



### Jean-Claude EMIN

ancien sous-directeur de l'évaluation à la DEPP et ancien secrétaire général du Haut Conseil à l'évaluation de l'école



**"les évaluations dans le primaire, ce qu'elles disent, ce qu'elles changent"**

Invité du SNUipp-FSU 73 le 16 février 2011, JC Emin nous parle des évaluations CM2 et CE1 (et des questions que posent certains usages des évaluations) mais également des résultats et des problèmes du primaire à partir des évaluations disponibles (les sérieuses !).

On ne peut parler d'évaluations aujourd'hui sans parler des évaluations nationales CM2 et CE1.

#### Quelques généralités sur les évaluations pour commencer :

L'évaluation pose des problèmes techniques et politiques, mais **c'est d'abord un problème politique : qu'est ce qu'on va en faire etc ?**

Aujourd'hui, où on parle sans arrête de pilotage par les résultats, si on n'a pas capable de montrer de bons résultats, on risque d'être contestés. Il y a donc des enjeux politiques extrêmement forts pour le pouvoir et pour le ministre.

Il y a un problème de recherche de légitimité au travers des évaluations, qui rend d'autant plus compliqué les discussions sur les évaluations et leur mise en place. Il y a finalement une politisation des usages des évaluations qui est liée à une très forte médiatisation, et qui n'est pas scandaleuse en soi : il peut être légitime que les citoyens français demandent à leur école où elle en est, par exemple. En même temps, ça peut conduire à de s dysfonctionnements.

C'est lié également à une idée libérale, discutable, qu'on commence à voir assez nettement en France, notamment avec les évaluations CM2, et qui est la suivante :

"je suis ministre de l'éducation, je constate qu'il y a des problèmes à tel endroit, je mets en place une évaluation" et à partir de là les choses sont censées se régler, un peu comme par miracle, de façon automatique.

Or, ce n'est pas vrai !



L'évaluation en elle-même ne règle strictement rien, et cela, il faut s'y attarder. Deux exemples :

- 1) D'une part, il n'est prouvé nulle part, à travers tous les travaux internationaux, que la mise en place d'une évaluation entraînera, par elle-même, de façon automatique, une amélioration du système et de ses résultats.

On peut même avoir des contre-effets : il y a des enjeux très forts, cela a été cité il y a un instant. Attardons-nous sur l'exemple anglais : les anglais ont mis en place des évaluations qui servent à la fois au pilotage, au classement des écoles – ça a d'ailleurs été longtemps associé chez eux à la liberté de choix très complète de l'école de la part des parents, qui revient d'ailleurs à permettre aux écoles de choisir leurs élèves. Cela s'est traduit, des travaux universitaires l'ont montré, et cela a même été dénoncé par le parlement, que ces écoles,

dans la mesure où, à travers ces évaluations, il y a des sanctions très fortes pour elles, en terme de rémunération des directeurs, en terme de crédits pour l'école etc par des **contre-effets extrêmement dangereux** :

- d'un côté, ce que les anglais appellent le **"teaching for the test"**, c'est à dire que l'enseignement tend à se concentrer dans la préparation des évaluations. De ce point de vue là, il est d'ailleurs extrêmement inquiétant d'avoir lu dans un rapport de l'Inspection Générale du premier degré sur la rentrée dernière, et notamment la mise en place des évaluations, que c'était très bien parce que les enseignants allaient se mettre à travailler les items. C'est pas les items qu'il faut travailler, ce sont les apprentissages ! Il y a une vraie question derrière ça.
  - deuxième contre-effet si l'enseignement se résume à préparer des évaluations, dans le choix de l'école. C'est ce qui a été constaté en Grande-Bretagne : les écoles ont tendance à sélectionner les élèves, et à prendre plutôt des élèves moyens, pour une raison très simple : c'est que ces élèves moyens, on a d'assez bonnes chances de les faire progresser. Par contre, il faut éviter les élèves "mauvais", parce que ceux-là, c'est beaucoup plus difficile, beaucoup plus coûteux, et cela risque d'être moins rentable au sens propre du terme. Quant aux bons élèves, à la limite, ce n'est pas très intéressant car la marge de progression n'est pas terrible. On voit bien les contre-effets, notamment de ce à quoi on échappé pour le moment, d'une publication école par école, surtout si elle est liée à une liberté de choix de l'école.
- 2) Deuxième exemple, tout à fait récent, qui montre bien l'enjeu politique d'une évaluation et de la communication sur l'évaluation. Ceux qui ont eu la curiosité de regarder sur internet la conférence de presse de Luc Chatel présentant les résultats de PISA. Elle était en deux parties. La première partie était extrêmement intéressante. C'était une présentation des résultats de PISA, qu'à la limite quelqu'un de la DEPP aurait pu faire, tout à fait correcte. Elle insistait sur une baisse de résultats dans le temps, surtout en maths, et le fait que les écarts s'aggravent. C'est notre problème très fort en France, avec les élèves les plus faibles en proportion de plus en plus importante, et avec des niveaux de plus en plus bas, et à l'autre bout de l'échelle, une petite augmentation des très bons. L'augmentation est cependant plus sensible du côté de la baisse. Cela a donc été présenté de façon tout à fait correcte.

Puis, deuxième partie : et là, se pose la question de l'utilisation des évaluations. On peut résumer grossièrement comme suit : "circulez, y'a plus rien à voir, la politique que je mène c'est exactement ce qu'il fallait faire". En bref, on ne regarde pas les évaluations, on ne travaille pas à partir de là. Elles ne servent qu'à communiquer.

On voit bien avec ça les enjeux en matière d'évaluations. Cela conduit à discuter de la question du vrai danger, si on mélange tout en matière d'évaluations.

- **Les évaluations CM2 et CE1**

Il faut revenir à une distinction fondamentale entre deux types d'évaluations. C'est la question forte des **évaluations CM2 et CE1 qui mélangent tout, entre une évaluation "diagnostique", qui elle est un outil de travail pour les enseignants, qui peut les aider à percevoir les difficultés des élèves ; et des évaluations qui servent au pilotage, alors que l'évaluation CM2 et CE1 mélange allègrement tout.**

Les évaluations CM2. Les critiques sont parfaitement justifiées (*lire l'intervention du SNUipp*). Leur objectif n'est pas clair. Le flou est entretenu même si il y a eu un peu des reculs.

Elles sont à la fois mauvaises pour faire des diagnostics :

- le binaire ne permet pas de faire du diagnostic
- elles ne sont pas construites pour repérer des difficultés
- elles n'ont pas de qualité technique
- elles conduisent à dire "il va falloir travailler les items"

C'est véritablement l'antithèse du diagnostic.

Elles sont aussi très mauvaises du point de vue du pilotage. Elles n'apportent pas grand-chose de ce point de vue là. Compter des items réussis sans tenir compte de la difficulté des items n'apporte pas véritablement de renseignements sur le niveau d'un élève.

Donc du point de vue d'un bilan des connaissances, elles apportent extrêmement peu, voire elles disent des bêtises.

De plus, ces évaluations, du moins celle de CM2, ont lieu en janvier : la stupidité conduit à ne pas plus discuter, si ce n'est que des IEN ont pu dire "le programme de maths, vous pouvez le faire en quatre mois" (sic)... Cela ne demande sans doute pas de commentaires, mais démontre à la fois la stupidité de l'exercice, son caractère dangereux, et peut-être embêtant du point de vue du fonctionnement de l'éducation nationale : faire dire à des IEN des stupidités de ce type, c'est quand même grave.

**La plupart des améliorations apportées cette année sont des améliorations en trompe l'œil. A**

regarder d'un peu plus près :

- le fait d'ouvrir la palette des modes de correction, en sortant du binaire, ne conduit pas à grand-chose puisqu'effectivement ce n'est pas introduit dans les remontées
- la question de la remédiation, enfin des outils qu'on peut fournir aux enseignants en termes de remédiation est assez caricatural. On voit bien que le problème c'est pas d'aider les enseignants à repérer les symptômes ; ce à quoi il faut l'aider, c'est à repérer des difficultés, à repérer des points forts et des points faibles, et à partir de là lui proposer des outils, des aides à la remédiation, des aides à la construction de ses apprentissages. De ce point de vue là, ça reste extrêmement faible.

Les exercices de cette année ont à nouveau été faits par le DGESCO, avec le problème d'être à la fois juge et partie, ce qui est problématique. Du point de vue du fonctionnement du ministère, ou de l'administration centrale de l'Éducation Nationale, ça pose un problème. A chacun son métier. Il vaudrait mieux que la DEPP fasse les évaluations, ce qui était le cas auparavant, et que la direction de l'enseignement scolaire, avec l'inspection générale, propose les exercices de remédiation, propose les analyses des résultats. A l'heure actuelle, ceux qui ne sont pas compétents là-dessus font les évaluations, au détriment de la proposition de remédiation.

Une chose a été faite cette année, qui risque de poser problème après. On a demandé à la DEPP de préparer d'avance, alors que l'an dernier on l'avait fait après, la correspondance entre les exercices de l'an dernier et ceux de cette année. C'est à dire que lorsque les résultats seront dépouillés, on sera capable de faire la comparaison. Sous une réserve : c'est qu'effectivement on continue d'ajouter des items et des pourcentages de réussite à ces items sans tenir compte de leur difficulté. La conséquence pour les enseignants, c'est que vous n'aurez pas, ce qui prouve bien que ce n'est pas un bon outil de pilotage, les barres du genre "bon", "très bon", "pas bon", parce que ça, ça ne pourra être fait qu'après dépouillement des résultats national.

### **On a là un outil mal fichu, ambigu politiquement et qui continuera à poser des problèmes.**

Juste pour vous faire sourire un peu, les gens de la DEPP se félicitent de voir qu'aujourd'hui on se dit "ah, les évaluations CE2 et 6è, c'était le bon temps" ! Elles restent améliorables mais ça mérite de réfléchir à ces évaluations, la façon dont elles étaient construites et à la façon dont elle étaient utilisées pour véritablement poser le problème de ce qu'il faudrait dans notre système en matière d'évaluation diagnostique.

Le projet d'origine, qui avait été envisagé en matière d'évolution de ces évaluations, n'est pas du tout ce qu'on a aujourd'hui avec les évaluations CM2 et CE1. Il était différent. Il consistait à partir de l'acquis que constituait les évaluations CE2 et 6è en essayant d'améliorer le dispositif.

L'idée était la suivante :

- faisons ces évaluations en les rapprochant de la fin d'un cycle, donc il y avait bien l'idée de les faire au CM2, mais l'idée de les faire en début de CM2, non pas d'ailleurs juste à la rentrée mais plutôt au retour des vacances d'octobre. Cela pour permettre le temps de remise en route des élèves, du début d'année, qui est important, car on sait qu'il y a des pertes d'acquis importants pendant les vacances, et que ces pertes sont socialement situées. Les élèves de milieu défavorisé perdent plus pendant les vacances que les autres. Donc il est très important si on veut faire une évaluation d'avoir laissé le temps de cette remise en route de début d'année.

- L'idée était d'en faire une véritable évaluation diagnostique, c'est à dire quelque chose qui devait permettre d'aller repérer les difficultés des élèves, avec des modes de correction autre, et de proposer aux enseignants de véritables pistes de remédiation, à un moment où il restait bien 6 ou 7 mois sur l'année pour s'efforcer d'atteindre ou de conduire le maximum d'élèves aux objectifs de fin de cycle.

Voilà le dispositif qui était envisagé, et qui a été, dans une certaine mesure, "trahi", déformé avec les évaluations CM2 et CE1.

A propos des évaluations CE1, on peut faire les mêmes critiques que pour celles du CM2, si ce n'est concernant la période, avec un danger : placer les évaluations en fin d'année peut faire courir le risque de se transformer en examen d'entrée dans la classe suivante. C'est pour ça qu'il faut attirer l'attention là-dessus ; des évaluations en fin de CM2 appliquées à tout le monde, cela ressemble fichtrement à un examen d'entrée en sixième. Il faut l'avoir en tête.



## **A partir d'évaluations sérieuses, quelles questions peut-on se poser aujourd'hui sur l'enseignement du premier degré ?**

Certaines évaluations posent de vraies questions sur le pilotage du premier degré, sur son état, sur son fonctionnement.

Un mot sur les deux types d'évaluations, déjà évoqués et sur lesquels il faut insister. Elles sont fondamentalement différentes dans leurs objectifs et leurs conceptions, même si elles se traduisent à un moment par des items ou des exercices qui se ressemblent.

### **Évaluation bilan**

Objectif : faire le point des acquis des élèves à un moment donné

Elles sont plutôt faites en fin de cycle, en tout cas en fin d'année, c'est à dire au moment où on a pu développer l'ensemble des apprentissages.

Elles se font généralement sur des échantillons. Ce n'est pas la peine d'empoisonner toute une génération d'élèves : avec entre 5 000 et 10 000 élèves, on se fait une assez bonne idée de l'état de leurs acquis.

C'est un outil de pilotage politique : "je constate que les acquis des élèves sont particulièrement défaillants en géométrie, je me pose des questions sur les programmes, sur les pratiques, sur la formation et à partir de là j'essaie de mettre quelque chose en place.

### **Évaluation diagnostique**

C'est ce devraient être les évaluations CM2 / CE1. Le projet est différent.

Objectif : repérer les difficultés des élèves, repérer aussi leurs points forts, de façon à mieux les connaître, faire le point sur les problèmes qu'ils rencontrent dans leurs apprentissages de façon à pouvoir construire des progressions adaptées, éventuellement un enseignement diversifié etc.

On a donc deux choses différentes.

L'un est un outil de nature politique, de conduite du système ; l'autre est un outil professionnel de l'enseignant, et il prend tout son sens si il est associé à des propositions de remédiation derrière.

Un des gros reproches qu'on peut faire aux évaluations CM2 / CE1, c'est le mode de correction.

Dans les évaluations CE2 / 6è, on avait un mode de correction qui avait pour objectif de repérer l'erreur, pour s'appuyer sur elle pour l'apprentissage.

Là, l'intérêt n'est pas de savoir ce que l'élève a réussi ou pas réussi, c'est de savoir ce qui coince, à quoi il va falloir s'attaquer. Et le problème sera pas de lui faire travailler un item derrière, pour qu'il ait juste à l'item ; le problème, c'est de s'attaquer à ses difficultés.

Nous sommes donc face à deux logiques différentes.

Ce qui serait intéressant, c'est que les évaluations nationales actuelles évoluent vers ce type de correction et se rapprochent au plus du début de l'année, afin de permettre de repérer et de s'attaquer à ces difficultés.

## Les autres types d'évaluation

Il existe diverses évaluations sur le plan national et international, qui donne des éléments d'état des lieux sur l'enseignement du premier degré, les questions qu'on peut se poser, les hypothèses qu'on peut faire.

### Ces évaluations sont diverses.

- Il y en a en CM2 et en 3<sup>e</sup> qui sont faites tous les 6 ans depuis 2003, sur un échantillon d'élèves, avec comme objectif de voir où en sont les acquis des élèves par rapport aux objectifs des programmes.

Elles sont faites tous les 6 ans pour un domaine ou une discipline, en faisant défiler les disciplines au cours de la période de 6 ans.

- On a des indicateurs annuels, construits dans le cadre de la loi de finances, qui sont destinés à apprécier le nombre d'élèves maîtrisant ce qui a été appelé des "compétences de base".
- Il y a aussi des comparaisons historiques, qui permettent d'apporter un éclairage à la question éternelle : le niveau monte, le niveau baisse ? Elles sont extrêmement révélatrices.
- Enfin, il y aussi les comparaisons internationales. Une sur laquelle la presse communique beaucoup, PISA, dont l'OCDE est le promoteur. Une deuxième, dans le premier degré, qui s'appelle PIRLS.

### • Évaluations fin de CM2

Ces évaluations bilan ont donc pour objectif de faire le points sur les acquis des élèves en référence aux programmes.

Elles ont lieu en fin de CM2 et en fin de 3<sup>e</sup>. Ce sont des échantillons nationaux.

Cela permet de dire des choses au niveau national, sur les différences public / privé, ZEP / hors ZEP etc.

Elles sont organisées sur un cycle de 6 ans en faisant tourner les disciplines. Il devrait y avoir d'ici peu les résultats de l'évaluation 2009, en terme de maîtrise de la langue et du langage, permettant la comparaison avec 2003. Ce n'est pas encore totalement dépouillé.

C'est une évaluation faite avec des QCM, ce qui permet de faire un certain nombre de choses, mais pas tout. Il faut d'ailleurs le garder en tête : une évaluation, ça n'épuise pas le sujet.

C'est une évaluation assez classique, avec les compétences suivantes : prélever des informations, les synthétiser etc. *Diaporama*

Elle a été construite avec des enseignants, à partir des programmes. Il y a des compétences qui n'ont pas pu être évaluées, parce que c'est trop coûteux, même si c'est possible techniquement. Participer à un débat sur un texte littéraire, par exemple... mais ce serait un dispositif beaucoup plus lourd que papier – crayon.

En 2003, cette évaluation a cependant comporté une partie orale, avec une observation faite par les enseignants avec une grille d'observation, sur les capacités des élèves à s'exprimer, à faire un résumé etc. On peut donc faire beaucoup de choses, mais il faut garder en tête qu'**une évaluation, c'est très coûteux**. Il faut dire au passage qu'on peut s'interroger sur la meilleure utilisation qui pourrait être faite de ce que coûte en terme de papier, de temps de préparation, de la prime discutable donnée à certains avec ce qu'on fait en matière de CM2 et CE1. **On peut discuter sérieusement de l'utilisation de l'argent public.**

Deux mots de technique sur ces évaluations fin de CM2.

Lorsqu'on s'intéresse aux résultats des élèves, il faut mettre en corrélation le nombre d'items réussis avec la difficulté des items. Ajouter des pourcentages d'items réussis sans tenir compte de la difficulté de chaque item, c'est un peu comme additionner des choux et des carottes !

C'est pourquoi on construit des échelles qui mettent en correspondance les compétences des élèves et les difficultés des items. *Diaporama*

Un élève est d'autant plus compétent qu'il a des chances de réussir des items de plus en plus difficiles. C'est d'ailleurs pourquoi les évaluations CM2 sont un très très mauvais outil du point de vue de l'appréciation des acquis des élèves.

Les évaluations bilan sont donc travaillées en terme de résultats, ce qui explique que ce soit si long à dépouiller.

Cela permet de construire des échelles qui ont du sens. *Diaporama* Au fur et à mesure que l'on monte

dans l'échelle, on trouve des groupes d'élèves de plus en plus compétents.

Les barres bleues et rouges correspondent à

- bleu : ce que savait faire le groupe précédent
- rouge : ce que le groupe regardé sait faire en plus du précédent

Avant de regarder plus en détail, il faut préciser, car il y a souvent des bêtises dites à ce sujet, qu'une échelle de ce type n'a strictement aucune valeur normative. La moyenne de l'échelle a été construite de façon arbitraire. C'est une valeur relative, et c'est très important en terme d'exploitation.

Si on regarde le bas de l'échelle :

Il y a deux groupes d'élèves, qui font 15 % de l'ensemble des élèves.

- **Le groupe 0** : très franchement, ils auraient répondu au hasard qu'ils auraient eu les mêmes résultats. C'est le noyau dur des élèves en difficulté scolaire, qui sont de l'ordre de 3 à 4 %. ils ne maîtrisent pratiquement rien des compétences théoriquement exigibles à ce moment.
- **Le groupe 1** : élèves qui ont du mal à rentrer dans la globalité d'un texte et qui se trouvent en difficulté dès que celui-ci dépasse une certaine longueur...

Si on grimpe dans l'échelle :

Il y a deux groupes qui se situent "au milieu", qui font à peu près un quart de la population des élèves.

- **le groupe 2** : les élèves sont capables de trouver des informations explicites, et plutôt dans un support discontinu que continu, présentés sous forme de tableaux, graphiques, schémas. Par contre, ils peuvent prélever des informations à partir d'un texte continu entendu. **La différence entre support écrit et support oral est également quelque chose d'important.**

Ceci dit, ils restent sur l'aspect global d'un texte.

- **le groupe 3** : les élèves deviennent plus compétents, sont capables d'identifier le thème d'un texte, son sujet principal ; ils connaissent les règles d'accord sujet-verbe / nom-adjectif. On se rapproche des objectifs des programmes.

Si on monte en haut de l'échelle :

On a des élèves qui, n au moins pour le groupe 5, maîtrisent l'ensemble des objectifs des programmes.

Cela n'a pas de valeur normative. L'évaluateur ne peut dire si « c'est bien » ou si « c'est mal ». La barre qui dit où doivent être les compétences d'un élève en fin de CM2 est une question à laquelle l'évaluateur non seulement ne peut mais ne doit pas répondre. C'est fondamentalement une question de politique :

- soit de politique au sens social du terme : qu'est-ce que la société française estime indispensable que les élèves aient acquis, notamment à la fin de la scolarité obligatoire ?
- Ou c'est une question de politique « pédagogique », éducative.

Les responsables politiques sont souvent tentés de renvoyer aux évaluateurs cette question politique.

Or c'est une question pour les enseignants, pour l'Inspection Générale, pour la société française. L'évaluateur, lui, décrit une situation.

A partir de ces éléments, on peut discuter en termes de pilotage, de politique éducative. On constate qu'il y a un pourcentage d'élèves en très très forte difficultés à la fin de l'école primaire, 3 à 5 %. Ces élèves, souvent, existaient dès le début. On n'a pas pu ou pas su régler leurs difficultés.

Deuxième question qui se pose : on a des élèves dont les compétences peuvent être extrêmement diversifiées. Entre ceux du bas de l'échelle et ceux du haut de l'échelle, vous avez un monde,

Pour la suite de leur formation, il va falloir en tenir compte. Cela renvoie à une question : **à quoi ça sert les programmes**, dans la mesure où cette évaluation est construite en collant au plus près aux programmes. Manifestement, si on s'imagine que tous les élèves sont capables à la fin du CM2, ou du collège, de « réciter » les programmes de A à Z, on se fait des illusions.

Poser des questions comme celle du socle a donc du sens : penser un « standard minimum » dont on estime que si on ne l'a pas, on a beaucoup de difficultés à continuer, mais qu'on peut faire mieux. Les programmes ont pour objectif de tirer vers le haut, il serait illusoire de vouloir que tous les élèves les atteignent. La question d'un « seuil minimum » est ouverte. Il pourrait se situer dans le groupe 2, si on regarde ce qui a été vu précédemment.

Il faut regarder aussi de plus près l'intérêt de cette évaluation. Elle comporte des épreuves de compréhension / expression écrite et orale.

**En termes de réussite à l'écrit et à l'oral, on voit bien que les élèves sont, notamment dans les groupes les plus faibles, moins mal à l'aise lorsqu'on les sollicite à l'oral.**

Cela pose le problème de ce qui va se passer quand ces élèves vont arriver au collège : si on les

« jette » directement dans Polyeucte, ça risque de « faire mal ». Il se pose la question de ce que sont les exigences qu'on peut avoir avec ces élèves. Ils continuent à conforter leur maîtrise de la langue à l'oral. C'est un des problèmes de notre système éducatif, avec ce passage un peu « compliqué » de l'école au collège.

Même type de remarque si on s'intéresse aux supports, entre supports continus et discontinus. **Il est manifeste, même pour les élèves les plus performants, que les élèves sont beaucoup plus à l'aise, ou moins mal à l'aise, si on les sollicite sur des supports discontinus.** Ce sont aussi des éléments importants à prendre en compte pour les pratiques.

*Précisions : un support continu, c'est un texte, plus ou moins long ; un support discontinu, ce peut être un graphique, un mode d'emploi... ce n'est pas un texte rédigé en paragraphes. C'est vrai qu'au fil des dernières décennies, on a considérablement diversifié les supports d'apprentissage de la lecture, ce qui est sans doute une bonne chose. C'est un élément qui va faciliter la maîtrise de la lecture, et les élèves seront plus à l'aise sur ces supports, ce qui est important pour les enseignants de collège ! Si on les « jette » tout de suite dans une pièce de Racine, qui est un support continu sophistiqué, ça va être difficile ! Il faudra travailler autre chose si on veut accroître leur niveau de compétences.*

Voilà donc le type de conclusions et d'enseignements que l'on peut tirer d'une **évaluation bilan, qui est une vraie évaluation de pilotage.**

- **Évaluation des compétences de base.**

A quoi sert-elle ? **C'est une évaluation construite pour nourrir les indicateurs associés au budget.** L'idée était la suivante : pour suivre dans le temps les performances du système, c'est intéressant de savoir quelles proportions d'élèves maîtrisent ou ne maîtrisent pas les compétences de base à différents niveaux.

Elle a lieu en français et en maths en fin de CM2 et en fin de collège. Il serait d'ailleurs plus juste de dire en fin d'école, car elle « ramasse » des éléments au-delà du CM2. Elle est également faite sur des échantillons, environ 8 000 élèves à chaque niveau.

Il s'agit d'une évaluation annuelle, ce qui honnêtement ne sert à rien. C'est coûteux. Sur une évaluation de ce type, si on constate une grande disparité d'une année sur l'autre, c'est que l'évaluation est mauvaise ! La question la plus intéressante de cette évaluation, ce sont ses résultats mais c'est aussi le type de discussion auxquelles elle a donné lieu.

**Résultats** : on est en fin de CM2, et on estime qu'à peu près 9 élève sur 10 maîtrisent les compétences de base, encore faut-il les définir, en français et en maths, avec des différences significatives entre garçons et filles et entre élèves « à l'heure » et élèves « en retard », sans grande surprise.

La mise en place de cette évaluation était intéressante car elle a associé les évaluateurs et les responsables pédagogiques. Elle a été faite à partir d'une série de thèmes. Une échelle a été construite en associant des compétences à des items plus ou moins difficiles, comme dans l'évaluation précédente.

Pour les évaluateurs, il a été demandé une évaluation sur les compétences de base, mais c'est aux responsables pédagogiques de dire quelles sont ces compétences et où se place la barre. Le cabinet du ministère n'a jamais répondu à la question, or c'est une responsabilité politique importante. L'Inspection Générale a répondu, il y a eu une discussion où cette dernière a dit : « voilà ce que j'estime les compétences de base », appelons ça des pré-requis, que devraient avoir un élève en fin de CM2 en français et maths pour poursuivre de manière convenable sa scolarité.

Cela a donné lieu à des discussions intéressantes.

Est-il indispensable de savoir combien il y a de « p » à apercevoir ? Faut-il maîtriser le théorème de Thalès ou la proportionnalité etc ?

On a là une vraie question pédagogique. On peut l'illustrer par une anecdote extrêmement significative, sur le côté politique de cette discussion :

s'agissant du collège, on est allé voir les différentes inspections en lettres et maths, en leur présentant l'échelle et en leur demandant où ils faisaient passer la barre.

Les maths, eux, se sont calés pour arrêter leur décision, sur ce qu'un élève devait maîtriser pour pouvoir suivre correctement une scolarité de CAP, ce qui paraît être assez logique avec la démarche ; ils ont pris la formation post scolarité obligatoire la moins exigeante, en disant la base c'est ça. Du côté des littéraires, ils n'ont pas fait la même chose, ce qui est discutable. Eux se sont calés sur les compétences nécessaires, les acquis nécessaires pour suivre avec profit une seconde générale et technologique, en oubliant totalement l'enseignement professionnel. Cela montre bien que derrière une question de ce type, on a une véritable question sociale et politique, qui exige les responsables de la politique éducative prennent précisément leurs responsabilités.

- **Comparaisons historiques**

Ces comparaisons ont été maintenues sous le coude par le ministre, Darcos à l'époque, pendant plus de 6 mois, et finalement ont été rendues publiques par une fuite, sans doute savamment orchestrée, dans le Café Pédagogique.

C'est une évaluation sur la lecture et le calcul en fin de CM2, qui est intéressante car elle porte sur une période longue : 20 ans. Elle a porté sur un échantillon d'environ 4 000 élèves, en 1987 avec reprise en 1997 et en 2007, donc comparable sur une période relativement longue.

**Résultats** : les résultats moyens des élèves sont restés les mêmes entre 1987 et 1997. **En 2007, en revanche, il y a eu une baisse relativement sensible. Dégradation de la performance moyenne, notamment sur les 10 dernières années** entre 1997 et 2007.

Deuxième élément, **l'écart type a sensiblement augmenté en 2007**, ce qui signifie que les élèves sont plus dispersés : il y en a plus « en haut » et plus « en bas ».

Si on regarde de plus près, ce n'est pas exactement ça, c'est plus subtil : si on prend les 10 % d'élèves les moins forts en 1987, au même niveau, on est 11 % en 1997, en revanche, en 2007, il y a deux fois plus d'élèves au niveau des 10 % les moins forts de 1997, c'est à dire que la branche du bas s'est considérablement abaissée.

Même logique, les 10 % qui ont les plus forts résultats en 1987, ça a peu bougé : 10 % en 1997 et encore 8 % en 2007.

C'est donc la branche du bas qui s'est écartée, celle du haut est pratiquement restée au même niveau. Il y a là une véritable interrogation sur le système.

Ce qui a conduit à garder ces résultats sous le coude, c'est cette baisse vers le bas. Un des problèmes de notre système éducatif aujourd'hui, c'est qu'on a tendance à laisser sur le bord du chemin de plus en plus d'élèves, et de les laisser en plus mauvais état... **La proportion d'élèves en difficultés augmente, et leur niveau est plus faible**, ce que montre également PISA.

Ce « divorce » n'est pas neutre socialement. L'évolution au fil des 20 ans des performances moyennes des élèves par catégorie sociale montre qu'il y a une baisse pour toutes les catégories qui correspondent à la baisse moyenne, sauf pour les professions intellectuelles (cadres, chercheurs, enseignants...).

A ce stade-là, il faut bien garder en tête que cela ne traduit pas un déterminisme social ! Cela traduit simplement que notre système est mieux adapté à certains qu'à d'autres, et que les choses se sont plutôt aggravées au fil des dernières années.

Autre enseignement de cette évaluation, c'est la diminution assez sensible des élèves entre 1987 et 1997 à une dictée, avec une dégradation de leurs performances sur la grammaire, ce qui pose là aussi une question sur notre enseignement voire la valeur qu'on accorde à certains éléments.

- **Évaluation sur les difficultés de lecture à l'entrée en sixième**

Très intéressante du point de vue pédagogique, elle a été faite en 1997 et 2007 sur deux échantillons d'élèves.

2007 est la dernière année où il y avait les évaluations CE2 et sixième.

Ces évaluations se sont intéressées aux élèves qui avaient les moindres résultats à l'entrée en sixième. On a donc mis en place une « sous-évaluation » pour ces élèves, du moins une évaluation plus fine pour aller plus loin dans l'appréciation des difficultés qu'ils rencontraient.

L'idée est de faire un diagnostic vraiment fin des difficultés rencontrées, et cela questionne pas mal.

- Par exemple, si on prend le critère de la rapidité de traitement des informations, dont on sait quelle est importante pour la lecture, on se rend compte qu'il y a une partie des mauvais lecteurs qui sont en fait des lecteurs lents, parce qu'on a une exigence et que « si il n'est pas allé assez vite, c'est parce qu'il était trop lent » (!)... Par contre, il n'y a pas de baisse significative entre 1997 et 2007 pour ce critère.
- De même, il n'y a pas de baisse significative concernant les connaissances lexicales, les mots fréquents, ni sur les connaissances phonologiques ; pas de baisse sur les connaissances morphologiques.
- En revanche, lorsqu'on passe à des domaines plus « sophistiqués » de connaissances, comme les connaissances lexicales sur les mots rares, on a une augmentation significative des élèves en difficultés ; on peut faire le même constat pour les connaissances orthographiques et la compréhension d'énoncés.

Ces éléments permettent d'aller plus loin dans la réflexion. **Nos élèves en difficultés ne le sont pas sur les compétences « basiques », mais sur des compétences plus « sophistiquées »,** ce qui renvoie



peut-être aussi à des questions d'enseignement ?

**Le phénomène est particulièrement marqué en ZEP.**, plus que dans le reste du système éducatif. Ce qui renvoie à une autre question, importante, liée au fonctionnement de notre système éducatif et aux ZEP. Plusieurs travaux montrent que les enseignants en ZEP ont tendance à travailler et « surtravailler » les compétences les plus basiques, en se disant « au moins, ils auront ça », au détriment peut-être des acquis des élèves, parce qu'on laisse de côté des travaux plus fins ou qu'on abaisse ses exigences.

Cela renvoie à des questions plus larges sur notre système éducatif ; au niveau d'exigences qu'on a à avoir à-travers nos élèves, à la question de mixité sociale et scolaire des établissements ; la question de la liberté de choix des établissements etc.

**Tous les travaux scientifiques montrent qu'en ayant des élèves de faible niveau, on a tendance à avoir des exigences plus faibles** que si l'on a des élèves de niveau plus important. D'où l'importance d'avoir des « têtes de classe », pour employer un vocabulaire classique, des classes hétérogènes dans lesquelles une partie tire l'autre et dans lesquelles maintenir des exigences relativement élevées.

A-travers ce type d'évaluation, on voit bien, notamment en ZEP, que les élèves les plus en difficulté le sont d'autant plus qu'on est sûr des compétences qui ne sont, peut-être, pas assez travaillées.

- **L'enquête internationale PIRLS**

PIRLS n'est pas menée par l'OCDE, mais par une autre institution.

Il existe une autre enquête internationale : PISA, une enquête sur les élèves de 15 ans., où qu'ils soient dans le système éducatif. Cela signifie qu'en France, on peut avoir des élèves de seconde, quelques uns en première, et pas mal encore en troisième voire une partie d'entre eux en quatrième, avec notre système de redoublement fort important.

PIRLS est une enquête qui porte sur un niveau, donc la logique n'est pas exactement la même. Il s'agit du niveau dans lequel il y a le plus d'élèves de 9/10 ans, donc pour nous le CM1, sur un échantillon national. Cette enquête s'intéresse à la lecture, avec des comparaisons temporelles puisqu'elle a déjà eu lieu en 2001 et 2006, et qu'elle aura lieu également en 2011.

C'est une évaluation plus proche de notre conception de la lecture que PISA, avec des compétences renvoyant à différents supports (textes littéraires, continus, et textes discontinus).

De très nombreux pays participent à cette évaluation. La moyenne des résultats internationaux est arbitrairement fixée à 500, et la France est très au-dessus, aux alentours de 530. Cela paraît plutôt « sympathique » comme résultat », à ceci près que l'enquête comprend un très grand nombre de pays, et qu'il y en a quand même quelques uns qui nous aident à être considérablement au-dessus de la moyenne, car la moyenne est un peu basse (Afrique du Sud, Koweït, Qatar etc). Ces pays n'ont pas un système éducatif et un développement comparable au notre.

Par ailleurs, cela montre bien qu'une enquête quelle qu'elle soit n'a en elle-même aucun caractère normatif. Du coup, les scores moyens ont été reconstruits en ne gardant que les pays comparables à la France, c'est à dire les pays de l'Union Européenne. 21 sur 27 avaient participé à l'enquête. A ce moment-là, la position de la France est un peu moins « triomphante » puisqu'on se trouve en-dessous de la moyenne des pays européens. Ce n'est pas catastrophique ; ceci dit.

C'est d'ailleurs cette enquête qui avait conduit Xavier Darcos à supprimer le samedi matin, en disant « nos résultats sont pas bons, on va à la fois supprimer le samedi matin et faire des évaluations le samedi ». Bon. C'est tout le problème de l'utilisation des résultats des évaluations et plus globalement le problème de l'utilisation par les responsables politiques des résultats de la recherche en éducation et des évaluations. Il n'y a aucune réflexion véritable à partir de travaux de recherche ou de travaux d'évaluation.

**Pour terminer...** Partant de là, il faut **se poser un certain nombre de questions sur notre enseignement**, interrogations que devraient légitimement avoir les politiques.

- 1) **la question des moyens** (est-ce que je mets assez d'argent?). Si on s'intéresse à la dépense d'éducation (tout confondu, c'est à dire dépenses des collectivités territoriales, des familles etc liées directement à l'éducation), la France est au-dessus de la moyenne de l'OCDE.

Si on regarde le premier degré, pour un élève de l'élémentaire (à partir de 6 ans, données de l'OCDE) : on est sensiblement en-dessous de la moyenne de l'OCDE, ce que vont montrer aussi d'autres indicateurs. La répartition en France des dépenses d'éducation est différente de celles des autres pays ; l'enseignement secondaire est fortement loti par rapport à l'élémentaire, et le supérieur plutôt moins loti. C'est une caractéristique qu'une étude récente du CEREQ a remis à l'ordre du jour.

La façon dont on dote notre enseignement primaire est un problème : on met moins de priorités que

dans d'autres pays. Cela constitue un véritable problème parce que c'est bien au début de la scolarité que se constituent des difficultés qu'on a ensuite beaucoup de mal à récupérer.

Si on prend le rapport postes d'enseignants pour 100 élèves, il y a aussi une baisse sensible de ce ratio (le fameux P/E dont on se sert lors de la carte scolaire). Le taux d'encadrement diminue.

- 2) **la question des effectifs** ; une étude a montré que des CP à effectif réduit avaient un petit plus, mais qu'il disparaissait dans la durée. En outre, cette expérience avait posé beaucoup de problèmes aux enseignants : passer d'un groupe d'une vingtaine à 10/12 change la donne.

Au début, la disposition de classe restait la même, par exemple : ça n'avait pas été mis à profit pour modifier, se mettre en cercle ou autre ; les rapports devenaient plus affectifs, ce qui posait problème par moment. La plupart des travaux internationaux, pour des tailles de classe du type des nôtres, une faible diminution de 1 ou 2 élèves n'apporte rien : il faut diminuer beaucoup plus, aller jusqu'à la moitié, ce qui est beaucoup plus coûteux.

En revanche, les résultats de cette enquête montre que les acquis en fin de CP sont extrêmement liés aux acquis en début de CP, d'où question sur la maternelle.

- 3) **la question de la maternelle**. Depuis les années 90, tous les 3 ans sont scolarisés. La scolarisation des 2 ans a diminué de moitié pendant la dernière décennie.

En 2000, 35 % des enfants de 2 ans étaient scolarisés, soit la moitié des enfants de cette tranche d'âge scolarisables. On est aujourd'hui à 15 %, et on va encore régresser à cette rentrée, puisque c'est ce qui sert de variable d'ajustement. Vraisemblablement, la scolarisation à 3 ans va commencer à en prendre un coup dans les régions les plus compliquées, notamment en couronne parisienne.

Il y a une vraie question. Il est clair que l'amélioration importante des résultats de notre système éducatif depuis la dernière guerre est pour une part due à la généralisation de la pré-scolarisation, un progrès à long terme. Il y a une question forte à se poser.

- 5) **la question du redoublement**. Cela fait discuter, clairement. Tous les travaux nationaux, internationaux convergent pour constater que le redoublement ne rend service ni à l'élève qu'on fait redoubler, ni au système éducatif. C'est un peu une loterie, en fonction de l'école, de l'élève, de la position dans tel ou tel tiers de la « constante macabre » etc. A niveau égal, selon l'endroit, selon le PCS (catégories socio-professionnelle), vous redoublez ou pas : il y a là une forme d'injustice.

Quand on compare quelques années plus tard les performances d'élèves de même niveau, l'un ayant redoublé l'autre pas, celui qui n'a pas redoublé a plutôt un meilleur niveau que celui qui a redoublé. On n'a donc pas rendu service à celui qui a redoublé.

Lorsqu'on fait des comparaisons internationales, **les systèmes éducatifs qui pratiquent peu, voire pas, le redoublement, ont plutôt tendance à de meilleurs résultats et plus homogènes.**

Par exemple, pour PISA, les élèves de 15 ans en France : 60 % sont « à l'heure », 40 % en retard, ce qui est quand même une petite moitié. Le redoublement en France fonctionne comme une filière, avec des conséquences sur lesquelles il faut s'interroger, notamment les redoublements en début de scolarité : l'élève qui a redoublé est marqué. En fin de collège, à niveau égal, cet élève sera systématiquement moins noté que l'autre, il ne sera pas orienté de la même manière, parce qu'il est redoublant. Il y a un vrai problème. Il faut s'interroger.

**Les enseignants ont besoin d'instruments qui permettent la comparaison, de repères en termes d'évaluations bien faites.**

**Une réelle politique d'évaluation implique d'être clair sur les objectifs :**

- **il ne faut pas confondre les outils de pilotage et les outils professionnels**
- **il faut être rigoureux méthodologiquement**
- **il faut garder en tête que l'évaluation n'apporte pas de solution ; elle permet de s'interroger, mais il faut ensuite analyser, réfléchir...**
- **l'abus d'évaluation nuit ; d'abord, cela coûte cher : on peut s'interroger sur ce qu'on fait de l'argent...mais il ne faut pas jeter l'évaluation avec l'eau du CM2.**
- **Les enseignants ont besoin d'outils pour se repérer, s'améliorer.**