommandations. L'usage de l'évaluation pour rendre compte est donc très marqué. La question est de savoir ce qu'en font effectivement les destinataires. La principale faiblesse est le manque d'éléments de comparaison entre les établissements : en particulier, on ne sait rien sur les compétences des titulaires d'un même diplôme dans les diverses universités. Notons cependant que, dans le cadre de l'évaluation de disciplines (voir 2.4.1), le CNE effectue des « cotations comparatives » des établissements sur divers critères (Comité national d'évaluation, 1999).

L'usage pour aider à progresser est moins clair. Dans les établissements, le processus d'évaluation est en général bien accueilli, il est perçu comme une aide et parfois certaines recommandations sont appliquées avant même qu'il soit achevé. Mais il semble que l'évaluation a surtout son impact au niveau de l'équipe de direction, et éventuellement par des discussions dans les Conseils : elle a une influence sur le « gouvernement » de l'établissement plus que sur les activités d'enseignement et de recherche directement. D'ailleurs, les rapports ont un aspect structurel marqué et, en particulier, ils ne dissipent pas l'obscurité qui règne sur les modalités et la qualité de l'enseignement. Pourtant, l'évaluation interne conduit à une introspection utile, le CNE recommande qu'elle associe tous les personnels et la manière dont elle a été menée fait partie de l'évaluation. En outre, douze à dix-huit mois après la diffusion du rapport, il est prévu d'organiser une réunion dans l'établissement pour apprécier la mise en œuvre des recommandations, qui est également étudiée lors de l'évaluation suivante.

L'usage par les directions ministérielles est moins clair encore, et on peut voir là le revers de l'indépendance du Comité. En effet, la volonté de promouvoir l'échelon de l'établissement, dont l'évaluation est une des manifestations, se traduit aussi par la contractualisation quadriennale que le ministère mène avec les établissements. Mais, jusqu'à présent, ces deux opérations ont été conduites indépendamment l'une de l'autre, ne serait-ce que pour leurs calendriers, de sorte

16 Par exemple, les indicateurs INFOSUP fournis par la DPD sur l'enseignement supérieur et, pour la recherche, les indicateurs de l'Observatoire des sciences et des techniques (Barré, 2000 ; Mustar, 2000).

17 En tout, 600 exemplaires sont diffusés pour chaque rapport.

18 300 à 400 consultations par jour au premier semestre 2001 et 60 000 téléchargements de publications entre la création du site en juin 1996 et août 2001 (Comité national d'évaluation, 2001 c).

que les discussions conduisant au contrat ne tiennent guère compte de l'éventuelle évaluation de l'établissement. Or, comme le dit B. Alluin (Conférence des présidents d'université, 1998), « l'évaluation ne prendra vraiment corps dans les universités que si on la lie aux conditions dans lesquelles s'élaborent et le projet et le contrat ». Ici encore, l'évaluation n'est pas intégrée dans un processus de régulation dont fait partie la contractualisation.

Il faut ajouter que d'autres instances procèdent à des évaluations plus partielles, mais guère coordonnées avec celles du CNE : l'Inspection générale de l'administration de l'éducation nationale et de la recherche, la Commission du titre d'ingénieur, la Mission scientifique universitaire, le Comité national du CNRS et les instances des autres établissements publics de recherche, le Comité national d'évaluation de la recherche, la Cour des comptes et les Chambres régionales des comptes, la Direction de l'enseignement supérieur pour les habilitations de formations et à travers ses conseillers d'établissement... Cette dispersion n'est sans doute pas favorable à un véritable impact des évaluations.

Cependant, dans ses orientations pour la période 2000-2004, le CNE, sans abdiquer son indépendance, notamment sur les méthodes, prévoit de mettre en phase ses évaluations avec la procédure de contractualisation, en les faisant immédiatement en amont (Comité national d'évaluation, 2000 b). Cela conduira à deux conséquences : procéder à l'évaluation de chaque établissement tous les quatre ans, comme le prévoit la réglementation, même si pour des raisons pratiques une opération sur deux serait plus approfondie et l'autre plus légère ; comme pour la contractualisation, effectuer en même temps l'évaluation des établissements d'une même académie, ce qui permettra aussi des synthèses géographiques : déjà, en 2001, le site d'Aix-Marseille a été examiné.

En conclusion, si l'évaluation des établissements d'enseignement supérieur est délicate (Duru-Bellat et al., 2000), celle que fait le CNE a le mérite d'exister et d'avoir ouvert une voie associant un volet interne et un volet externe, cohérente avec l'autonomie des établissements. Il reste à approfondir son usage, tant par les établissements et leurs personnels que par les autorités de tutelle. Une étude pourrait donc être menée sur l'usage des rapports par leurs destinataires à titre de compte-rendu d'une part et pour le progrès à l'intérieur des établissements d'autre part.

2.3.2 Les établissements scolaires : indicateurs et évaluation

Des indicateurs de pilotage ont été construits pour les établissements secondaires (IPES). Ils diffèrent légèrement selon qu'il s'agit d'un collège, d'un lycée d'enseignement général et technologique, d'un lycée professionnel. Dans tous les cas, ils portent sur la structure de l'établissement, les caractéristiques des élèves, leur origine scolaire, les enseignants, les résultats et le devenir des élèves, l'environnement de l'établissement. Ils portent sur l'année scolaire en cours et sur un historique.

Les indicateurs sont calculés par la DPD, notamment à partir d'informations fournies par les établissements pour alimenter la base de données « Scolarité » concernant les élèves. Ils se présentent sous la forme de tableaux statistiques. Les établissements ont accès aux valeurs de leurs indicateurs et à des moyennes académiques et nationales. En outre, ils disposent de références nationales pour des indicateurs, plus qualitatifs, qu'ils calculent eux-mêmes pour leur établissement (indicateurs de compte rendu). Enfin, chacun d'eux a la possibilité de construire d'autres indicateurs en fonction des besoins qu'ils ressentent, et des outils leur sont fournis pour cela.

Un sondage fait en 1995 auprès d'un panel de chefs d'établissement montre qu'une large majorité d'entre eux ressent un besoin d'indicateurs pour piloter leur établissement et dialoguer avec le Conseil d'administration, les collectivités locales et les services académiques. Et pourtant, il semble bien que les indicateurs IPES soient très peu utilisés. D'après une enquête

récente, seul un établissement sur douze en ferait usage¹⁹. Au niveau central, on pense que les raisons sont techniques : l'évolution des logiciels informatiques, du système DOS à Windows, aux réseaux Intranet et à Internet, n'a pas été maîtrisée ; il existe plusieurs modes d'accès dont aucun ne donne satisfaction ; les questions de protection et de droit d'accès sont laissées aux techniciens alors qu'il s'agit de questions de nature politique. Ces raisons sont sans doute réelles, mais elle sont partielles : d'ailleurs, même avant ces difficultés techniques, l'utilisation de l'outil était très minoritaire. Il serait donc illusoire de penser qu'une solution technique soit suffisante. À la base, on parlerait plutôt d'un outil trop lourd et compliqué : près de 50 tableaux statistiques pour les lycées et de 30 pour les collèges, tous sur le même plan, sans même parler des indicateurs à calculer ou à construire par les établissements. Dans bien des cas, les chefs d'établissement préfèrent des outils plus simples fournis par les services académiques ou qu'ils se définissent eux-mêmes. Il semble donc que ces indicateurs ne répondent pas aux besoins ressentis par leurs utilisateurs potentiels, besoins qui se transforment d'ailleurs avec les évolutions pédagogiques et la plus grande autonomie laissée aux établissements en ce domaine. Il demeure aussi une méfiance envers un contrôle possible par la hiérarchie, en particulier à travers la remontée des indicateurs dits « de compte rendu ».

Dans l'enseignement primaire, des indicateurs (INPEC) sont donnés en plus petit nombre sur la population, l'environnement et les cheminements des élèves à l'école puis au collège, à plusieurs niveaux : école individuelle, secteur de collège, circonscription, département, académie, France. Ces indicateurs semblent peu utilisés par les inspecteurs, mais encore moins par les écoles à qui ils ne sont d'ailleurs pas en principe accessibles : vue hiérarchique des choses ?

Dans les deux cas, il s'agit d'indicateurs destinés au « pilotage » (Landsheere, 1980). Ils peuvent être utilisés dans le cadre d'une évaluation, mais ils ne constituent pas en eux-mêmes une évaluation. Ils ne sont d'ailleurs pas publics.

Il existe cependant une exception. Trois indicateurs sont publics et diffusés largement. Ils concernent les performances des lycées au baccalauréat (Émin et al., 1998) :

- taux de réussite, rapportant les candidats reçus aux candidats présentés,
- taux d'accès au diplôme des élèves entrés en seconde (en première pour les lycées professionnels),
- proportion des bacheliers parmi les sortants.

De tels indicateurs de performance conduisent à une évaluation, mais ils n'en constituent que le référent. Le référent ne saurait être une moyenne nationale, alors que les lycées n'ont pas des populations comparables. Parmi les caractéristiques des élèves déterminantes pour la réussite figurent l'âge et l'origine socioprofessionnelle. Les valeurs brutes des trois indicateurs sont donc accompagnées de valeurs de référence, ou « valeurs attendues » : celles qu'ils prendraient si, en moyenne, les élèves du lycée connaissaient les mêmes succès que les élèves de même âge et de même origine socioprofessionnelle dans l'ensemble des lycées français.

On a donc ici une évaluation des lycées rendant compte de leurs performances. Elle est de bien meilleure qualité que les « palmarès » publiés par la presse, qui ne mentionnaient que le taux brut de réussite à l'examen, masquant par là les politiques sélectives en cours de scolarité et oubliant l'influence des caractéristiques de la population scolaire. Cependant, s'il est vrai que l'âge puis l'origine socioprofessionnelle sont des caractéristiques déterminantes de la réussite, il en est une autre qui est sans doute plus déterminante encore : le niveau scolaire à l'entrée²⁰. Il est donc dommage qu'il ne soit pas pris en compte, à cause d'une certaine timidité dans l'utilisation des résultats de l'évaluation des élèves en début de lycée.

19 Information communiquée par la mission de l'évaluation de la DPD.

20 C'est ce que semblent montrer les études faites à d'autres niveaux d'enseignement : voir par exemple (Duru-Bellat et Mingat, 1993 ; Schmitt-Rolland et Thauriel-Richard, 1996 ; Thauriel-Richard et Verdon, 1996).

Chaque année, la presse fait bien écho à la publication des trois indicateurs. En 1995, un panel de proviseurs a été interrogé sur eux par la DEP : une grande majorité en a eu connaissance, pense qu'ils constituent un progrès par rapport aux classements publiés par la presse et considère qu'ils permettent de mieux situer leur lycée par rapport aux autres ; une majorité indique que la réputation des établissements se fondera à l'avenir sur des bases plus objectives, mais que cela ne revient pas à mettre les établissements en concurrence ni à accentuer la fuite des élèves ; pour les deux tiers des répondants, ces indicateurs ont donné lieu à des discussions avec les enseignants de leur lycée ; cependant, ce n'est qu'une courte majorité qui croit qu'ils tiennent compte du fait que les lycées n'accueillent pas des élèves de même niveau. On peut cependant noter une nuance entre les proviseurs des lycées généraux et ceux des lycées professionnels, les seconds apparaissant un peu moins convaincus : la raison est peut-être que l'efficacité d'un lycée professionnel ne tient pas seulement à la réussite aux examens, mais aussi à l'insertion professionnelle des élèves sortants sur le marché de l'emploi.

Le but de rendre compte est donc bien atteint à travers ces indicateurs. Mais ils ne portent que sur les lycées et même, pour les lycées professionnels, sur le seul cycle conduisant au baccalauréat ; et ils ne couvrent pas l'ensemble des objectifs du lycée. Par exemple, l'insertion professionnelle n'apparaît pas, même pour les lycées professionnels. Surtout, ils ne font que constater sans guère contribuer à aider les établissements à progresser, alors que le « pilotage » devrait avoir ce but. Mais, pour cela, comme nous l'avons dit au chapitre 1, il faudrait qu'il soit fondé sur une régulation à base de concertation, dont l'une des composantes est une évaluation.

Il faudrait donc évaluer le fonctionnement des établissements, ce qui n'est pas fait systématiquement, contrairement à ce qui se passe dans l'enseignement supérieur²¹. Diverses tentatives ont été effectuées ici ou là : par les inspections générales au début des années 90, sur trois échantillons d'une centaine de collèges, de lycées d'enseignement général et technologique, de lycées professionnels ; mais le but était moins d'évaluer chaque établissement que d'en tirer des leçons au plan national sur chacune des catégories ; par quelques académies, depuis plus de dix ans (Casteran et Ouevrard, 1990), sur certains établissements, souvent ceux qui rencontraient des difficultés. On peut encore citer la procédure d'homologation des établissements scolaires français à l'étranger²².

Le seul exemple systématique est celui de l'académie de Lille qui, entre 1991 et 1995, a effectué un « audit » (selon le terme utilisé) de ses 520 collèges et lycées. Ici, le référent est double : les règles et objectifs nationaux complétés par les priorités académiques, d'une part, le projet de l'établissement d'autre part. Le résultat est un court rapport comportant une description du fonctionnement et une liste d'environ cinq points forts, cinq points faibles, cinq recommandations. L'usage visé est d'aider au progrès de l'établissement en premier lieu, et aussi de l'académie dans son ensemble. Les destinataires du rapport sont donc d'abord l'établissement, puis les autorités académiques. Puisqu'il s'agit d'impulser un progrès, l'évaluation associe une participation des acteurs de l'établissement et un regard extérieur. Pour ce second aspect intervient une équipe d'audit formée de quatre personnes : deux inspecteurs, le chef d'un autre établissement, une personnalité de l'Éducation nationale ou extérieure. Ces équipes d'audit reçoivent une formation et disposent d'outils. L'audit lui-même comprend sept étapes :

- le recueil des informations, avec la participation de l'établissement, des inspecteurs y ayant évalué l'enseignement, des services académiques ;
- l'exploitation des informations par l'équipe d'audit pour déterminer les points à approfondir ;
- une visite de l'équipe dans l'établissement pour rencontrer les différents acteurs ;
- la rédaction d'un rapport provisoire ;

21 (Dubet, 1999) le propose pour les collèges. L'académie de Besançon a entrepris de mettre cette recommandation en pratique.

22 Elle fait l'objet d'un dossier établi par les ministères de l'Éducation nationale et des Affaires étrangères (Agence pour l'enseignement français à l'étranger), accessible sur leurs sites Internet.

- la restitution de ce rapport à l'établissement, et d'abord à son chef, avec discussion ;
- la mise au point du rapport définitif.

On remarque une parenté avec la procédure du CNE pour l'enseignement supérieur. Un peu de la même façon, une synthèse des rapports est produite chaque année. Mais, contrairement au CNE, l'académie de Lille n'a pas rendu publics les rapports : l'usage est d'aider au progrès, et pas de rendre compte, si ce n'est aux acteurs directs. Cette opération a fait l'objet d'une double description et évaluation, par l'académie elle-même (Cahiers de l'académie de Lille, 1997) et par une équipe de recherche (Demailly, 1999). Au-delà des retombées sur les établissements, les chercheurs constatent que « l'audit a été (...) un outil d'émergence d'acteurs collectifs aussi bien au niveau de certains établissements que de l'Académie elle-même, et un incontestable outil de formation des cadres ».

En conclusion, il existe un trou majeur dans l'évaluation de l'École en France : celle des établissements scolaires. Pourtant, nous disposons d'un exemple qui, même s'il peut être perfectionné, sur la clarté de ses objectifs notamment, a été effectué avec succès pour tous les établissements secondaires d'une académie. Le cas de l'enseignement du premier degré est sans doute plus difficile à cause de la dispersion des écoles et de l'absence de statut clair des directeurs ; cependant, certains inspecteurs commencent à évaluer des écoles. Les indicateurs fournis par le Ministère peuvent être une base pour ces évaluations, mais elle ne saurait s'y réduire. Ils sont d'ailleurs peu utilisés : une étude pourrait être entreprise auprès des établissements et des services académiques pour cerner les obstacles et les besoins réels, besoins qui émergeraient sans doute mieux s'il existait une évaluation du fonctionnement des établissements. Car il existe une dialectique entre indicateurs et évaluation : les indicateurs sont une base pour l'évaluation, mais l'évaluation peut aider à définir des indicateurs de pilotage ; ainsi, les évaluations d'établissements par les inspections générales, mentionnées plus haut, ont contribué à la définition des indicateurs IPES²³.

2.3.3 Les académies

En 1998, un rapport faisait au ministre de l'Éducation nationale la recommandation suivante : « Évaluer le fonctionnement des académies :

- la direction en charge de l'évaluation rassemble et organise des informations ;
- l'inspection générale évalue périodiquement le fonctionnement de chaque académie ; cette évaluation est conclue par un rapport sur lequel le recteur fait ses observations ;
- le rapport est la base d'un séminaire entre le ministre, entouré des directeurs, et le recteur, accompagné par l'équipe par direction de l'académie. » (Pair, 1998).

Les inspections générales se saisissaient de cette recommandation et proposaient au Ministre de la mettre en application. Dès l'année scolaire 1998-99, elles entamaient une opération expérimentale, avec les académies de Bordeaux et de Rouen, pour « cerner le champ des questions à aborder, formuler les règles de l'investigation et élaborer la méthode d'enquête adaptée à un type de travail nouveau »²⁴. Les deux premières évaluations, en 1999-2000, portaient sur les académies de Limoges et de Rennes, deux académies du « haut du panier » selon les indicateurs de résultats²⁵. Cinq autres académies sont en cours d'évaluation en 2000-2001²⁶. Il s'agit donc d'une opération récente, sur laquelle il est encore difficile de porter un jugement, mais qui apparaît prometteuse.

L'évaluation porte seulement sur l'enseignement scolaire, le supérieur étant traité par le CNE qui, selon ses nouvelles orientations, évaluera la même année les établissements d'une même

23 Le Haut Conseil a demandé une étude à P. Dasté sur l'évaluation des lycées.

24 Avant-propos du rapport sur l'académie de Limoges (IGEN et IGAENR, 2000).

25 Voir *Géographie de l'École*.

26 Amiens, Lyon, Orléans-Tours, Poitiers, Strasbourg.

académie et pourra donc faire à partir de là un rapport de synthèse ; mais il paraît difficile à court terme de coordonner les calendriers du CNE et des inspections générales.

Les deux rapports publiés en septembre 2000 (IGEN et IGAENR, 2000) annoncent dans leur titre : « évaluation de l'enseignement ». Cependant, comme l'indique le rapport de Limoges : « Plus que d'une étude de l'état des disciplines, qui ne paraît pas réellement constituer un élément discriminant d'une académie à l'autre, les évaluateurs ont tenté de mesurer la réussite des élèves dans l'académie et de rechercher ce qui concourt à produire leurs résultats et contribue à expliquer la qualité de leurs apprentissages. Ce qui se passe dans les classes est évidemment essentiel, mais s'inscrit dans un ensemble plus large incluant par exemple les politiques d'orientation des élèves, d'élaboration de la carte des formations, de gestion des enseignants, de formation continue... comme le contexte sociologique dans lequel s'inscrivent les établissements ». C'est donc bien le fonctionnement de l'académie qui est évalué, en partant d'un état des lieux mais en étudiant les grandes problématiques stratégiques et en les replaçant dans le contexte régional²⁷. Le référent, s'il n'apparaît pas explicitement dans les rapports, est constitué par les orientations nationales et le projet de l'académie.

Les évaluations sont effectuées par des équipes d'inspecteurs généraux qui, dans les deux cas considérés, comportaient onze et huit personnes ; elles bénéficient de contributions des autres inspecteurs généraux intervenant dans l'académie. La méthode est décrite dans l'avant-propos du rapport de Limoges : elle « s'appuie en premier lieu sur l'analyse d'un dossier constitué par l'académie et sur des données fournies par l'administration centrale²⁸, puis, à partir des hypothèses que cette analyse conduit à élaborer, sur des entretiens conduits en académie ainsi que sur des visites d'établissements ou de circonscriptions. (...) Les entretiens ont évidemment concerné les différents responsables et acteurs de l'éducation nationale à l'échelon du rectorat comme des inspections académiques mais également grand nombre de ceux qui sont directement intéressés par le fonctionnement du système scolaire (...) : personnels, parents d'élèves, autres services de l'État, conseil régional, conseils généraux, représentants des maires ».

Le résultat est formé de rapports riches et intéressants. S'ils partent d'indicateurs statistiques, ils vont beaucoup plus en profondeur par des analyses quantitatives et qualitatives. Cela leur permet, d'une part, de questionner les statistiques²⁹, d'autre part de dégager des interrogations pour l'avenir en attirant l'attention sur les limites de certaines méthodes pédagogiques qui ont jusqu'à présent conduit à de bons résultats³⁰ : ce dernier élément est particulièrement important pour la politique des académies.

Le rapport est conclu par des recommandations à l'adresse de l'académie et du niveau national. Pour ce dernier, il s'agit moins de recommandations concernant le cas particulier de l'académie que le système éducatif en général : pour le rapport de Rennes, par exemple, elles sont présentées comme des « pistes de réflexion pour le pilotage du système éducatif ». Il y a donc deux destinataires. Le recteur a l'occasion de faire des observations avant la rédaction finale du rapport, mais il ne lui est pas demandé de réponse officielle et aucune discussion n'est formellement organisée entre l'académie et les responsables nationaux au plus haut niveau, comme cela avait été proposé en 1998.

27 Le rapport sur Rennes comprend trois parties : l'état des lieux (analyse du contexte, du fonctionnement, des résultats et des performances) ; les grandes problématiques stratégiques ; la Bretagne, École et société. Celui de Limoges est organisé en six chapitres : l'académie dans la région Limousin ; l'efficacité de l'enseignement, les résultats des élèves ; la scolarité des élèves, les acquis et les flux d'orientation ; l'organisation de l'académie ; la vie pédagogique et éducative ; les ressources.

28 Il s'agit surtout de la DPD : les inspections générales ont élaboré avec elle un ensemble d'indicateurs pour servir de base à l'évaluation des académies. En 2001, la DPD a décidé de relancer un produit, nommé OPERA, donnant des indicateurs de pilotage des académies.

29 « Les données statistiques disponibles présentent des contradictions difficilement explicables » (Limoges). « Le croisement des données montre que des précautions sont nécessaires dans l'utilisation des outils statistiques (...) il y a là un champ de réflexion pour la direction de la programmation et du développement et pour le Haut Conseil de l'Évaluation » (Rennes).

30 « Un modèle menacé ? » (titre d'un paragraphe du rapport de Rennes).

Il est encore difficile de se prononcer sur les usages de l'évaluation à l'intérieur de l'académie et dans ses relations avec l'administration centrale, et en particulier de savoir si l'évaluation entrera bien dans un processus de régulation du fonctionnement des académies par l'administration centrale, accompagnant la politique contractuelle qui se met en place. On peut seulement dire que, pour les deux académies évaluées, après une période d'expectative, le rapport est utile pour le pilotage de l'académie et qu'il est intervenu dans le processus de contractualisation.

Cette évaluation des académies est une opération novatrice très importante, même si elle n'a pas toujours été bien vécue dans les académies concernées. Une étude pourrait être menée sur les usages dès que l'on disposera d'un plus grand nombre de cas.

2.3.4 En conclusion

L'évaluation des unités, si elle reste un point faible pour les écoles et établissements scolaires, s'est imposée depuis quinze ans dans l'enseignement supérieur et elle vient d'être mise en place pour les académies. Ces deux opérations, ainsi que l'audit des établissements secondaires de l'académie de Lille, utilisent une méthodologie proche qui, comme l'indique le CNE, « ne croit pas possible le recours exclusif à des indicateurs de performance et prend en compte le contexte, la situation spécifique, l'évolution, les objectifs particuliers de l'institution évaluée ». Cette convergence est certainement une force pour développer ce type d'évaluation et en faire davantage un outil de progrès des unités considérées et de régulation du système éducatif. Il n'en demeure pas moins que l'emploi d'indicateurs de performance des élèves et étudiants est nécessaire pour permettre des comparaisons dans le temps et dans l'espace.

2.4 Le système éducatif national

Alors que l'évaluation des unités, lorsqu'elle existe, cherche à aboutir à un jugement global, pour le système éducatif national il existe de nombreuses évaluations thématiques de divers aspects de son fonctionnement (politiques et pratiques éducatives en particulier) plutôt qu'une évaluation d'ensemble.

2.4.1 Évaluations thématiques

Un certain nombre d'acteurs concourent à ces évaluations, chacune ayant son domaine :

- l'inspection générale de l'Éducation nationale (IGEN), pour les aspects pédagogiques dans le premier et le second degrés ;
- l'inspection générale de l'administration de l'Éducation nationale et de la Recherche (IGAENR), pour les aspects administratifs, financiers et organisationnels, à tous les niveaux, y compris sur la recherche depuis 1999 ;
- le comité national d'évaluation (CNE) sur divers aspects de l'enseignement supérieur³¹ ;
- la direction de la programmation et du développement (DPD) du Ministère, et auparavant la direction de l'évaluation et de la prospective (DEP), en particulier sur les pratiques et innovations pédagogiques dans le premier et le second degrés ;
- des équipes de recherche des universités, de l'INRP, du CEREQ...
- la Cour des comptes sur l'utilisation des moyens³².

Pour les inspections générales, et en particulier l'IGEN, la part de ces évaluations au niveau national dans leur activité s'est notablement accrue au cours des années 90 et elle continue à croître. Pour le CNE, elle est très minoritaire par rapport à l'évaluation des établissements. Dans les actions d'évaluation réalisées par la DPD, elle est significative mais minoritaire.

31 Le Haut Conseil a demandé un rapport à J. Dejean sur l'évaluation des enseignements dans les universités.

32 Par exemple la gestion des enseignants du secondaire.

Le mode de choix des thèmes explorés est variable selon les organismes. Le CNE les détermine de manière totalement indépendante ; les inspections générales proposent un programme de travail annuel qui est arrêté par le(s) ministre(s), mais en outre des commandes plus ponctuelles leur sont passées en cours d'année ; pour la DPD, le programme de travail résulte en partie de demandes des autres directions ministérielles ; la direction de l'enseignement scolaire (DESCO) passe des contrats avec des équipes de recherche.

C'est ainsi que le CNE a publié des évaluations disciplinaires (la géographie en 1989, les sciences de l'information et de la communication en 1993, l'odontologie en 1994, la formation des cadres de la chimie en 1996, le troisième cycle de médecine générale et la formation des pharmaciens en 1998) et des rapports thématiques concernant les enseignants du supérieur, la valorisation de la recherche, le sport à l'université... De même, les sujets abordés par l'IGEN se partagent entre des évaluations relatives aux disciplines et d'autres qui leur sont transversales. Pour l'année 2000, on trouve par exemple, dans la première catégorie, la place des travaux pratiques en physique-chimie au lycée, la formation continue des professeurs de sciences économiques et sociales, les techniques de l'information et de la communication en sciences de la vie et de la terre ; dans la seconde, la place de l'oral dans les enseignements de l'école au lycée, le contrôle en cours de formation en lycée professionnel, l'évaluation de l'opération « la main à la pâte » (IGEN, 2000). L'IGAENR a traité la même année, entre autres, de la situation sociale des étudiants, de la sécurité lors des sorties scolaires, de la gestion des personnels, du pilotage par les projets et la contractualisation, de la politique de la recherche au niveau régional (IGAENR, 2000 b). À son programme de travail, la DPD a placé en 2001 le travail en commun des enseignants, l'éducation à l'orientation, les apports des aides-éducateurs, le suivi des élèves en classes-relais, l'aide individuelle et le travail personnel des élèves, etc. Le programme d'études de la DESCO concerne essentiellement les filières et diplômes professionnels.

On peut constater que les thèmes abordés sont bien liés aux préoccupations d'actualité. Ils réussissent moins à avoir un caractère prospectif ou à attirer l'attention sur des points négligés : par exemple, on semble avoir découvert brutalement il y a quelques années que l'enseignement des sciences était négligé à l'école élémentaire, sans que l'on dispose d'aucun élément chiffré sur ce sujet. Une autre faiblesse est l'absence de coordination entre les travaux des uns et des autres. On note une certaine sous-représentation des sujets concernant l'enseignement technique, sauf dans les travaux confiés par la DESCO à des équipes de recherche.

Pour la plupart des études, le référent, explicité plus ou moins clairement, est trouvé dans les objectifs nationaux et les textes officiels, mais il existe aussi des présupposés implicites. La méthode associe un recueil préalable d'informations (éléments statistiques, évaluations et rapports antérieurs...), des enquêtes, des observations, des entretiens. La DPD procède largement par questionnaires : elle recueille ainsi sur les pratiques pédagogiques des déclarations plutôt que des observations, et ne porte pas véritablement de jugement évaluatif (voir par exemple Thauvel-Richard, 1999). L'IGEN, grâce en particulier à l'appui des corps d'inspections académiques, dispose davantage d'observations, sans que l'on sache très bien jusqu'à quel point elles reflètent des pratiques effectives. L'IGAENR effectue aussi chaque année des observations dans près de 200 établissements et dans les services de toutes les académies. Les équipes de recherche sont mieux placées pour effectuer des observations systématiques qui exigent un travail important, fournissent une masse de données lourde à exploiter et conduisent souvent à des descriptions plus qu'à des évaluations (Altet et al., 1996).

Une faiblesse de la plupart de ces travaux est que les performances du système, en particulier l'efficacité de l'enseignement fondée sur un bilan des acquis des élèves ou des étudiants, figurent rarement parmi les critères d'évaluation ; d'ailleurs, l'évolution de ces acquis dans le temps ne fait guère l'objet de bilans systématiques. Même s'il existe des exceptions³³, on en

33 Parmi les études de la DEP, certaines ont mis en relation les résultats des élèves avec les pratiques déclarées par leurs enseignants : voir (Schmitt-Rolland et Thauvel-Richard, 1996 ; Thauvel-Richard et Verdon, 1996).

reste plutôt à des évaluations de conformité ou à des descriptions de pratiques. La raison est peut-être de faisabilité : une technique expérimentale classique, faisant appel à une population témoin comparable à la population testée et placée dans les mêmes conditions, est souvent hors de portée pratique. Cependant, il existe des méthodes analytiques, utilisant une modélisation multivariée, permettant d'estimer l'effet sur la progression des élèves d'un facteur étudié « toutes choses égales par ailleurs » et donc d'évaluer l'efficacité ou le rendement d'une mesure (Duru-Bellat et Mingat, 1993 ; Jarousse, 2000) : Meuret (2001) parle de méthodes « corrélationnelles ». Une telle étude a été faite, par exemple, pour l'effet de la scolarisation à deux ans sur les acquis des élèves en français et en calcul au cours des trois premières années de l'école élémentaire (Jarousse et al., 1992)³⁴. On pourrait suggérer par exemple une étude sur la relation entre les moyens accordés aux académies et les résultats obtenus.

Ces approches analytiques, courantes dans les pays anglo-saxons, ont été développées en France par l'Institut de recherche en économie de l'éducation (IREDU). Elles ont notamment contribué à mettre en évidence des « effets maîtres » et des « effets établissements ». Elles restent pourtant peu utilisées dans notre pays. Leur limite tient au processus de modélisation qui, d'une part, est forcément réducteur pour le critère d'évaluation³⁵ et qui, d'autre part, peut laisser échapper des facteurs importants. En outre, les études faisant appel à ces méthodes ne prennent pas toujours assez en compte le temps nécessaire pour qu'une innovation ait des conséquences appréciables sur les acquis des élèves. Mais la limite tient aussi à l'obscurité pour les utilisateurs de la technique statistique utilisée, rendant difficile la discussion des résultats lorsqu'ils semblent contre-intuitifs.

Pour l'ensemble des évaluations considérées dans cette partie, les résultats sont des rapports, des notes, des articles. Ils fournissent une information très riche, ce qui est une force pour le but de rendre compte. Mais rendre compte à qui ? Les rapports des inspections générales s'adressent formellement au(x) ministre(s) et sont accompagnés d'une liste des destinataires proposés. Les articles sont en principe destinés à un public plus large³⁶. En fait, on a l'impression que les destinataires ne sont pas réellement identifiés. Dans beaucoup de cas, on peut penser qu'il s'agit de responsables du système, mais les rapports les atteignent à un moment où les décisions correspondantes ne sont pas nécessairement à l'ordre du jour. C'est une des raisons pour lesquelles le but d'aider le système éducatif à progresser n'est sans doute guère rempli par ces travaux.

L'IGAENR (2000 a), examinant les suites données à ses rapports parle de « manque de visibilité des résultats ». Ce terme paraît adéquat : les rapports sont bien lus par un certain nombre de décideurs qui en retirent une information qu'ils réinvestissent ultérieurement, sans que le lien de cause à effet soit en général bien clair. Comme le dit l'IGAENR : « des procédures tombées en désuétude, des niveaux de décision multipliés, un fort accroissement du nombre des enquêtes, tout contribue aujourd'hui à rendre flou le paysage de l'exploitation des résultats ». Et elle constate qu'entre 1995 et 1999, près de la moitié de ses rapports sont restés sans suite et seulement 20% ont reçu une exploitation largement positive. Encore, parmi eux, seule une minorité concerne-t-elle l'ensemble du système éducatif et non une académie ou un établissement particulier. Ces rapports ont alors une caractéristique commune : il s'agit de « l'accompagnement d'une politique ministérielle » où « la commande passée à l'inspection tendait à évaluer les conditions de mise en œuvre d'une action décidée par l'échelon politique (...) l'IGAENR était invitée à participer à la construction d'un dispositif en temps réel, en apportant les éléments d'information nécessaires à l'autorité responsable »³⁷.

34 Voir aussi, sur d'autres sujets, (Grisay, 1993 ; Danser et al., 2001).

35 D'ailleurs, pour Perrenoud (1998), « aucune évaluation du système éducatif ne peut se référer à une définition à la fois précise et consensuelle des finalités de l'enseignement, donc à une appréciation unanime et fiable de son efficacité ».

36 On doit signaler l'importance de revues comme *Éducation et Formations* et *Les Dossiers d'Éducation et Formations*.

37 Une décision ministérielle vient de mettre en place pour les rapports d'inspection générale une « commission des suites ». Attendons pour voir...

Ces constats s'étendent aux autres évaluateurs. Souvent, les rapports ne servent qu'à nourrir d'autres rapports. On se trouve rarement dans un processus de pilotage régulé où des évaluations précèdent les décisions, qui sont suivies de nouvelles évaluations conduisant à des décisions correctrices³⁸. On pourrait par exemple rêver que des évaluations-bilans des connaissances acquises par les élèves soient utilisées pour faire évoluer les programmes. Comme le dit Philippe Perrenoud (1998), « les réformes sont décidées suivant une rationalité limitée, qui tient compte de multiples logiques, toutes intelligibles, sinon avouables, mais dans lesquelles la qualité et l'efficacité du système ne sont pas les seuls critères » (voir aussi Legrand, 1994). Pour A. Bouvier (1998), « il est fort délicat, pour ne pas dire impossible, de (...) concevoir le processus d'évaluation par rapport au processus de prise de décisions ». Une autre explication se trouve dans la large indépendance de la plupart des organes d'évaluation par rapport aux décideurs, non seulement pour l'évaluation elle-même, mais aussi dans le choix des sujets ; et lorsqu'il existe des commandes, elles interviennent plutôt en aval des décisions, pour étudier leur mise en œuvre, qu'en amont ou en accompagnement. En revanche, pour les inspections générales et pour la DPD, la publication des rapports est soumise à l'accord du Ministre, ce qui constitue une dépendance n'allant pas dans le sens de la transparence du compte rendu.

On peut en déduire que les destinataires les plus importants ne sont pas les responsables du système, mais les personnels pour lesquels il s'agit « d'amplifier le processus de professionnalisation et le mouvement vers des pratiques réflexives et coopératives » (Perrenoud, 1998). Il reste à en tirer les conséquences pour utiliser les évaluations en ce sens dans le cadre d'une régulation concertée.

2.4.2 Le système éducatif dans son ensemble

Les travaux concernant le système éducatif dans sa globalité sont rares. Les rapports annuels des inspections générales constituent une mine d'informations, mais pas une évaluation d'ensemble : celui de l'IGEN est une juxtaposition de travaux sectoriels ; celui de l'IGAEN a davantage un caractère de synthèse, mais seulement sur les points examinés au cours de l'année. Pour l'enseignement supérieur, le CNE adresse chaque année un rapport au Président de la République, comme les textes réglementaires lui en font obligation : c'est, pour l'essentiel, un rapport d'activité du Comité (Comité national d'évaluation, 1999). Le seul document portant sur l'ensemble du système est « l'état de l'école », publié chaque année par la DPD. Il s'agit d'un ensemble d'indicateurs commentés, au nombre de trente, sur les coûts, les activités, les résultats, plus que d'une évaluation proprement dite. Mais les évolutions temporelles qu'il indique en font un excellent outil de compte-rendu. Une question est cependant de savoir jusqu'à quel point il est connu et utilisé à l'extérieur du système éducatif, par exemple par les Parlementaires, et même à l'intérieur par les personnels de base.

À ce niveau, le but essentiel est bien de rendre compte, le progrès demandant des études plus ciblées. Les méthodes analytiques décrites ci-dessus pourraient cependant, comme l'écrit Jarousse (2000), « permettre un arbitrage entre différents modes d'organisations alternatifs en se référant à leurs effets aux plans pédagogiques et économiques ». Mais elles ne sont pas utilisées en ce sens.

Sur l'évaluation du système éducatif national, la faiblesse peut être exprimée de la manière suivante : il existe beaucoup d'évaluations concernant des aspects particuliers de la politique éducative, mais elles ne dessinent pas une évaluation globale ; l'état de l'école donne des indicateurs pertinents et les place dans une évolution temporelle, mais ne les relie pas à la politique éducative. Il y aurait place pour un rapport périodique d'évaluation de la politique éducative française examinant les conséquences des arbitrages effectués sur la réalisation d'objectifs déterminés. La loi d'orientation prévoyait d'ailleurs un rapport annuel au Conseil supérieur de l'éducation sur son exécution.

³⁸ En termes techniques, on distingue des évaluations ex-ante, à mi-parcours ou d'accompagnement, ex-post.

Le Haut Conseil pourrait prendre une initiative en ce sens. La question est de savoir si un tel rapport peut être réalisé par un organisme français, ne serait-ce que parce qu'aucun n'a vocation à couvrir l'ensemble du champ, ou s'il ne vaudrait pas mieux suivre l'exemple des examens de politiques éducatives nationales de l'OCDE, confiés à des experts d'autres pays sur la base d'un rapport préliminaire fourni par les autorités nationales.

2.5 En résumé

Dans les tableaux qui suivent sont résumées les réponses aux questions posées en 1.4 pour les divers types d'évaluation des quatre domaines envisagés.

Les élèves et les étudiants

	Évaluation intégrée à l'acte éducatif	Examens et autres modes de délivrance des diplômes	Évaluations de masse
Sur quoi porte l'évaluation ?	Compétences de l'élève, manifestées par des productions écrites et orales. Réfèrent : programmes, livrets scolaires à l'école primaire	Compétences des élèves et étudiants Réfèrent : programmes, mais critères peu explicites ;référentiels pour diplômes professionnels. Problèmes de représentativité des épreuves et de docimologie	Début CE2-6 ^e : compétences en français et mathématiques Début 2 ^{nde} : dans 4 disciplines. Réfèrent : compétences représentatives des programmes Épreuves : exercices regroupant des items
Résultats	Notes surtout ; corrections, appréciations, compétences acquises	Diplôme	Réussites aux items, avec divers regroupements possibles
Qui évalue ?	Enseignants de l'élève Conseil de classe	Enseignants extérieurs à l'établissement et/ou ceux de l'élève	Enseignants de l'élève, avec grille nationale
Comment ?	Évaluation en grande partie spontanée	Enseignement secondaire : réglementation nationale, organisation académique, contrôle continu au brevet Enseignement supérieur : compétence de l'établissement, parfois sans examen terminal	Épreuves nationales, QCM
Usages	Usages pédagogiques variés : diagnostiques, formatifs, sommatifs, probatoires, pronostiques (pour orientation). Compte rendu aux familles	Rendre compte des compétences. Brevet : pas de conséquences. Baccalauréat : droit d'entrée. Diplômes attestant une compétence professionnelle	Stratégies d'apprentissage et actions de remédiation. Compte rendu aux familles. Traitements statistiques locaux. Synthèses nationales, mais pas de bilan dans le temps
Conclusion	« Un point obscur » Forces : - compte-rendu aux élèves et parents. - intégration dans la régulation du travail des élèves. Faiblesse : intégration dans la régulation du travail des enseignants (impulsion, animation, suivi, évaluation des enseignants et surtout formation)	Forces : - acceptation sociale - réaction sur la formation Faiblesses : - qualité des épreuves et incertitudes sur la correction - incitation au bachotage Risque d'ébranlement de l'acceptation par l'évolution socio-économique	Forces : - nouveauté marquante, notoriété - CE2-6 ^e : acceptation, reconnaissance Faiblesses : - CE2-6 ^e : intégration à la pédagogie - 2 ^{nde} : évaluation pas toujours effectuée et peu exploitée - n'entre pas dans un processus de régulation du travail des enseignants
Études possibles	Compréhension par les familles des livrets d'évaluation et bulletins Usage des exercices d'évaluation à l'école primaire Place de l'évaluation des élèves dans la formation des enseignants et dans les pratiques d'inspection Liaison entre évaluation et orientation	Réalité de l'acceptation sociale Compétences effectivement évaluées par les sujets de baccalauréat Comparabilité des diplômes de même intitulé dans l'enseignement supérieur	Usages des évaluations de masse ; comment procéder à une relance ?

Les personnels

	Enseignement scolaire : Personnels enseignants et d'éducation	Enseignement scolaire : Personnels de direction	IATOS	Enseignement supérieur : Personnels enseignants
Sur quoi porte l'évaluation ?	Qualité du travail Pas de référent officiel	Qualité du travail Un référent vient d'être défini	Qualité du travail Pas de référent réellement utilisé	Qualité des résultats de recherche et, à un moindre degré, d'enseignement et d'animation Pas de référent explicite
Résultats	Notes Appréciation, rapport	Pas d'évaluation officielle depuis 1993	Note, appréciation	Qualification Promotion de classe
Qui évalue ?	Chefs d'établissements Inspecteurs	Inspecteurs Inspecteurs d'académie, Recteurs	Chefs d'établissements	Conseil national des universités Établissement
Comment ?	Évaluation spontanée Inspection	Pas d'évaluation officielle depuis 1993	Évaluation spontanée	Examen des travaux
Usages	Essentiellement avancement (changement d'échelon, de corps)	Avancement et mutations	Avancement	Avancement
Conclusion	Pas de véritable évaluation dans l'état actuel des choses Ne contribue guère au progrès			Progrès sur la diversité des critères et la clarté

Les unités

	Établissements d'enseignement supérieur	Établissements scolaires	Académies
Sur quoi porte l'évaluation ?	Réalisation des missions Fonctionnement des établissements	Conditions de fonctionnement et résultats	Fonctionnement des académies
Résultats	Rapport, avec réponse du responsable de l'établissement	Indicateurs, dont certains sont publics (résultats des lycées au baccalauréat)	Rapport, conclu par des recommandations
Qui évalue ?	Comité national d'évaluation (et autres instances)	DPD	Inspections générales
Comment ?	Auto-évaluation et évaluation externe	Calcul et mise à disposition d'indicateurs de pilotage	Constitution d'un dossier, visites, entretiens
Usages	Rendre compte publiquement Progresser ?	Rendre compte publiquement Peu d'usage interne	Par l'académie et l'administration centrale: A voir dans l'avenir
Conclusion	Forces : - existence, ouverture d'une voie cohérente avec l'autonomie des établissements - volonté d'articuler avec la contractualisation Faiblesses : - manque d'éléments de comparaison entre établissements - usages interne et externe pour progresser	Forces : - mise à disposition d'indicateurs - compte rendu public pour les lycées Faiblesses : - une évaluation ne se réduit pas à une collection d'indicateurs - indicateurs publics très partiels - autres indicateurs peu utilisés - pas intégrés dans une régulation Expériences d'évaluation d'établissements dans certaines académies	Encore difficile de porter un jugement sur une opération naissante Forces : - opération novatrice - peut être un appui à des relations contractuelles entre le ministère et les académies
Études possibles	Usages des rapports comme compte-rendu et aide au progrès	Obstacles à l'emploi des indicateurs et besoins réels	Usages par l'académie et par l'administration centrale

Le système éducatif national

	Évaluations thématiques	Le système global
Sur quoi porte l'évaluation ?	Thèmes choisis par les organismes d'évaluation ou déterminés par des commandes institutionnelles Réfèrent plus ou moins explicite, issu des textes officiels plus souvent que des niveaux de performances des élèves	Résultats, efficacité, efficience
Résultats	Rapports, notes, articles Destinataires mal identifiés, souvent les décideurs	« L'état de l'école » Rapport annuel du CNE
Qui évalue ?	IGEN (aspects pédagogiques, enseignement scolaire) IGAENR (aspects administratifs, financiers et organisationnels) CNE (enseignement supérieur) DPD (pratiques et innovations pédagogiques, enseignement scolaire) Équipes de recherche Cour des comptes	DPD CNE, pour l'enseignement supérieur
Comment ?	Enquêtes, observations, entretiens Bilan des acquis d'une population d'élèves (plus rare) Modélisation (plus rare)	Regroupement et analyse d'indicateurs et d'observations
Usages	Rendre compte Aide au progrès et à la décision difficile à cerner	Rendre compte
Conclusion	Rarement placé dans un processus clair de pilotage	Les évaluations concernant des thèmes particuliers ne dessinent pas une évaluation globale du système éducatif « L'état de l'école » donne des indicateurs et une évolution temporelle, mais ne les relie pas à la politique éducative
Études possibles		Mise en place d'un rapport périodique d'évaluation de la politique éducative française

3 - L'évaluation, ailleurs

3.1 L'évaluation en éducation : un cas particulier ?

Dans les divers pays, l'activité d'évaluation du système éducatif résulte de la convergence de deux courants :

- l'évaluation des élèves et des étudiants, spécifique à l'Éducation et enracinée dans des traditions nationales variées, qu'il s'agisse de l'évaluation en classe ou de la délivrance des diplômes ;
- le développement plus récent de l'évaluation des politiques et des organismes publics en général : ce courant dépasse l'Éducation et il existe des facteurs de convergence internationale, mais l'Éducation présente la spécificité de la couverture du territoire par un grand nombre d'établissements.

3.1.1 L'évaluation des politiques et organismes publics

Pour le Premier ministre Lionel Jospin, dans sa lettre du 7 septembre 1998 sur l'évaluation interministérielle des politiques publiques, « l'évaluation est une préoccupation plus récente en France que dans d'autres pays, notamment les pays anglo-saxons où elle s'inscrit dans une tradition déjà bien établie ». Dans son premier rapport, concernant l'année 1999, le Conseil national de l'évaluation retrace un historique à travers plusieurs contributions. L'évaluation est née aux Etats-Unis avec le *Planning Programming Budgeting System* (PPBS) et a été importée en Europe dans les années 1970 : en France, on a alors parlé de rationalisation des choix budgétaires (RCB). Après une période de remise en cause et d'incertitude, une relance sur de nouvelles bases s'est produite dans les années 1990 et l'évaluation a trouvé sa place face à la mesure des performances. On s'éloigne du contrôle, de la recherche de la faute et du blâme. L'enjeu devient de convaincre les agents que l'évaluation, c'est-à-dire la mesure mais aussi l'interprétation des résultats, fait partie d'une nouvelle culture de l'efficacité à laquelle ils doivent prendre part. Les incitations peuvent venir, selon les pays, des usagers, des Parlements, des Cours des comptes, des Gouvernements, mais aussi des organisations internationales : c'est ainsi que l'OCDE publie un guide des meilleures pratiques pour l'évaluation (1998) ; l'Union européenne incite à l'évaluation, en particulier à travers ses fonds structurels. De manière schématique, on peut distinguer trois phases : d'abord, l'évaluation est vécue comme une obligation, une contrainte, une charge supplémentaire ; puis, elle devient une aide à la conception et à la gestion ; enfin, elle est un acte politique dont les résultats font l'objet de débats publics. Il n'en demeure pas moins que son origine budgétaire continue à peser, que les ministères des finances jouent un rôle important dans sa mise en œuvre et que le slogan de Madame Thatcher « *value for money* » (en avoir pour son argent) reste d'actualité dans bien des cas. Aujourd'hui, le budget de la France est accompagné d'objectifs et d'indicateurs pour chaque département ministériel ; et la réforme de l'ordonnance de 1959 sur l'exécution de la loi de finances, actuellement à l'étude, devrait renforcer cette tendance.

3.1.2 Pour l'éducation, un paysage varié

Pour ce qui concerne les systèmes éducatifs, « l'évaluation et les indicateurs sont récemment devenus une préoccupation majeure dans l'Union européenne »³⁹. Selon A. Barblan, secrétaire général de la Conférence des recteurs européens, « partout en Europe, la question de l'évaluation est posée » (Conférence des présidents d'université, 1998). Dans l'enseignement supérieur, le Conseil de l'Europe a recommandé le 24 septembre 1998 l'introduction de méthodes d'assurance-qualité et une coopération dans ce domaine⁴⁰. L'évaluation des élèves fait référence à des normes dont la forme a évolué dans les dernières décennies (OCDE, 1995).

39 Lettre du réseau européen des responsables des politiques d'évaluation des systèmes éducatifs, no 9, décembre 2000.

40 The European Network for Quality Assurance Newsletter, no 1, juin 2000.

Une note figure dans le document de présentation du Haut Conseil (15 novembre 2000)⁴¹. Elle montre qu'il est difficile de dresser un tableau clair et complet, même en Europe, à cause de la variété des situations. « La forme et le statut des structures chargées de l'évaluation du système éducatif dans chaque pays sont fortement conditionnées par la sociologie et la tradition administratives, ainsi que par les objectifs assignés aux systèmes. Il est clair que les questions de centralisation, d'autonomie des établissements scolaires, etc., jouent un rôle dans les choix qui sont faits. De même, la logique de mise en concurrence des établissements et du rôle de l'État dans l'éducation ne sont pas sans influence sur ces décisions. (...) Dans beaucoup de pays, l'évaluation des acquis des élèves relève d'une institution différente de celle qui s'occupe de l'évaluation des établissements, voire des enseignements. La présence ou l'absence d'un inspectorat dans ces pays complique encore le paysage »⁴². Les structures chargées de l'évaluation peuvent être internes au ministère chargé de l'éducation, en être plus ou moins indépendantes, de type « agence », faire partie des universités, être de statut privé.

L'évaluation de l'enseignement supérieur (Heywood, 2000), où les universités ont une longue tradition d'autonomie, relève le plus souvent d'une structure unique, contrairement à ce qui se passe en France, et différente de celles qui s'occupent de l'enseignement scolaire. Elle est surtout axée sur les établissements et les formations qu'ils délivrent. Cette préoccupation est née aux Etats-Unis, il y a un siècle, dans un système peu coordonné, avec l'apparition du concept d'accréditation, « processus par lequel un organisme reconnaît un collège, une université, un programme d'études comme ayant satisfait certains standards de qualification » (Seldin, 1960), « reconnaissance par un organisme externe compétent, après évaluation, de la valeur des objectifs d'un établissement et des moyens qu'il prend pour les atteindre » (Sylvestre, 1998). L'évaluation peut aussi avoir, ou non, des conséquences sur le financement. Elle est effectuée à partir d'indicateurs de performance internes et externes, ou par un organisme extérieur, ou par auto-évaluation. La tendance va vers l'association de ces trois modes. Cependant, on doit distinguer entre des pays qui s'intéressent davantage au fonctionnement qualitatif et aux moyens mis en œuvre, et d'autres pays, en particulier anglo-saxons, qui cherchent à garantir les résultats, par « l'assurance-qualité » : en France, par exemple, on se borne à « l'habilitation » des formations, alors que l'accréditation irait avec une assurance de qualité... et qu'elle est sans doute nécessaire pour permettre une mobilité des étudiants.

L'évaluation des établissements scolaires s'est développée plus récemment dans de nombreux pays, en parallèle avec l'affirmation de l'établissement comme unité de décision autonome. On peut, pour cette « méso-évaluation », distinguer deux courants (Crahay, 1994) : analyser le fonctionnement et caractériser l'identité de l'établissement, ou rendre compte de son degré d'efficacité et chercher à l'expliquer par ses caractéristiques. La Commission européenne a développé un projet pilote sur l'auto-évaluation des établissements (MacBeath et al., 1999).

Les structures chargées de l'évaluation du système éducatif national ont aussi la mission de participer aux nombreuses opérations d'évaluation comparative impulsées par des organisations internationales comme l'OCDE, l'UNESCO, l'Union européenne (European Commission, 2001) ou encore l'IEA, *International Association for the Evaluation of Educational Achievement* (Orivel, 2000). On peut recommander au Haut Conseil de faire un bilan de ces travaux internationaux et de la part qu'y prend la France.

41 « Les structures chargées de l'évaluation des systèmes éducatifs chez nos principaux partenaires européens et étrangers ».

42 Les questions de l'autonomie des établissements, de la concurrence entre eux et de la place corrélatrice de l'évaluation, évoquées dans cette citation, sont étudiées pour les systèmes éducatifs étrangers par (Duru-Bellat et Meuret, 2001).

3.2 Trois études de cas

Pour être plus précis, nous examinerons le cas de trois pays européens, parmi les plus avancés sur l'évaluation, qui montrent trois modèles différents : l'Angleterre, l'Espagne, la Suède⁴³.

3.2.1 En Angleterre

L'évaluation des élèves est effectuée à travers des tests nationaux et des examens.

Depuis l'introduction d'un curriculum national autour de 1990, il existe des tests obligatoires en fin d'année scolaire à 7 ans, 11 ans et 14 ans : à 7 ans, ils portent sur l'anglais et les mathématiques et sont organisés et notés par les enseignants ; à 11 et 14 ans s'ajoutent les sciences et il s'agit d'examens formels. En outre, depuis 1998 ont été introduits des tests à l'âge de 5 ans pour permettre des comparaisons à l'entrée de l'école obligatoire. Les écoles ont aussi à leur disposition des tests facultatifs destinés aux élèves de 8, 9, 10 et 12 ans.

Il existe des examens à la fin de la scolarité obligatoire, à 16 ans (*General Certificate of Secondary Education*, GCSE), puis pour couronner l'enseignement secondaire (*General Certificate of Education*, GCE, *AS level* et *A level*), normalement à 17-18 ans. Contrairement à ce qui se passe en Europe continentale, les diplômes ne sont pas globaux ; ils se passent discipline par discipline : en moyenne, un élève présente huit ou neuf disciplines au GCSE et trois *A levels* au GCE. Les épreuves sont notées de A à G pour le GCSE, de A à E pour le GCE. Ces diplômes sont délivrés par des organismes certificateurs indépendants (*awarding bodies*) entre qui se partage la certification : trois fonctionnent aujourd'hui en Angleterre de manière concurrente.

Il existe également des diplômes professionnels. Traditionnellement, ces diplômes portaient le nom de l'organisme certificateur qui les délivrait et qui en était propriétaire. Cette situation subsiste, avec des organismes certificateurs beaucoup plus nombreux que pour les diplômes académiques, plusieurs centaines ; certains sont purement commerciaux, mais beaucoup sont liés aux corps professionnels. À partir de 1988, un effort de clarification et d'unification a été entrepris, avec la mise en place des *National Vocational Qualifications*, NVQs, définies en termes de compétences, c'est-à-dire de savoir-faire professionnels attachés à des métiers, et validées en situation de travail en fonction de critères de résultats. Ces qualifications sont construites à partir de recommandations d'organisations représentant les secteurs professionnels, en fait surtout les employeurs.

L'organisme responsable de ce système de curriculum et de diplômes (en anglais, *qualifications*) est une autorité publique indépendante, le QCA (*Qualifications and Curriculum Authority*).

Les résultats des tests et des examens servent aussi à évaluer les établissements. Ils sont récapitulés par établissement et comparés à des normes nationales. Pour les tests à 11 ans et les résultats au GCSE et GCE, les données sont publiées par le ministère de l'éducation et reprises par la presse sur le modèle des classements sportifs (*league tables*). Cependant, ce type d'évaluation, qui risque de négliger d'autres aspects de l'éducation moins mesurables et de ne fournir aux établissements aucune analyse de leur fonctionnement, est complété par des inspections d'établissements, mises en œuvre par une autre autorité publique indépendante, l'OFSTED (*Office for Standards in Education*).

43 Dans le numéro cité plus haut de la lettre du réseau européen, on trouvera aussi une présentation de la nouvelle organisation danoise, avec la création d'un Institut chargé de l'évaluation à tous les niveaux du système éducatif. Les trois études qui suivent ont bénéficié de communications de Nicholas Tate, Alejandro Tiana et Ulf Lundgren, membres du Haut Conseil, portant davantage sur l'enseignement scolaire que sur le supérieur. Pour ce dernier, le prochain rapport du CNE au Président de la République traitera de comparaisons internationales. En ce qui concerne l'Angleterre, nous avons aussi utilisé une note de l'OCDE (1999) ; voir encore (Broadfoot, 2000).

Les établissements sont inspectés régulièrement, à un intervalle qui a varié entre 4 et 6 ans. L'OFSTED engage pour cela des équipes, souvent formées d'anciens enseignants accompagnés par une personne extérieure à l'éducation. La procédure d'évaluation est définie par l'OFSTED. Elle fait appel à des observations et à des entretiens. Il s'agit d'évaluer tous les aspects du fonctionnement : performances des élèves, qualité de l'enseignement, discipline, utilisation des ressources, gestion et gouvernement de l'établissement. Sur les différents points, le rapport précise des niveaux d'efficacité. Il contient des recommandations destinées à l'établissement. Le rapport est public. Lorsque de graves déficiences sont relevées, l'établissement est mis en observation et réinspecté peu de temps après. Si les déficiences persistent, il peut être fermé et réouvert avec un nouveau personnel : ce cas est rare, mais il s'est déjà produit.

L'évaluation des universités est de la responsabilité d'une agence (*Quality Assurance Agency for Higher Education*) qui dépend de l'ensemble des institutions d'enseignement supérieur (QAA, 1998). Cette agence définit les exigences requises par les diplômés et évalue les universités, avec publication des résultats. L'évaluation a des conséquences sur le financement.

Les tests, les examens et les inspections fournissent aussi des résultats nationaux. Les gouvernements successifs ont fixé des objectifs concernant les niveaux de qualification atteints par les sortants du système, et plus récemment pour les résultats aux tests d'anglais et de mathématiques à 11 ans : pour que cela ait une signification, des efforts sont faits afin que les tests restent comparables d'une année à l'autre. Cependant, la question se pose de savoir si les progrès sont réels ou s'ils ne reflètent qu'un meilleur entraînement, tant pour les élèves aux tests et examens que pour les établissements relativement aux procédures d'inspection.

Une unité de recherche se trouve au ministère de l'éducation, portant le nom significatif de *Standards and Effectiveness Unit*. En outre, de nombreuses équipes de recherche en sciences de l'éducation collaborent à l'évaluation, sans pourtant que leur influence soit clairement établie.

3.2.2 En Espagne

Le système éducatif est organisé par une loi organique de 1990 (LOGSE), complétée par une autre loi organique de 1995 (LOPEG). Ces lois accordent une grande importance à l'évaluation, à laquelle elles fixent trois buts : l'amélioration de la qualité de l'éducation ; un appui pour la prise de décisions ; la contribution au débat social sur l'éducation. Elles s'appliquent dans le cadre d'une organisation décentralisée où les régions (*Comunidades*) autonomes jouissent de prérogatives d'importance variable mais toujours fortes, conduisant quasiment à un régime fédéral. Aussi faut-il distinguer entre l'évaluation générale du système national et celle des systèmes des diverses régions. Par exemple, il existe un curriculum national prévu par la loi, mais les régions peuvent le compléter. En outre, les établissements scolaires prévoient dans leur projet les programmes enseignés, qui sont encore précisés pour chaque classe⁴⁴. Ils sont entièrement responsables de l'évaluation des élèves, en tenant compte des objectifs de résultats fixés dans le curriculum national et régional : il n'y a aucun examen formel avant celui qui permet d'entrer dans les universités.

Les établissements sont évalués par l'Inspection de l'éducation, selon un plan systématique. L'Inspection collecte les résultats obtenus par les élèves dans leur scolarité. Le but est d'abord d'aider les établissements à progresser en y impulsant une évaluation interne tout en apportant un regard externe sur leur fonctionnement. Mais l'Inspection ne publie ni les rapports d'évaluation ni les résultats des élèves par établissement : les établissements et les régions peuvent décider de le faire. Il n'existe pas d'évaluation individuelle des enseignants, sauf dans des cas exceptionnels.

44 Des décrets récents, parus en décembre 2000, restreignent le rôle des régions et des établissements (Salamé Sala, 2001).

Les observations faites dans les établissements conduisent l'Inspection à des synthèses qui permettent une connaissance de l'état et de l'évolution du système éducatif sur divers thèmes, par exemple les résultats des élèves, les examens d'entrée dans les universités, les abandons scolaires, la formation professionnelle en entreprise, la mise en place de programmes de réforme, l'organisation de l'administration scolaire....

Au niveau national, un organisme indépendant, l'Institut national de la qualité et de l'évaluation (*Instituto nacional de calidad y evaluación*, INCE), est chargé d'évaluer l'enseignement du curriculum national, d'élaborer des systèmes d'évaluation, de réaliser des études et recherches, de faire des propositions aux autorités éducatives pour contribuer à améliorer la qualité de l'enseignement. Il est conçu comme un organisme fondé sur la collaboration de toutes les administrations éducatives et respectueux des compétences de chacune d'elles. Ses principales tâches sont actuellement : l'évaluation des connaissances des élèves à la fin de l'enseignement primaire et de l'enseignement secondaire obligatoire, à partir de tests sur échantillons ; l'élaboration d'un système d'indicateurs de l'éducation, analogues à ceux de l'OCDE ou de l'état de l'école en France, mais détaillés par région. En outre, il effectue des études spécifiques, par exemple sur l'enseignement des langues étrangères ou sur les fonctions de direction d'établissement. Il est aussi le correspondant espagnol pour les études internationales. Enfin, il a une activité de publication et de formation de spécialistes en évaluation.

Les régions ont la charge de l'évaluation de leur propre système éducatif et elles le font de diverses manières. Certaines ont créé des institutions d'évaluation indépendantes (Canaries, Catalogne, Valence...). D'autres, comme l'Andalousie, ont mis en place un organisme d'évaluation à l'intérieur de leur administration. D'autres encore, comme la Galice, la Navarre, le Pays basque, ont confié la mission à une direction de l'administration exerçant aussi d'autres fonctions.

Tout ce qui précède concerne l'enseignement scolaire. Pour l'enseignement supérieur, le ministère compétent est en train de développer un ensemble d'indicateurs et il existe des évaluations des universités volontaires. De même, les enseignants peuvent solliciter une évaluation pour favoriser leur carrière, par l'université pour leur activité d'enseignement, par une commission nationale pour la recherche.

3.2.3 En Suède

Le système éducatif, qui était organisé de manière centralisée pour l'enseignement scolaire, a été radicalement décentralisé à partir de 1991. Aujourd'hui, le niveau central se borne à fixer les programmes nationaux et des objectifs de résultats, laissant aux municipalités le soin d'organiser l'enseignement. En particulier, l'État n'apporte plus de financement spécifique à l'éducation, sauf à la marge pour des programmes spéciaux. Il y a donc eu un changement complet dans les responsabilités, passant du « centre » à la « périphérie », d'un pilotage centralisé par les règles et le financement à un pilotage par les objectifs et les résultats. La question était alors : comment garantir la réalisation des objectifs ? L'évaluation se place dans ce cadre.

On peut distinguer quatre niveaux de décision :

- le niveau politique national, Parlement et Gouvernement : il fixe les buts, les règles générales et le curriculum national qui précise les grands objectifs et les principes d'application (une vingtaine de pages pour l'ensemble de la scolarité obligatoire) ;
- une Agence nationale de l'éducation (*Skolverket*), indépendante du ministère et beaucoup plus importante que lui en nombre de personnes ; elle est l'organe de pilotage, avec les tâches suivantes : développer les programmes nationaux (« syllabi ») des disciplines, exprimés en termes d'objectifs pédagogiques terminaux et non de contenus (une centaine de pages au total pour la scolarité obligatoire) ; fournir aux enseignants des critères et des outils d'évaluation des élèves ; suivre le fonctionnement et l'évolution du système éducatif

en collectant l'information nécessaire ; évaluer le système éducatif relativement aux programmes et objectifs nationaux et s'assurer que chaque municipalité y satisfait ; rendre compte des résultats au Parlement, au Gouvernement, aux municipalités, aux enseignants et au public ;

- les municipalités, qui organisent l'enseignement pour répondre à la demande locale, fixent aux établissements des objectifs complémentaires aux objectifs nationaux, organisent l'évaluation des établissements scolaires, synthétisent les résultats obtenus et en rendent compte au public et à l'Agence ;
- les établissements : ils définissent leur curriculum propre, concrétisant le curriculum national, et assurent entièrement l'évaluation des élèves, ce qui leur permet de s'auto-évaluer et de rendre compte de leurs résultats à la municipalité.

On voit comment l'évaluation se trouve au centre de ce processus de régulation. La base est l'évaluation des connaissances des élèves. Pour y contribuer, l'Agence nationale de l'éducation met en place un programme de tests, portant sur le suédois, l'anglais et les mathématiques. Les tests sont construits par des institutions universitaires. Ils fournissent une référence nationale pour aider les enseignants dans la notation des élèves, qui comporte trois niveaux. Ils sont passés par tous les élèves de cinquième et neuvième années de l'enseignement obligatoire ; il existe aussi des tests de diagnostic pour chaque année de l'enseignement obligatoire et pour les cours de l'enseignement secondaire post-obligatoire. Une banque d'items de test est en construction pour diverses disciplines et sera disponible sur Internet.

En outre, l'Agence dispose d'un corps d'inspection. Le Gouvernement lui fixe chaque année un programme de travail comportant trois thèmes, à étudier éventuellement avec des spécialistes extérieurs. À cette occasion, les inspecteurs visitent des établissements. Il n'existe pas de processus formel d'évaluation des enseignants.

L'Agence publie un rapport annuel, « l'Image de l'École », comprenant des éléments statistiques, des indicateurs de qualité, des résultats d'évaluation, d'inspection et de recherche.

Pour l'enseignement supérieur, qui dépend d'un autre ministère, il existe aussi une agence qui évalue les universités.

3.3 Similitudes et différences

Ce paragraphe compare les trois cas précédents entre eux et avec celui de la France.

Pour l'évaluation des connaissances des élèves, il existe des tests nationaux passés à différents niveaux dans les disciplines considérées comme les plus déterminantes, par tous les élèves en Angleterre, Suède et France, par un échantillon en Espagne : dans ce dernier cas, l'évaluation ne vise pas directement une aide au progrès des élèves, mais le but est seulement de faire un bilan national et régional. La particularité de la France est que ces tests sont passés en début d'année, donnant à l'évaluation un caractère de diagnostic et pas simplement de constat. Dans l'enseignement scolaire, seules l'Angleterre et la France ont en outre des examens formels ; ailleurs, l'évaluation sommative est de l'entière responsabilité des établissements. Concernant l'évaluation des élèves, la situation française apparaît assez singulière : pour l'enseignement secondaire, elle associe une indépendance totale des enseignants dans leur classe à un examen externe global, le baccalauréat, dont le nombre d'épreuves n'a pas d'équivalent au monde ; pour le supérieur, les diplômes sont nationaux, mais leur délivrance ne fait l'objet d'aucune garantie de qualité.

Seule la France effectue une inspection individuelle des enseignants pour tenter de les évaluer.

Dans les trois pays étrangers étudiés a lieu une évaluation systématique des établissements scolaires, alors qu'en France ce n'est le cas que pour l'enseignement supérieur. La France est

aussi sur ce point un cas singulier, puisque, comme nous l'avons déjà dit, l'évaluation des établissements scolaires s'est développée dans un grand nombre de pays.

Les évaluations faites au niveau national, qui existent dans les quatre cas, sont difficiles à comparer. Il semble cependant qu'on puisse dire que, tant pour leur abondance que pour leur variété, la France soit bien placée.

Mais c'est surtout sur les buts et les usages de l'évaluation qu'apparaissent les différences. Le très efficace modèle britannique est centré sur l'établissement, évalué à la fois par les performances des élèves aux tests nationaux et aux examens, et par des inspections ; la régulation du système repose sur la publication des résultats de ces deux types d'évaluation et sur la concurrence entre établissements : on peut dire que l'on est dans un modèle de marché. En Suède, on a une organisation multi-niveaux, avec une régulation par des objectifs et des programmes, en « descendant », d'une part, par l'évaluation d'autre part, en « remontant » : on peut parler de modèle systémique. En Espagne, dans un pays très décentralisé, à la limite du fédéralisme, sur les régions mais aussi sur les établissements, il s'agit de mettre des informations à la disposition des administrations éducatives tout en respectant leur liberté : le modèle vise l'aide à la décision. On peut d'ailleurs ajouter que, dans les pays fédéraux, la mise en place d'un système d'évaluation apparaît délicate : c'est par exemple le cas en Allemagne.

En France, le modèle d'évaluation n'est pas facile à qualifier car il apparaît très multiforme, voire foisonnant : à cause de sa richesse sans doute, mais ce n'est pas la seule raison. Par rapport aux pays anglo-saxons ou nordiques, on se méfie davantage de la transparence et on est plus critique sur les indicateurs quantitatifs. Tous ces éléments conduisent à une certaine frilosité et à de multiples précautions, qui rendent plus difficiles les conceptions systémiques. La concurrence ouverte est mal acceptée, et d'ailleurs la concurrence entre établissements est rejetée par la Loi. Enfin, la répartition des pouvoirs est souvent imprécise et controversée, comme le montrent par exemple les compétences partagées dans le cadre de la décentralisation politique et les incertitudes sur l'autonomie des établissements.

4 - Forces, faiblesses et marge de progrès

Ce chapitre de conclusion tente, à partir des analyses sectorielles du chapitre 2 et des comparaisons internationales du chapitre 3, de porter un jugement de valeur global sur l'évaluation de l'École en France, ou tout au moins de dégager des pistes de discussion.

Pour cela, on peut prendre comme références :

- une comparaison dans le temps et dans l'espace ;
- les deux buts mis en évidence au chapitre 1 : rendre compte et aider au progrès.

4.1 Comparer dans le temps et dans l'espace

Depuis une vingtaine d'années se sont produites des avancées considérables :

- avec l'inscription de l'évaluation dans des textes législatifs, en 1984 pour l'enseignement supérieur puis en 1989 pour le système éducatif en général ;
- avec la mise en place du CNE en 1985 ;
- avec la création en 1987 de la direction de l'évaluation et de la prospective (DEP), aujourd'hui incluse dans la DPD, qui a mis des outils à la disposition des acteurs ; déjà, en 1985, était apparu le service de la prévision, des statistiques et de l'évaluation (SPRESE), succédant au service de l'informatique de gestion et des statistiques (SIGES) qui avait fait avancer la notion de « tableau de bord » ;
- avec un élargissement des missions des inspections générales, et notamment de l'IGEN, qui, à partir de 1982 et plus encore de 1989, se sont étendues à des évaluations thématiques.

Tout cela a conduit à un progrès dans la qualité du débat public sur l'éducation et a contribué à la pénétration d'une culture de l'évaluation (Thélot, 1993) dans les services académiques et chez les personnels de direction. La situation est plus disparate parmi les enseignants, avec une implication forte d'une minorité et une moindre crainte chez la plupart des autres ; il existe cependant une autre minorité qui considère que, non seulement l'évaluation est une perte de temps, mais qu'elle est dangereuse car elle dépossède les enseignants de leurs prérogatives au profit des forces du marché.

Aujourd'hui, l'Éducation nationale française est dans une situation favorable par rapport à la plupart des systèmes éducatifs étrangers pour la variété et la qualité des évaluations : c'est la leçon qui peut être tirée du chapitre 3. Parmi les forces, on peut relever la qualité de l'appareil statistique sur lequel peut s'appuyer l'évaluation, ainsi que l'organisation des structures qui en sont chargées. Il est d'ailleurs significatif que le réseau européen des responsables des politiques d'évaluation des systèmes scolaires soit animé par la France ; la situation est sans doute moins favorable pour l'enseignement supérieur, où le réseau européen fait référence à « l'assurance-qualité » qui n'est pas précisément une notion d'origine française.

La position de l'Éducation nationale est également favorable par rapport aux autres administrations françaises : c'est la perception qui règne au Conseil national de l'évaluation et au Comité du coût et du rendement des services publics. Cette appréciation flatteuse s'accompagne cependant du regret que, comme souvent, l'Éducation nationale semble être réticente à une ouverture interministérielle.

Lorsque l'on compare dans le temps, en se référant au passé, et dans l'espace, en se référant à d'autres pays, on peut donc parler de forces.

4.2 Se référer aux buts poursuivis

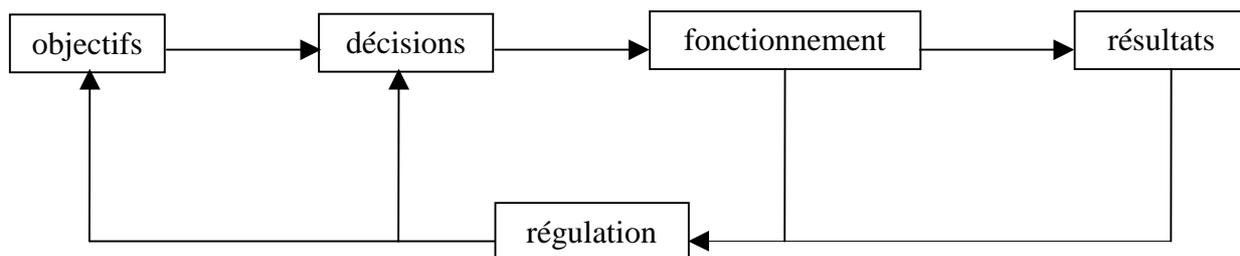
Relativement aux buts poursuivis, en revanche, nous avons vu au chapitre 2 qu'il existe une grande marge de progrès, autrement dit certaines faiblesses. Lorsque l'on passe en revue les divers types d'évaluation, l'intensité des efforts et la richesse des résultats contrastent avec une impression de dispersion et une certaine faiblesse dans les usages. Il faudrait *tendre vers une meilleure correspondance entre les efforts consacrés à la réalisation des évaluations et à la concrétisation de leurs usages*.

On devrait d'ailleurs se poser la question de l'efficacité et du rendement de l'évaluation. Nous ne l'avons pas fait dans ce rapport, à cause de la difficulté de cerner les coûts : cette question, abordée par la DPD à l'occasion de notre rapport, pourrait donner lieu à une étude du Haut Conseil.

Globalement, sur les deux buts, rendre compte et aider au progrès, c'est le premier qui est le mieux atteint. Ici, *une faiblesse est que les indicateurs et les référents portent surtout sur le fonctionnement et les résultats internes au système éducatif, mais rarement sur les résultats après la sortie du système ou vis-à-vis de l'extérieur* : l'économie et la société. D'autre part, il n'est pas certain qu'en dehors du système éducatif, par exemple chez les élus, chez les journalistes, dans les milieux économiques, dans le grand public, les destinataires soient réellement atteints. C'est que l'information est difficile dans le domaine de l'éducation, qui ne s'accommode pas d'idées simplistes. On peut alors recommander une étude de la notoriété et de l'impact des publications destinées à l'extérieur, par exemple « l'état de l'école » ; cela permettrait ensuite de mieux savoir quoi communiquer et comment, peut-être en passant par des professionnels extérieurs à l'Éducation nationale qui feraient un examen critique de ce qui peut être publié et par quels canaux.

Sur le second but, comme nous l'avons déjà indiqué au chapitre 1, l'évaluation n'entraîne pas de progrès à elle seule. Or, comme l'a dit un de nos interlocuteurs, « elle a été portée par des techniciens, heureux de faire progresser le débat public, pas par des responsables éducatifs pour faire progresser le système ». *L'enjeu est que l'évaluation conduise à des décisions des acteurs qui permettent des progrès. Ce n'est le cas que lorsqu'elle est intégrée dans un processus de pilotage* (Bouvier, 1998) où sont fixés des objectifs à atteindre, sont prises des décisions pour y parvenir, est exercée une régulation pour tenir compte des résultats obtenus et du fonctionnement qui y conduit : l'évaluation entre dans la régulation ; le rapport annexé à la loi d'orientation lui associe d'ailleurs le terme de « régulation permanente ». De manière voisine, le Conseil national de l'évaluation (2000) « conseille aux administrations de penser les politiques publiques en termes de système c'est-à-dire de passer des objectifs généraux à des objectifs concrets et spécifiques, mesurés à travers des systèmes d'informations, alimentés par des indicateurs qui montrent si les objectifs sont réalisés et complétés par une évaluation qui analyse les causes de l'écart entre les objectifs et les résultats ».

Un pilotage raisonné



4.3 Divers modes de pilotage

G. de Landsheere (1994) distingue le pilotage « administratif ou monitoring de conformité », qui fait une place importante au contrôle de l'application de normes, le pilotage « par le rendement scolaire » et le pilotage « formatif, axé sur les individus » : le premier est traditionnel dans l'administration ; le second entre dans un modèle de gestion « technologique » qui peut conduire à confondre les objectifs avec les indicateurs ; le troisième fait référence à un modèle « humaniste ».

Dans un pilotage administratif, seuls les dirigeants fixent les objectifs et les normes de mise en œuvre. Il n'y a pas d'appel à une évaluation, plutôt un contrôle qui ne porte pas sur les résultats mais seulement sur le fonctionnement, vérifiant si les actions des agents sont conformes aux normes et prenant des sanctions dans le cas contraire. Ce type de pilotage s'applique mal à un système éducatif où beaucoup de décisions sont prises au quotidien dans la classe. Il conduit alors à un compromis où seules sont vérifiées des règles bureaucratiques ne portant pas sur les décisions réellement importantes. En outre, dans un tel pilotage, les objectifs et les décisions des dirigeants ne sont pas remises en cause : cela peut se justifier dans un monde immobile mais compromet l'adaptation à un contexte changeant.

Dans un pilotage par le rendement, les dirigeants fixent un cadre d'objectifs et de décisions qui conditionne une mise en œuvre dont sont responsables les agents. L'évaluation porte sur les résultats mais pas sur le fonctionnement, considéré comme une « boîte noire ». Elle est destinée aux dirigeants pour lesquels elle est une aide à la décision pour adapter le cadre. Les agents ont alors l'impression d'exercer une responsabilité sans réel pouvoir. Ils considèrent qu'ils n'ont pas les moyens d'appliquer les décisions et que l'évaluation est une sanction qui les obligera à transformer leurs pratiques sans qu'ils aient pu faire valoir les difficultés qu'ils rencontrent ; l'évaluation peut aussi les amener à se considérer comme inférieurs à la tâche ou à chercher des boucs émissaires. Tout cela n'est pas favorable à ce que leurs décisions aillent dans le même sens que celles des dirigeants ; c'est ce qui conduit à des « effets pervers » des décisions hiérarchiques.

Une variante peut faire passer une partie importante du pouvoir aux usagers, qui prennent des décisions conformes à leurs objectifs individuels, en particulier dans le choix des établissements (Duru-Bellat et Meuret, 2001). On peut parler de pilotage par le marché, le rôle des dirigeants étant de fixer le cadre de la concurrence, en particulier en publiant les résultats des établissements : l'évaluation sert alors à rendre compte aux usagers. Dans cette conception, la concurrence est considérée comme le principal facteur de progrès. Mais nous avons vu (§ 1.1) que le rapport annexé à la loi d'orientation rejette cette idée. C'est qu'en matière d'éducation, on peut lui faire un certain nombre d'objections. L'une d'elles est que l'optimum global, en admettant qu'on sache le définir⁴⁵, ne résulte pas nécessairement de la conjonction d'optimums partiels. Par exemple, si tous les établissements délivrent le même type de formations, celles qui attirent les élèves, la pertinence globale vis-à-vis du marché de l'emploi ne sera pas bonne. Ou si chacun veut augmenter son nombre d'élèves, cela peut conduire à une augmentation de l'ensemble des effectifs s'il y a pénurie d'éducation ; cela a pu être un objectif global dans le passé, mais aujourd'hui, dans notre pays, on ne cherche plus à accroître le nombre d'élèves ou à allonger encore la durée de scolarisation ; on vise plutôt à améliorer l'équité et la qualité. Or la concurrence entre établissements peut provoquer une ségrégation des populations et les résultats de recherche⁴⁶ montrent que cette situation accroît les écarts, contre l'équité, et ne mène pas à la meilleure qualité globale. En outre, la concurrence conduit à viser des résultats à court terme (par exemple, avoir le plus possible d'enseignants disponibles plutôt que de les envoyer en formation continue), alors que l'éducation demande des investissements, de tous ordres, à long terme. Pour les services publics en général, le Conseil national de l'évaluation

45 Sur la difficulté de définir les finalités de l'éducation, voir (Perrenoud, 1998).

46 Voir par exemple (Duru-Bellat et Mingat, 1997).

(2000, contribution de S. Trosa) montre, au plan international, les limites de ce mode de pilotage.

On peut faire le même genre d'objection à un autre mode de pilotage, « corporatiste », où le pouvoir se trouverait essentiellement entre les mains des agents du système. L'évaluation sert alors à leur fournir une information pour influencer leurs décisions. Pour être efficace, elle doit porter sur les performances de leurs élèves et non pas sur le système en général. Mais, là encore, il n'y a aucune raison pour que les décisions locales conduisent à un optimum global : par exemple, elles ont toute chance d'inciter à une sélection des élèves.

Aucun type de pilotage où le pouvoir appartient à l'un ou l'autre des principaux acteurs – les dirigeants, les usagers, les agents – n'est donc favorable au progrès. Il faut parvenir à un pilotage concerté qui fasse partager un sens, un « pilotage formatif » dans un « modèle humaniste » selon les expressions de G. de Landsheere⁴⁷.

Vis-à-vis des usagers, le pilotage concerté repose sur la responsabilité du système éducatif qui consiste, comme nous l'avons dit au chapitre 1, à fixer des objectifs après un débat et à apporter des corrections en fonction des résultats obtenus, pour progresser vers les objectifs. Cette responsabilité s'exerce à tous les niveaux, de la classe au système national et, à chacun d'eux, ce sont les dirigeants qui en sont les garants.

Cependant, la correction requiert des décisions des acteurs que sont, à chaque niveau, les agents et les dirigeants : le mot « agent » ne doit donc pas prendre une connotation de passivité ; le dirigeant d'un niveau devient d'ailleurs agent au niveau suivant : par exemple, dans une classe, les agents sont les élèves et les dirigeants les enseignants qui deviennent agents au niveau de l'établissement. La correction est permise par une régulation « permanente » car les décisions des agents de base dans leur travail sont quotidiennes. Il ne s'agit plus de fixer des normes, comme dans le pilotage administratif, ni seulement de confronter des résultats à des objectifs, comme dans un pilotage par le rendement, mais de rendre cohérents les objectifs et les décisions des uns et des autres en créant du sens et en développant un diagnostic partagé et une culture commune.

Une telle régulation est donc un processus complexe qui fait un pont entre les agents et les dirigeants. Nous en avons ailleurs (Pair, 1998) détaillé les composantes :

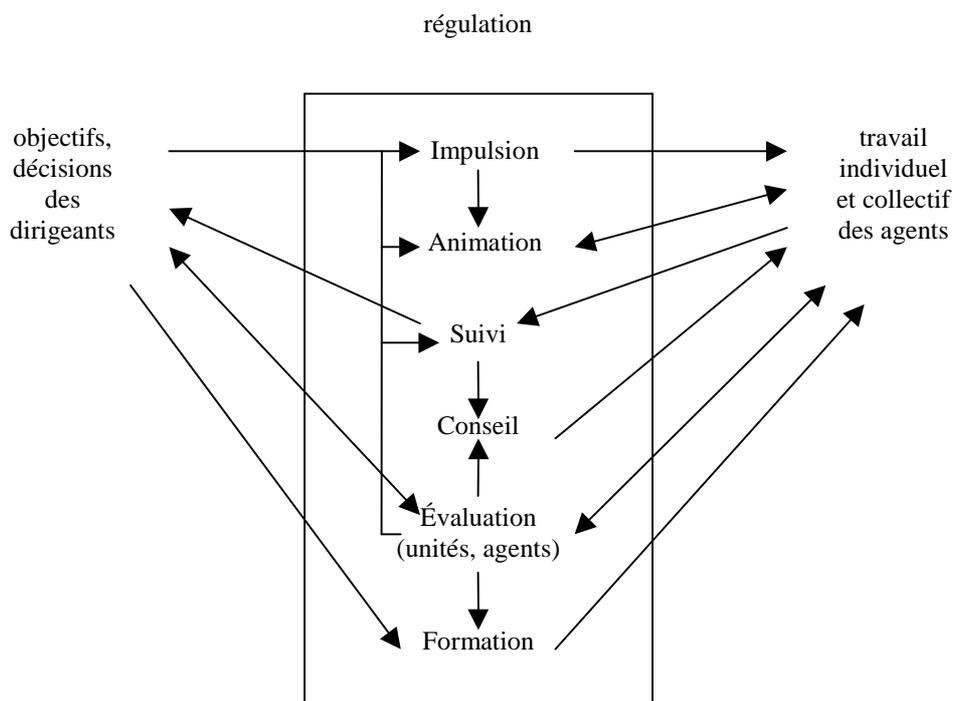
- l'impulsion par les dirigeants, dont le but est l'intégration des objectifs par les agents ;
- le suivi dont le but est la connaissance et la reconnaissance par les dirigeants des actions menées, des difficultés rencontrées, des solutions adoptées par les agents ;
- l'animation qui organise l'échange entre les agents ;
- le conseil que peut donner un dirigeant comme personne extérieure ;
- la formation pour améliorer les compétences professionnelles des agents et leur proposer des outils ;
- et l'évaluation : celle des unités qui mesure leurs résultats et analyse leur fonctionnement ; celle des agents qui, avec eux, analyse leur activité pour favoriser une progression.

4.4 L'évaluation dans un pilotage concerté

D'abord, l'évaluation nourrit le débat, puis elle juge des résultats, ce qui conduit à un nouveau débat : il s'agit de son but de rendre compte. D'autre part, elle entre dans la régulation qui permet le progrès, en nourrissant l'impulsion, le suivi, l'animation, le conseil, la formation. Elle joue donc un rôle fondamental, surtout lorsque la responsabilité est clairement répartie entre les divers niveaux du système, comme c'est par exemple le cas en Suède. Réciproquement, elle bénéficie de l'existence de la régulation : il existe pour elle une impulsion, un suivi, une animation, une activité de conseil, une formation... et une évaluation.

⁴⁷ Voir aussi (Demailly, 2001), p. 135-138.

Le pilotage concerté



C'est parce que l'évaluation est trop peu intégrée dans une régulation au sens précédent qu'elle est trop peu facteur de progrès. Par exemple, il en est ainsi pour les évaluations de masse (§ 2.1.3), à deux niveaux : celui de la pédagogie, avec les élèves ; celui d'une régulation globale de l'opération, qui ne laisse pas les enseignants seuls devant les usages de l'évaluation. Cela explique aussi que les types d'évaluation nouveaux aient du mal à s'imposer par rapport aux plus traditionnels qui pourtant sont souvent moins clairs, davantage de l'ordre du spontané ou de l'implicite (§ 2.1.4) ; alors qu'il faudrait allier la solidité de la tradition et le caractère plus rigoureux des nouveautés. En revanche, les modalités qui viennent d'être définies pour l'évaluation des personnels de direction (§ 2.2) vont dans le sens d'une telle régulation : il reste à les mettre en œuvre dans cet esprit. Il en est de même pour les travaux de la DPD qui permettent, à travers des enquêtes, de mieux connaître les représentations que les enseignants se font de leur métier, les ressources qu'ils utilisent et les modalités de fonctionnement des dispositifs innovants (§ 2.4.1) : encore faudrait-il que les résultats de ces travaux soient diffusés aux formateurs et discutés avec les acteurs de terrain.

On peut alors dégager les caractéristiques d'un système d'évaluation intégré dans la régulation d'un pilotage concerté, à partir des questions posées en 1.4.

1. Sur quoi porte l'évaluation ? Elle doit porter à la fois sur les résultats, pour rendre compte et servir à l'impulsion, et sur le fonctionnement pour nourrir le suivi, l'animation, le conseil, la formation, et ainsi aider au progrès. Et en conséquence associer aspects quantitatifs et qualitatifs.
2. Qui évalue ? L'évaluation doit être « participative », en associant évaluation externe, pour garantir la qualité du compte rendu, et auto-évaluation pour avoir un caractère « formatif » qui augmente ses chances d'influer sur les décisions prises par les agents⁴⁸.

48 « L'approche participative : un outil au service de l'amélioration des politiques publiques », formule de S. Trosa (Conseil national de l'évaluation, 2000).

3. Quels sont ses destinataires ? Dans un pilotage concerté, l'évaluation ne peut être destinée seulement aux dirigeants, ou essentiellement aux usagers, mais à tous les acteurs, même s'il existe diverses sortes de résultats pour permettre de mieux les atteindre dans leur diversité. Alors qu'aujourd'hui, par exemple, les évaluations concernant l'ensemble du système ne s'adressent pas aux acteurs de base, dont elles proviennent pourtant... tout en ne servant que médiocrement aux décisions des dirigeants.
4. Quels usages ? L'évaluation est utilisée dans les autres composantes de la régulation qui, réciproquement, la confortent.
5. Quelles modalités ? Comme la régulation est exercée à tous les niveaux du système éducatif, tous les niveaux participent à l'évaluation : il est essentiel qu'elle soit impulsée de manière continue au niveau national et que l'administration centrale soit organisée pour en utiliser les résultats ; mais elle ne peut pas être uniquement conçue comme une opération nationale ; en particulier, elle doit occuper à l'échelon académique une place plus importante qu'aujourd'hui. D'autre part, la régulation est un processus unique : l'évaluation ne doit pas être une collection d'opérations isolées, comme c'est largement le cas aujourd'hui. Par exemple, les évaluations de masse sont surtout des micro-évaluations, axées sur les relations entre élèves et enseignants, et il n'y a pas de communication organisée avec les méso-évaluations, celles des établissements ; il n'y en a pas non plus avec une évaluation du système sous forme de bilan des connaissances acquises et de leur évolution dans le temps ; de même, aucun lien n'existe entre l'évaluation des personnels et celle des établissements ; les évaluations d'établissements universitaires par le CNE conduisent à des monographies isolées et sans grand rapport avec la politique globale ; il n'y a guère d'interaction entre les nouveaux types d'évaluation et les plus traditionnels... Peut-être pour éviter d'effaroucher, on est loin de la situation décrite par C. Thélot (1993) : « toute l'année, les enseignants évaluent leurs élèves (...), ces derniers passent des examens, (qui évaluent) indirectement leurs enseignants et (...) le système dans son entier ».

Ces recommandations dessinent un idéal auquel confronter les évaluations. Si on devait leur chercher un point commun, ce serait que l'évaluation doit toujours davantage entrer dans une perspective systémique, celle de la régulation du système éducatif avec ses trois types d'acteurs et ses différents niveaux d'action : classe, établissement, académie, national.

Enfin, l'évaluation ne peut progresser seule. Cela éclaire la question de l'indépendance des organismes d'évaluation. Pour rendre compte de manière crédible, une indépendance est souhaitable dans le choix des objets évalués, les méthodes d'évaluation, la publicité et l'usage des résultats. Mais alors, il est difficile à l'évaluation d'entrer dans une régulation, surtout dans le cadre d'un pilotage concerté. Il y a donc un compromis à trouver qui tienne compte du caractère plus ou moins crucial des éléments de l'indépendance : sur les méthodes et la publicité des résultats, plus que sur le choix des objets et l'usage des résultats. La situation française n'est pas si mauvaise, ou plutôt elle ne le serait pas si l'articulation était davantage pensée entre les divers organismes qui interviennent dans l'évaluation et dans ses usages, et les rôles mieux partagés entre eux en fonction de leurs liens plus ou moins forts avec les dirigeants administratifs et politiques. Parmi ces organismes figurent en particulier les directions ministérielles – qui font partie de l'appareil de décision – et les corps d'inspection, généraux – indépendants de l'administratif sinon du politique – et régionaux – dépendant des recteurs et jouant un rôle de premier plan dans la régulation.

Dans cette articulation, le Haut Conseil a un rôle à jouer : peut-il être un garant d'indépendance du rendre compte, grâce à la multiplicité des origines de ses membres, pour que d'autres organismes puissent, par une plus grande intégration au système, contribuer au progrès ?

Bibliographie

- Altet, M., Bru, M., Bressoux, P. & Lambert, C. (1996). Étude exploratoire des pratiques d'enseignement en classe de CE2 : deuxième phase. *Les Dossiers d'Éducation et Formations*, no 70.
- Andrieux, V. & Colmant, M. (2000). Les élèves en début de CE2, évaluation de septembre 1999. *Note d'information 00.01*. Paris : Ministère de l'Éducation nationale.
- Andrieux, V., Dupé, C. & Robin, I. (2000). Les élèves en début de sixième, évaluation de septembre 1999. *Note d'information 00.02*. Paris : Ministère de l'Éducation nationale.
- Baktavatsalou, R. & Pons, C. (1998). Comparaison des performances en lecture-compréhension des élèves en fin de CM2 à dix années d'intervalle (1987-1997). *Note d'information 98.39*. Paris : Ministère de l'Éducation nationale.
- Baluteau, F. (1993). *Le conseil de classe : "peut mieux faire !"*. Hachette Éducation, Paris.
- Barbier, J.-M. (1990). *L'évaluation en formation*, 2^e édition. Paris : PUF.
- Barré, R. (dir.) (2000). *Science et Technologie : Indicateurs, édition 2000*, Rapport de l'Observatoire des sciences et des techniques. Paris : Economica.
- Blanchet, R. (dir.) (1999). *La revalorisation du rôle des chefs d'établissement de l'enseignement secondaire*. Paris : Ministère de l'Éducation nationale.
- Bouvier, A. (1998). Évaluation ou pilotage au sein des organisations de formation ? in Pelletier, G. (dir.), *L'évaluation institutionnelle de l'éducation*. Montréal : Éditions de l'Affides.
- Broadfoot, P. (2000). Un nouveau mode de régulation dans un système décentralisé : l'État évaluateur. *Revue française de pédagogie*, no 130.
- Cahiers de l'académie de Lille (1997). *L'audit des établissements dans l'académie de Lille*. Lille : Rectorat.
- Casteran, F. & Oeuvarard, F. (1990). Les pratiques d'analyse et l'évaluation des établissements dans trois académies. *Éducation et Formations*, no 22.
- Champy, P. & Étévé, C. (2000). *Dictionnaire encyclopédique de l'éducation et de la formation*, 2^e édition. Paris : Nathan.
- Chauvet, J. (1991). L'opinion des enseignants sur l'évaluation CE2-6^{ème}. *Note d'information 91.49*. Paris : Ministère de l'Éducation nationale.
- Comité national d'évaluation (1999). *Enseignement supérieur : autonomie, comparaison, harmonisation*, Rapport au Président de la République 1995-1999. Paris : La documentation française.
- Comité national d'évaluation (2000 a). *Mémento de l'expert*. Paris : CNE.
- Comité national d'évaluation (2000 b). *Un nouveau contrat quadriennal : les orientations du Comité national d'évaluation*. Paris : CNE.
- Comité national d'évaluation (2001 a). *Guide de l'évaluation des universités*. Paris : CNE.
- Comité national d'évaluation (2001 b). *Les Instituts universitaires de formation des maîtres au tournant de leur première décennie : panorama et perspectives*. Paris : CNE.
- Comité national d'évaluation (2001 c). *Bienvenue sur le serveur internet du Comité national d'évaluation*. Paris : CNE.
- Conférence des présidents d'université (1998). *L'évaluation : pour quoi faire ?*. Paris : CPU.
- Conseil national de l'évaluation (2000). *L'évaluation au service de l'avenir*, Rapport annuel 1999. Paris : La documentation française.
- Conseil scientifique de l'évaluation (1996). *Petit guide de l'évaluation des politiques publiques*. Paris : La documentation française.
- Crahay, M. (1994). *Évaluation et analyse des établissements de formation : problématique et méthodologie*. Bruxelles : De Boeck-Wesmael.
- Crozier, M. (1963). *Le phénomène bureaucratique*. Paris : Seuil.
- Danner, M., Duru-Bellat, M., Le Bastard, S. & Suchaut, B. (2001). Évaluation d'une innovation pédagogique au lycée : l'aide individualisée en seconde. *Les Notes de l'IREDU*, no 01/4 et *Éducation et Formations*, no 58.

- Demilly, L. (1999). L'évaluation et l'auto-évaluation des établissements scolaires : un enjeu collectif – Le cas des audits des établissements scolaires dans l'académie de Lille. *Politiques et Management public*, 17, no 1.
- Demilly, L. (2001). Management et évaluation des établissements, in Van Zanten, A. (dir.), *L'école : l'état des savoirs*. Paris : La découverte.
- Dessus, C. & Gossot, B. (1993). L'opinion des enseignants sur l'évaluation en classe de seconde à la rentrée 1992. *Note d'information 93.24*. Paris : Ministère de l'Éducation nationale.
- Dessus, N., Jouvanceau, P. & Murat, F. (1996). Les connaissances des élèves en fin de troisième générale : évolution 1984-1990-1995. *Note d'information 96.36*. Paris : Ministère de l'Éducation nationale.
- Dubet, F. (dir.) (1999). *Le collège de l'an 2000*. Paris : Ministère de l'éducation nationale, de la recherche et de la technologie.
- Duru-Bellat, M., Jarousse J.-P., Leroy-Audoin, C. & Michaut, C. (2000). Écueils et enjeux de l'évaluation de l'enseignement supérieur universitaire. *Administration et Éducation*, no 2/2000.
- Duru-Bellat, M. & Meuret, D. (2001). Nouvelles formes de régulation dans les systèmes éducatifs étrangers : autonomie et choix des établissements scolaires. *Revue française de pédagogie*, no 135.
- Duru-Bellat, M. & Mingat, A. (1993). *Pour une approche analytique du fonctionnement du système éducatif*. Paris : PUF.
- Duru-Bellat, M. & Mingat, A. (1997). La constitution de classes de niveau par les collèges et les incidences sur les progressions et les carrières des élèves. *Revue française de sociologie*, 38, no 4.
- Émin, J.-C., Julienne, J. & Rondeau, M.-C. (1998). Trois indicateurs de performance des lycées à partir d'une exploitation des résultats au baccalauréat général, technologique et professionnel, session 1997. *Note d'information 98.07*. Paris : Ministère de l'Éducation nationale.
- European Commission (2001). *European report on the quality of school education : sixteen quality indicators*. Luxembourg : Commission européenne.
- Férole, J. (1994). L'opinion des enseignants sur l'évaluation nationale en classe de sixième à la rentrée 1993. *Note d'information 94.13*. Paris : Ministère de l'Éducation nationale.
- Ferrier, J. (rapporteur) (1996). *Évaluation et notation des personnels enseignants du premier degré*. Paris : IGEN.
- Flieller, A. (1997). Comparaison des compétences cognitives des élèves de 7 ans à vingt ans d'intervalle. *Note d'information 97.12*. Paris : Ministère de l'Éducation nationale.
- Flieller, A. (2001). Problèmes et stratégies dans l'explication de l'effet Flynn, in Huteau, M. (ed.), *Les figures de l'intelligence*. Paris : EAP.
- Grisay, A. (1993). Le fonctionnement des collèges et ses effets sur les élèves de sixième et cinquième. *Les Dossiers d'Éducation et Formations*, no 32.
- Heywood, J. (2000). *Assessment in higher education : student learning, teaching, programmes and institutions*. Londres et Philadelphie : Jessica Kingsley Publishers.
- IGAENR (2000 a). *Les suites données aux rapports de l'IGAENR : essai d'évaluation 1995-1999*, Rapport 00-0049. Paris : Ministère de l'Éducation nationale et Ministère de la Recherche.
- IGAENR (2000 b). *Rapport général 2000*. Paris : La documentation française.
- IGEN (1998). L'évaluation des élèves en classe, in *Rapport de l'inspection générale de l'éducation nationale*. Paris : La documentation française.
- IGEN (2000). *Rapport de l'inspection générale de l'éducation nationale*. Paris : La documentation française.
- IGEN & IGAENR (2000). *Évaluation de l'enseignement dans l'académie de Limoges et Évaluation de l'enseignement dans l'académie de Rennes*, Rapports 00-0070 et 00-0071. Paris : Ministère de l'Éducation nationale.
- Jarousse, J.-P. (2000). Comment évaluer les systèmes éducatifs, in Paul, J.-J. (dir.), *Administrer, gérer, évaluer les systèmes éducatifs*. Paris : ESF.
- Jarousse, J.-P., Mingat A. & Richard M. (1992). La scolarisation maternelle à deux ans : analyse des effets pédagogiques et sociaux. *Éducation et Formations*, no 31.

- Jeantheau, J.-P. & Murat, F. (2000). Les évaluations en lecture dans le cadre des journées d'appel de préparation à la défense. *Note d'information 00.09*. Paris : Ministère de l'Éducation nationale.
- Landsheere, G. de (1980). *Évaluation continue et examens : précis de docimologie*, 2^e édition. Paris : Nathan.
- Landsheere, G. de (1994). *Le pilotage des systèmes d'éducation*. Bruxelles : De Boeck-Wesmael.
- Legrand, A. (1994). *Le système E*. Paris : Denoël.
- Legrand, P. (1995). *Le Bac chez nous et ailleurs*. Paris : Hachette Éducation.
- Mac Beath, J., Meuret, D., Schratz, M. & Jakobsen, L.B., 1999. *L'évaluation de la qualité de l'enseignement*. Bruxelles : Commission européenne.
- Meuret, D. (2001). *Les recherches sur la réduction de la taille des classes*. Paris : Haut Conseil de l'évaluation de l'école.
- Michel, A. (2001). Évaluer pour piloter. *Administration et Éducation*, no 2/2001.
- Monteil, J.-M. (1999). *Propositions pour une nouvelle approche de l'évaluation des enseignants*. Paris : Ministère de l'Éducation nationale.
- Musselin, C. (2001). *La longue marche des universités françaises*. Paris : PUF.
- Mustar, P. (2000). *Les chiffres clés de la Science et de la Technologie*. Paris : Economica.
- OCDE (1995). *Les normes de résultats dans l'enseignement : à la recherche de la qualité*. Paris.
- OCDE (1998). *Guide des meilleures pratiques à suivre pour l'évaluation*, PUMA, Note de synthèse no 5. Paris.
- OCDE (1999). *Thematic review of the transition from initial education to working life, country note : United Kingdom*. Paris.
- Orivel, E. & F. (2000). Les comparaisons internationales de l'efficience interne des systèmes éducatifs, in Paul, J.-J. (dir.), *Administrer, gérer, évaluer les systèmes éducatifs*. Paris : ESF.
- Ovaert, J.-L. & Pietrzyk, G. (rapporteurs) (1996). *Évaluation et notation des personnels enseignants du second degré*. Paris : IGEN.
- Pair, C. (dir.) (1998). *Rénovation du service public de l'Éducation nationale : responsabilité et démocratie*, rapport au ministre de l'Éducation nationale, de la recherche et de la technologie, paru sous le titre *Faut-il réorganiser l'Éducation nationale ?*. Paris : Hachette.
- Pelletier, G. (dir.) (1998). *L'évaluation institutionnelle de l'éducation*. Montréal : Éditions de l'Afides.
- Peretti, A. de (dir.) (1980). *Recueil d'instruments et de processus d'évaluation formative*. Paris : INRP.
- Peretti, C. (1994). L'opinion des enseignants sur l'évaluation en classe de seconde à la rentrée 1993. *Note d'information 94.06*. Paris : Ministère de l'Éducation nationale.
- Perrenoud, P. (1998). Évaluer les réformes scolaires, est-ce bien raisonnable ?, in Pelletier, G. (dir.), *L'évaluation institutionnelle de l'éducation*. Montréal : Éditions de l'Afides.
- Pons, C. (1996). Connaissances en français et en calcul des élèves des années 20 et d'aujourd'hui. *Note d'information 96.19*. Paris : Ministère de l'Éducation nationale.
- QAA (1998). *Quality Assurance in UK, Higher Education, a brief guide*. Gloucester : Quality Assurance Agency for Higher Education.
- Rémond, M. (1996). Pour une pédagogie de la compréhension au cycle III, in Observatoire national de la lecture, *Regards sur la lecture et ses apprentissages*. Paris : Ministère de l'Éducation nationale.
- Salamé Sala, J. (2001). Une réforme qui ne dit pas son nom. *Administration et Éducation*, no 2/2001.
- Salines, M. & Vrignaud, P. (2001). *Apprécier et certifier les acquis des élèves en fin de collège : diplôme et évaluations-bilans*. Paris : Haut Conseil de l'évaluation de l'école.
- Schmitt-Rolland, S. & Thauvel-Richard, M. (1996). Pratiques pédagogiques de l'enseignement du français en sixième et progrès des élèves. *Note d'information 96.39*. Paris : Ministère de l'Éducation nationale.
- Seldin, P. (1960). The change and scope of accreditation, in Young, K.E. (dir.), *Understanding accreditation*, San Francisco : Jossey-Bass.

- Sylvestre, S. (1998). L'évaluation des établissements d'enseignement : un tour d'horizon, in Pelletier, G. (dir.), *L'évaluation institutionnelle de l'éducation*. Montréal : Éditions de l'Affides.
- Thaurel-Richard, M. (1999). L'enseignement de la lecture au CP et au CE1. *Note d'information 99.14*. Paris : Ministère de l'Éducation nationale.
- Thaurel-Richard, M. & Verdon, R. (1996). Pratiques pédagogiques de l'enseignement des mathématiques en sixième et progrès des élèves. *Note d'information 96.44*. Paris : Ministère de l'Éducation nationale.
- Thélot, C. (1993). *L'évaluation du système éducatif : coûts, fonctionnement, résultats*. Paris : Nathan.
- Vogler, J. (coord.) (1996). *L'évaluation*. Paris : Hachette Éducation.
- Vugdalic, S. (1996). Les compétences en lecture, en calcul et en géométrie des élèves à l'entrée au CE2 et en sixième. *Note d'information 96.22*. Paris : Ministère de l'Éducation nationale.

Liste des personnes consultées

- M. Alain ABECASSIS, chef de service à la direction de l'enseignement scolaire au ministère de l'éducation nationale
- M. Patrick BARANGER, directeur de l'IUFM de Lorraine
- Mme France BAUMGARTNER, directrice des ressources humaines de l'académie de Nancy-Metz
- Mme Geneviève BECQUELIN, doyenne de l'inspection générale de l'Éducation nationale
- M. Gilles BERTRAND, professeur à l'université de Bourgogne, président du Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel
- M. Gérard BONNET, chargé de mission à la direction de la programmation et du développement
- M. Norberto BOTTANI, professeur à l'université de Genève
- M. Guy BOUISSOU, directeur de cabinet de la rectrice de l'académie de Limoges
- M. Bernard CAZEAU, président du Conseil général de la Dordogne et de la commission de l'éducation de l'Assemblée des départements de France
- Mme Véronique CHANUT, rapporteure générale du Conseil national de l'évaluation
- M. Jean-Richard CYTERMANN, directeur de la programmation et du développement au ministère de l'éducation nationale
- M. Marc DEBÈNE, recteur de l'académie de Rennes
- M. Jean-Philippe DECROUX, proviseur du lycée Robert de Mortain, Mortain, Syndicat national des personnels de direction de l'éducation nationale
- M. Jean-Louis DEROUET, chef de département à l'Institut national de recherche pédagogique
- M. Philippe DESVAUX, IA-IPR de l'académie de Limoges
- M. Alain DUMONT, directeur de la formation au Mouvement des entreprises de France
- M. Éric DUPUY, chef du service d'information et d'orientation de l'académie de Limoges
- M. Michel DURAND, IA-IPR à l'inspection académique de Meurthe-et-Moselle
- Mme Marie DURU-BELLAT, professeure à l'université de Bourgogne (Institut de recherche en économie de l'éducation)
- M. Jean-Claude ÉMIN, chef de la mission de l'évaluation à la direction de la programmation et du développement
- Mme Marie-Christine FACCHINI, proviseur du lycée Stanislas, Villers-lès-Nancy
- M. Pierre FALLOURD, sous-directeur à la direction de la programmation et du développement
- M. Pascal FAURE, directeur de la pédagogie de l'académie de Nancy-Metz
- M. Éric FAVEY, secrétaire national de la Ligue de l'enseignement et de l'éducation permanente
- M. Jean-Pierre FINANCE, ancien président d'université, délégué général de la Conférence des présidents d'université
- M. André FLIELLER, professeur à l'université de Nancy II
- M. Jean-Jacques FRANCOIS, délégué aux enseignements techniques de l'académie de Nancy-Metz
- M. Jean-Marc GEBLER, ancien directeur d'IUFM, consultant au Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel
- Mme Bernadette GROISON, secrétaire nationale de la Fédération syndicale unitaire
- M. François-Régis GUILLAUME, chargé de mission auprès du directeur de la programmation et du développement
- M. Jean GUYOT, principal du collège Edmond de Goncourt, Pulnoy
- M. Hervé HAMON, écrivain, éditeur
- M. Jean-Jacques HAZAN, administrateur national de la Fédération des conseils de parents d'élèves des écoles publiques
- M. Daniel HUET, responsable de la division des études, de l'évaluation et de la prospective au rectorat de Nancy
- M. Christian JANET, président de la Fédération des parents d'élèves de l'enseignement public
- Mme Gisèle JEAN, secrétaire nationale de la Fédération syndicale unitaire

M. Jean-Loup JOLIVET, délégué général du Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel

Mme Liliane KERJAN, rectrice de l'académie de Limoges

M. Jean-Pierre KOROLITZKI, chef du service des contrats et formations à la direction de l'enseignement supérieur

M. Daniel LECLERCQ, chef du service d'information et d'orientation de l'académie de Nancy-Metz

M. Jean-Pierre LEMEUR, chef du dispositif de validation des acquis de l'académie de Nancy-Metz

M. Pierre LÉNA, professeur à l'université de Paris VII, membre de l'Académie des sciences

M. Bernard LEHALLE, doyen des inspecteurs de l'éducation nationale (enseignement technique) de l'académie de Nancy-Metz

Mme Jacqueline LEVASSEUR, chef de bureau à la direction de la programmation et du développement

M. Joseph LOSFELD, recteur de l'académie de Nancy-Metz

M. Ulf LUNDGREN, professeur à l'université d'Uppsala, ancien directeur de l'agence nationale de l'éducation de Suède

Mme Paulette MAILLARD, secrétaire nationale de la Fédération de l'éducation nationale

Mme Françoise MALLET, chef du service des formations à la direction de l'enseignement scolaire

M. William MAROIS, recteur de l'académie de Montpellier

M. Denis MEURET, professeur à l'université de Bourgogne (Institut de recherche en économie de l'éducation)

Mme Catherine MOISAN, inspectrice générale de l'éducation nationale, chargée de mission auprès du directeur de la programmation et du développement

Mme Françoise ŒUVRARD, chargée de mission à la direction de la programmation et du développement

Mme Thérèse PAPEIL, Union nationale des associations de parents d'élèves de l'enseignement libre

Mme Anne-Marie PERRIN-NAFFAKH, directrice de l'Institut national de recherche pédagogique

M. Pascal PIERRE, responsable des techniques de l'information et de la communication pour l'enseignement à l'inspection académique de Meurthe-et-Moselle

M. Michel PRIN, adjoint à la directrice des ressources humaines de l'académie de Nancy-Metz

Mme Annie QUINIOU, secrétaire nationale du Syndicat national des lycées et collèges

M. Joël RAVAILLÉ, responsable du département d'analyse de gestion et d'administration des données au rectorat de Limoges

M. René REGNAULT, maire de Saint-Samson-sur-Rance, président de la commission éducation, sport, culture de l'Association des maires de France

Mme Catherine RÉGNIER, chef de bureau à la direction de la programmation et du développement

Mme Martine RÉMOND, chercheuse à l'Institut national de recherche pédagogique

M. Bernard RICCI, conseiller pédagogique, Pompey

M. Jacky RICHARD, chef du service de l'inspection générale de l'administration de l'Éducation nationale et de la Recherche

Mme Marie-Claude RONDEAU, chef de bureau à la direction de la programmation et du développement

M. Jean-Louis ROUILLAN, délégué académique aux enseignements techniques de l'académie de Limoges

M. Xavier ROYER, directeur de la formation à l'Union des industries textiles

M. Jean-Pierre RUFFIER, IA-IPR à l'inspection académique de Savoie

Mme Martine SAFRA, inspectrice générale de l'éducation nationale

M. Claude SAUVAGEOT, chef de la mission éducation-économie-emploi à la direction de la programmation et du développement

Mme Edith SCHEURER, conseillère d'orientation – psychologue, Nancy

M. Laurent SIRANTOINE, secrétaire national du Syndicat général de l'éducation nationale-CFDT

- M. Nicholas TATE, professeur à Winchester Collège, directeur de la *Qualifications and Curriculum Authority* d'Angleterre
- M. Claude THÉLOT, conseiller maître à la Cour des comptes, Président du Haut Conseil de l'évaluation de l'école
- M. Michel THIRY, IPR-IA de l'académie de Nancy-Metz
- M. Alejandro TIANA, vice-recteur de l'université d'éducation à distance d'Espagne, président de l'association internationale pour l'évaluation des acquis des élèves
- Mme Gabrielle TORDJMAN, professeur au collège Edmond de Goncourt, Pulnoy
- M. Philippe TOURNIER, proviseur du lycée Guy Mollet à Arras
- M. Jean-Marc VINKLER, inspecteur de l'éducation nationale, Briey
- M. Jean VOGLER, chargé de mission à l'inspection générale de l'administration de l'Éducation nationale et de la Recherche
- M. Hervé ZÄGEL, inspecteur de l'éducation nationale, Longwy

Liste de sigles

ATOS : personnels administratifs, techniciens, ouvriers, de santé, sociaux et de service
BEP : brevet d'études professionnelles
BTS : brevet de technicien supérieur
CAP : certificat d'aptitude professionnelle
CE2 : cours élémentaire deuxième année
CEREQ : centre d'études et de recherches sur les qualifications
CNE : comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel
CNRS : centre national de la recherche scientifique
CP : cours préparatoire
DEP : direction de l'évaluation et de la prospective (ministère de l'éducation nationale)
DESCO : direction de l'enseignement scolaire (ministère de l'éducation nationale)
DPD : direction de la programmation et du développement (ministère de l'éducation nationale)
GCE : general certificate of education (Angleterre)
GCSE : general certificate of secondary education (Angleterre)
IA-IPR : inspecteur d'académie – inspecteur pédagogique régional
IATOS : personnels ingénieurs, administratifs, techniciens, ouvriers et de service
IEA : international association for the evaluation of educational achievement
IEN : inspecteur de l'éducation nationale
IGAENR : inspection générale de l'administration de l'éducation nationale et de la recherche
IGEN : inspection générale de l'éducation nationale
INCE : instituto nacional de calidad y evaluación (Espagne)
INPEC : indicateurs de pilotage de l'école au collège
INRP : institut national de recherche pédagogique
IPES : indicateurs de pilotage des établissements secondaires
IREDU : institut de recherche en économie de l'éducation
ITARF : personnels ingénieurs, techniciens, administratifs de recherche et de formation
IUFM : institut universitaire de formation des maîtres
IUT : institut universitaire de technologie
LOGSE : ley orgánica de ordenación general del sistema educativo (Espagne)
LOPEG : ley orgánica de la participación, la evaluación y el gobierno de centros docentes (Espagne)
NVQs : national vocational qualifications (Angleterre)
OCDE : organisation de coopération et développement économiques
OFSTED : office for standards in education (Angleterre)
PPBS : planning programming budgeting system (traduit en français par RCB)
QAA : quality assurance agency for higher education (Angleterre)
QCA : qualifications and curriculum authority (Angleterre)
QCM : questionnaire à choix multiple
RCB : rationalisation des choix budgétaires
SIGES : service de l'informatique de gestion et des statistiques (ministère de l'éducation nationale)
SPRESE : service de la prévision, des statistiques et de l'évaluation (ministère de l'éducation nationale)
UNESCO : organisation des nations unies pour l'éducation, la science et la culture

Rapports établis à la demande du Haut Conseil de l'évaluation de l'école

Directeur de la publication : **Claude THELOT**

Secrétariat général : 3/5 boulevard Pasteur 75015 PARIS

Tel : 01 55 55 77 14

Fax : 01 55 55 77 62

Adresse du site du HCéé : <http://cisad.adc.education.fr/hcee>

ISSN en cours

Conception et impression – DPD/BED



AVIS DU HAUT CONSEIL DE L'ÉVALUATION DE L'ÉCOLE

LES FORCES ET LES FAIBLESSES DE L'ÉVALUATION DU SYSTÈME ÉDUCATIF FRANÇAIS

L'appréciation des forces et des faiblesses de l'évaluation du système éducatif français peut être considérée comme la tâche permanente du Haut Conseil de l'évaluation de l'école. C'est pourquoi, il lui a paru important, dès sa création, d'esquisser un premier tableau général de ces forces et faiblesses, tableau qui se précisera chaque fois qu'il abordera des thèmes nouveaux et qui devrait évoluer en fonction des échos que rencontreront ses recommandations.

Pour étayer son premier avis sur ce sujet d'ensemble, le Haut Conseil a demandé un rapport de synthèse à Claude PAIR, rapport public qui peut être consulté sur le site du Haut Conseil : <http://cisad.adc.education.fr/hcee> à la rubrique « publications ».

En France, des pratiques diversifiées d'évaluation de l'École sont aujourd'hui largement développées.

Les deux formes principales d'évaluation que connaît traditionnellement le système éducatif

sont la notation des élèves et l'évaluation individuelle des prestations professionnelles des personnels. Leur validité est certainement discutable et elles n'ont jamais fait l'objet, elles-mêmes, d'une véritable évaluation. Le système éducatif français a préféré, plutôt qu'améliorer ces deux formes traditionnelles, les compléter par des travaux et des outils nouveaux. Les pratiques d'évaluation – et la diffusion des résultats des évaluations – ont ainsi connu un développement important au cours des deux dernières décennies.

Aujourd'hui, le système éducatif français présente, relativement à la plupart des systèmes éducatifs étrangers et par rapport aux autres services publics nationaux, une image favorable quant à la variété et à la qualité des évaluations qui y sont conduites. Certains outils ou réalisations y connaissent des développements particuliers, voire originaux, comme les évaluations diagnostiques « de masse » des acquis des élèves, organisées à plusieurs moments clés de la scolarité, ou les indicateurs de performance des lycées fondés sur une appréciation de la « valeur ajoutée » de ces établissements. Il est vrai

que, dans ce dernier cas, l'appréciation est imparfaite du fait de l'absence d'évaluation des progrès des lycéens.

Le coût de l'ensemble de ces travaux d'évaluation devrait être connu et diffusé et le Haut Conseil souhaite qu'il soit apprécié de façon plus fine et plus exhaustive qu'il l'a été jusqu'ici.

Ceci dit, l'évaluation du système éducatif français se présente sous la forme d'un foisonnement de travaux et de dispositifs variés qui se sont ajoutés les uns aux autres, sans constituer véritablement un ensemble organisé et cohérent qui permette de répondre, à la fois et à tous les niveaux, aux deux objectifs de l'évaluation : rendre compte aux citoyens et aux « usagers » de l'état du système éducatif, et donner aux responsables et aux acteurs les moyens d'une régulation et d'un pilotage.

La question majeure, aujourd'hui, est celle du très faible usage de ces évaluations.

La variété des dispositifs en place et la richesse de leurs résultats contrastent avec la faiblesse de leurs usages.

A part la notation des élèves qui intervient très fortement, en particulier dans la certification et l'orientation et qui mériterait d'être améliorée, les diverses évaluations sont efféctées très peu utilisées. Deux exemples de dispositifs dont la nature et les promoteurs sont différents l'illustrent. L'évaluation des personnels, qui mobilise une part importante de la force de travail des corps d'inspection et qui n'est ni assez homogène ni fondée sur les résultats, ne sert guère qu'à la notation « statutaire » de ces personnels et débouche rarement sur des infléchissements de leurs pratiques. Les indicateurs de performance des lycées, régulièrement rendus publics depuis huit ans, n'ont pas encore été utilisés comme point d'appui pour développer un processus de régulation du fonctionnement de ces établissements.

C'est pourquoi le Haut conseil considère que la création des conditions propres à permettre le développement des usages des dispositifs d'évaluation actuels est aujourd'hui un enjeu au moins aussi décisif, en termes de progrès, que

la recherche d'une amélioration de la qualité de ces dispositifs. Il est indispensable de mettre l'accent – et de faire porter les efforts – sur ces conditions.

Promouvoir un meilleur usage des dispositifs existant pour évaluer les politiques conduites et pour piloter le système éducatif à tous les niveaux - national, académique et local - suppose, de la part des évaluateurs, un effort important de communication, de diffusion et de vulgarisation de leurs travaux, de leurs outils et de leurs résultats, et, de la part des responsables et des acteurs, une réelle volonté d'intégrer ces outils et ces résultats dans les processus de régulation du système.

Pour encourager cet usage, un effort important en moyens tant humains que matériels est indispensable. Pour ce faire, le Haut Conseil recommande, indépendamment de ce qu'il conclura des thèmes qu'il inscrit à son programme de travail (cf. ci-dessous) :

- que la publication des résultats des travaux et la mise à disposition des outils d'évaluation soient systématiquement accompagnées, d'une part, par des informations qui permettent aux responsables et aux acteurs d'en débattre et de se les approprier et, d'autre part, d'une formation qui leur en facilite l'usage ;

- que soit régulièrement établies des synthèses de l'état de la connaissance dans des termes tels qu'elles soient réellement utiles aux responsables et aux acteurs du système éducatif ;

- que toute politique ou toute mesure nouvelle envisagée dans le système éducatif soit explicitement articulée avec un ou des dispositifs d'évaluation, et que ses conditions de mise en œuvre précisent l'usage qui doit être fait de ces dispositifs, notamment de ceux existant déjà ;

- qu'une formation professionnelle ciblée des enseignants soit développée dans tous les IUFM et en formation continue, notamment sur la question spécifique de la notation et de l'évaluation des élèves.

En tout état de cause, seul leur usage effectif, permettra aux dispositifs d'évaluation, quels que soient leurs concepteurs, d'être améliorés

et de constituer progressivement un ensemble cohérent au service du pilotage du système.

Le dispositif d'évaluation lui-même a des points faibles qui constituent autant de marges de progrès à explorer.

Tout d'abord, dans le légitime souci de rendre compte, qui constitue l'un des objectifs de l'évaluation, le ministère de l'Éducation nationale a engagé un effort important – et positif – pour apprécier et faire connaître les résultats du système éducatif et de ses composantes.

Le fait que ces travaux restent trop peu utilisés pour la régulation et le pilotage tient, pour une part, à ce qu'ils éclairent sur les résultats, mais pas encore – ou tout au moins pas encore assez – sur les processus qui conduisent à ces résultats, processus dont la connaissance et l'analyse sont indispensables pour réguler et agir. L'exemple de l'évaluation des lycées, déjà évoqué plus haut, illustre bien ce constat.

Au-delà de cette appréciation générale, et sans qu'il soit question ici de prétendre à l'exhaustivité, le Haut Conseil estime que parmi les marges à explorer figurent la question de l'évaluation des pratiques éducatives, pédagogiques ou non, ainsi que celle de l'évaluation des établissements scolaires qui est beaucoup moins abordée en France qu'à l'étranger. L'une et l'autre devraient permettre de mieux apprécier le fonctionnement du système éducatif au plus près des élèves, c'est-à-dire dans les établissements et les classes, là où l'enjeu d'une régulation efficace est le plus essentiel.

Pour sa part, le Haut Conseil se propose de faire un point plus poussé sur certaines de ces questions dont il précisera le contenu au cours des mois qui viennent :

- celle de l'évaluation des lycées ;
- celle de l'évaluation des enseignements universitaires ;
- celle de l'évaluation des pratiques éducatives ;
- celle de l'évaluation des enseignants du premier et du second degrés ;
- et celle de l'évaluation des acquis des étudiants.

Il faut multiplier les travaux d'évaluation du système éducatif, accroître le nombre et la diversité des lieux où ils peuvent être réalisés.

Le Haut Conseil ne peut qu'approuver et soutenir la volonté ministérielle clairement affichée de confirmer et renforcer un puissant « pôle d'évaluation » au sein du ministère de l'Éducation nationale. La proximité d'un tel pôle avec les lieux où se décide et se réalise à tous les niveaux l'action éducatif est certainement un facteur qui doit favoriser le développement des usages effectifs des travaux réalisés et une bonne adéquation de ceux-ci aux questions que le système éducatif doit résoudre.

Encore faut-il que l'on veuille à organiser et exploiter des synergies entre les différentes instances qui, d'une façon ou d'une autre, y contribuent : corps d'inspection, direction de la programmation et du développement, rectorats et organismes tels le Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel, ce qui n'est pas le cas aujourd'hui.

De ce point de vue, les évaluations des académies qui sont en cours actuellement, et pour lesquelles une approche des inspections générales prend notamment appui sur des données issues des travaux et des dispositifs de la direction de la programmation et du développement, semblent au Haut Conseil une amorce intéressante d'une telle synergie ; ceci d'autant plus que la démarche de contractualisation engagée avec les académies devrait s'articuler avec ces évaluations.

Mais il semble également au Haut Conseil qu'il serait souhaitable que d'autres « pôles », de toute nature – universitaires, administratifs ou privés – développent des travaux d'évaluation du système éducatif qui viendraient compléter et conforter ceux conduits au sein du ministère ou à sa périphérie. Ce sont la multiplication de tels travaux, leur confrontation et leur mise en perspective qui permettront, de façon relativement assurée, d'une part, un compte rendu du fonctionnement et des résultats du système éducatif aux « usagers » de l'école et à l'opinion, et,

d'autre part, une régulation et une amélioration du fonctionnement de ce système.



Promouvoir un usage plus intensif et efficace des évaluations et développer la qualité et la variété de celles-ci à hauteur des besoins impliquent des moyens importants dans plusieurs domaines.

Tout d'abord, en renforçant les structures en charge de l'évaluation au sein du ministère et à sa

périphérie, et en dédiant des moyens particuliers à l'animation des dispositifs d'évaluation et aux formations qui doivent les accompagner.

Ensuite, en suscitant le développement de nouveaux pôles d'évaluation et la multiplication des travaux, par exemple grâce à des appels à projets.

Enfin, en favorisant la confrontation et la capitalisation des travaux des uns et des autres, ce à quoi le Haut Conseil s'efforce, pour sa part, de contribuer ■

LES AVIS DU HAUT CONSEIL DE L'ÉVALUATION DE L'ÉCOLE

Ce troisième avis du Haut Conseil de l'évaluation de l'école, de portée générale, fait suite aux deux avis rendus précédemment sur « l'effet de la réduction de la taille des classes sur les progrès des élèves » (mars 2001) et sur « l'appréciation et la certification des acquis des élèves en fin de collège : diplôme et évaluations-bilans » (juin 2001).

Le premier rapport annuel du Haut Conseil de l'évaluation de l'école vient d'être remis au Ministre de l'éducation nationale. Comme tous les avis du Haut Conseil et les rapports qui ont été établis à sa demande, il est public et disponible sur le site <http://cisad.adc.education.fr/hcee>

AVIS du Haut Conseil de l'évaluation de l'école

Directeur de la publication : **Claude THÉLOT**

Secrétariat général : 3/5 boulevard Pasteur 75015 – PARIS

Tel : 01 55 55 77 14

Fax : 01 55 55 77 62

Adresse du site du HCéé : <http://cisad.adc.education.fr/hcee>

ISSN en cours

Conception et impression – DPD/BED