

Dissonances et convergences évaluatives

De l'évaluation dans la classe aux évaluations internationales : quelle cohérence ?

Antoine Bodin(*)

1. Introduction

L'évaluation est de nouveau à la mode et risque de le demeurer assez longtemps. Il est en effet de plus en plus courant d'entendre dire qu'une meilleure évaluation serait garante d'une meilleure formation des élèves. Tout le monde y va de son couplet : les organisations internationales, les ministères et autres responsables des systèmes éducatifs, les instances strictement politiques et, par voie de contagion, une bonne partie des personnels d'encadrement. L'intérêt pour la question est moins clair pour les enseignants, lesquels, bien qu'étant les principaux acteurs de l'évaluation, sont surtout écartelés entre des demandes contradictoires et servent trop souvent de boucs émissaires.

Loin de nous l'idée de dire que l'évaluation n'est pas importante. Elle l'est, pour le meilleur comme pour le pire ! Selon les formes qu'elle prend, elle peut stimuler les apprentissages des élèves, ou au contraire les contrarier. Tout comme elle peut faciliter l'action des enseignants ou au contraire les conduire à perdre de vue l'essentiel. L'essentiel étant, personne n'en doute, le développement du savoir des élèves, c'est-à-dire le développement de connaissances et de compétences.

Mais de quelle évaluation parle-t-on ? De celle qui est conduite par les enseignants dans leurs classes, à des fins diverses que nous analyserons plus loin ? De celle qui se manifeste par les examens plus ou moins externes ? De celle, externe, qui ne concerne plus les élèves individuellement, mais les élèves dans leur ensemble (évaluations dites « de masse », ou à grande échelle, ou institutionnelle, ...) : évaluation destinée en premier lieu au pilotage du système ? En fait, de toutes en général et de chacune en particulier, avec l'idée que l'ensemble fonctionne en synergie d'une façon qui peut être, selon les cas, plus ou moins positive ou plus ou moins négative.

Les réflexions qui suivent sont en partie issues de la confrontation du cas français avec les cas d'autres pays, européens ou non. Elles s'appuient sur la publication récente de plusieurs rapports importants de l'IGEN⁽¹⁾ [18]⁽²⁾ et [19], de la DEPP⁽³⁾ [14] et de

(*) Responsable de l'Observatoire EVAPM de 1986 à 2006.

(1) IGEN : Inspection générale de l'Éducation nationale.

(2) Les nombres entre crochets renvoient aux références bibliographiques.

(3) DEPP : Direction de l'évaluation de la prospective et de la performance. Avatar récent de la DEP, elle-même héritière du SPRESE, lui-même descendant du SIGES ! (Heureux pays dont les institutions du même type, moins dépendant des politiques, peuvent garder le même sigle malgré l'évolution de leurs missions !).

l'OCDE [21]. Elles s'appuient aussi, pour une part, sur l'expérience de 20 ans de l'Observatoire EVAPM [2].

Les rapports en question sont riches et chacun d'eux mérite une lecture attentive et critique. Les citer ici, selon les cas, à charge ou à décharge selon le cas, n'épuisera pas le sujet⁽⁴⁾. Dans l'ensemble, ils renferment des analyses pertinentes et proposent des orientations qui méritent d'être étudiées. D'une certaine façon, ils manifestent une évolution tout à fait intéressante du regard de certains responsables sur la question qui nous intéresse.

Mais d'abord, faisons une place particulière à un récent texte de positionnement de l'association EARLI⁽⁵⁾[15]. Ce texte attire l'attention des décideurs, des chercheurs, et des acteurs, sur les risques d'une évaluation essentiellement centrée sur les acquis des élèves (cas des évaluations globales, nationales ou type PISA, mais aussi, on le verra, de pas mal d'évaluations menées au niveau des classes).

En voici le point central : « *Les évaluations actuelles [dans le monde] sont centrées sur l'évaluation DE l'apprentissage plutôt que sur l'évaluation POUR l'apprentissage ; elles ne portent que sur une partie de ce qu'il conviendrait d'observer et conduisent à enseigner pour l'évaluation et NON à enseigner pour l'apprentissage.* »⁽⁶⁾ [15]

Il y a longtemps que chercheurs et observateurs des systèmes éducatifs mettent en garde sur les risques d'un enseignement POUR l'évaluation (tests ou examens). On sait assez que tout enseignement piloté par l'évaluation conduit au conditionnement et conduit, finalement, à privilégier les compétences qui relèvent de l'adaptation à l'examen. Le risque est encore accru avec les évaluations globales (telles PISA) pour lesquelles l'adaptation la plus évidente consiste à déplacer les objectifs, puis à les réduire pour, finalement, les identifier à ceux de ces études.

Notons encore que l'adoption récente, par le législateur, d'un « socle commun de connaissances et de compétences » ainsi que la définition nouvelle du Diplôme National du Brevet contribuent à renforcer le caractère à la fois sensible et actuel de la question de l'évaluation. Il en est de même pour tout ce qui touche à l'évolution en cours des épreuves mathématiques du baccalauréat (qu'il s'agisse de nouveaux types d'épreuves ou de nouvelles modalités de questionnement).

L'objet de cet article est de replacer les questions d'évaluation dans leur cadre systémique, de relever le manque de cohérence entre les différents types d'évaluation (la dissonance) et de mettre en évidence l'intérêt qu'il y aurait pour tous (élèves, enseignants, administration, société, ...) à réduire au maximum cette dissonance (et donc, à gagner en ... cohérence).

Bien que s'adressant de façon privilégiée à des enseignants de mathématiques, cet article ne présente pas d'exemples strictement mathématiques. Le lecteur en trouvera dans [6], [7], [9] et, plus généralement, dans les documents issus des études

(4) À eux seuls ils justifieraient la réactivation au sein de l'APMEP d'une commission évaluation (rôle que ne peut pas remplir l'Observatoire EVAPM).

(5) EARLI : Association Européenne pour la Recherche sur l'Apprentissage et l'Instruction – European Association for Research on Learning and Instruction.

(6) « Current assessments focus on assessment of learning rather than assessment for learning. They are limited in scope, and lead to teaching for assessment, NOT teaching for learning... »

EVAPM [2]⁽⁷⁾.

L'idée que l'évaluation serait le remède à tous les maux, supposés ou réels, du système est assez répandue. Bien que, habituellement, en France, par prudence, cette idée soit exprimée de façon plus nuancée, elle est bien résumée, en substance, dans cette citation d'un représentant du MEN français auprès des organisations internationales :

« ... à long terme, seule l'évaluation est susceptible de promouvoir les changements de pratiques professionnelles nécessaires à l'adaptation à la demande d'un nouvel et différent millénaire. » [8]⁽⁸⁾

Là encore, il ne sera pas question dans cet article de nier le rôle que peut prendre l'évaluation dans le pilotage des systèmes mais simplement de mettre en garde sur un pilotage exclusivement guidé par l'évaluation.

2. Activité des élèves et évaluation

Mettant à part les examens, nous partirons du principe que l'évaluation, à quelque niveau qu'elle se manifeste dans le système, a pour but ultime l'amélioration de la qualité de la formation donnée aux élèves. Dans ces conditions, il est naturel d'aller voir en premier lieu comment l'évaluation se manifeste dans la salle de classe.

2.1. Enseigner et apprendre

Avant tout, il convient de distinguer le processus d'enseignement, conduit par l'enseignant, du processus d'apprentissage, qui se tient tout entier du côté de l'élève. Cela éviterait déjà bien des confusions. Un enseignant francophone pourra dire, en parlant de ses élèves « *je leur ai appris que ...* » et même avec un peu d'amertume « *je leur ai pourtant bien appris que ... !* ». Un professeur anglophone dira plutôt « *je leur ai enseigné que...* ; *certaines peuvent avoir appris que ...* » « *I taught them that – some may have learnt* ». En anglais, enseigner (teach) et apprendre (learn) désignent en effet des processus bien différenciés et le risque de méprise est moindre qu'en français.

Ainsi lorsque dans l'étude [14], on pose aux enseignants la question :

« *Selon vous l'évaluation fait (-t-elle) partie du processus d'apprentissage de l'élève (?)* », type de question qui commande la réponse (!), on ne peut pas s'étonner d'apprendre que plus de 95% des enseignants (échantillon représentatif) répondent « OUI ». Mais comment la question a-t-elle été comprise ? Le processus en question est-il de l'ordre du « process(us) » de fabrication ? Auquel cas la question peut être comprise comme « *est-il normal que le processus d'enseignement — celui qui est du côté de l'enseignant — fasse sa place à l'évaluation ?* » ; il ne serait pas surprenant (mais trivial) que la réponse dans ce cas soit massivement positive. La question peut aussi être comprise ainsi : « *le processus d'apprentissage — celui qui est du côté de l'élève — inclut-il l'évaluation ?* » ; je ne sais pas quelle serait la réponse à cette

(7) Le lecteur trouvera aussi sur mon site personnel une taxonomie rénovée illustrée d'exemples du domaine mathématique.

(8) texte d'origine : « ... in the long term, evaluation alone is likely to promote the changes in professional practices necessary to adapt education to the demands of a new and different millennium. »[10]

question, mais une réponse raisonnable pourrait être : « *ça dépend du contexte, des apprentissages en jeu et du type d'évaluation pratiqué* ». Pour aller plus loin il serait nécessaire de soumettre la question à l'expérimentation.

Cette nécessaire distinction entre enseigner et apprendre n'est pas originale, mais toute réflexion de fond sur l'évaluation devra commencer par préciser les places et les rôles respectifs de l'enseignant et de l'enseigné.

2.2. Travail des élèves dans le cadre de la classe et évaluation

Les parts respectives, dans les pratiques de classe, de l'activité de l'enseignant et de l'activité de l'élève diffèrent d'un système éducatif à l'autre, mais aussi d'un niveau scolaire à l'autre, d'une discipline à l'autre (mais ici nous n'envisagerons que le cas des mathématiques) et d'un enseignant à l'autre.

Nombre d'éducateurs et de chercheurs en éducation sont convaincus que le lien entre l'activité de l'élève et la façon dont son travail et ses acquis sont évalués déterminent, pour une part, leur intérêt et, en fin de compte, leurs acquis.

Il est clair que nous parlons ici du travail comme processus (traduisant l'engagement de l'élève), et non comme produit. Le terme « travail » est en effet utilisé dans le registre de l'évaluation avec des acceptions très différentes, et cela de façon concomitante :

- « *Cet élève ne travaille pas assez* » : il s'agit sans doute de l'activité « travail ».
- « *Cet élève travaille de façon désordonnée* » : il s'agit plutôt là du processus « travail ».
- « *Il ne parvient pas à se concentrer sur son travail* » : il peut s'agir là d'une tâche particulière « activité à mener », ou bien d'un constat général qui concerne le processus « travail ».
- « *Son travail n'est pas satisfaisant, je lui ai mis 6 sur 20* » : il s'agit là, sans doute, du produit du travail.

Cette polysémie du terme, à elle seule, engendre bien des malentendus, avec les élèves, les parents, l'administration, ...

Dans de nombreux pays, le temps scolaire inclut, de façon plus ou moins réglementée, des périodes réservées au travail (activité) des élèves. Il s'agit de périodes pendant lesquelles les élèves sont confrontés à des tâches (exercices ou problèmes) qu'ils sont invités à effectuer seuls ou en groupe. L'enseignant peut intervenir, donner des conseils, ..., mais, en fin de compte, les élèves devront avoir le sentiment qu'ils ont eux-mêmes fait la majeure partie du travail (et pas simplement avoir retranscrit ce que l'enseignant a dicté ou écrit au tableau, ni même avoir exécuté pas à pas ce qui leur était demandé).

Ce travail (activité de l'élève) peut être organisé pour approcher une notion nouvelle, pour contrôler la compréhension des élèves, pour contrôler que certaines notions ou résultats ont bien été étudiés, pour entraîner un savoir déjà acquis, ...

Souvent, ces travaux (processus, acquisition, production, ...) sont utilisés à des fins purement formatives, ils peuvent faire l'objet d'appréciations, voire de notes, pour informer l'élève, mais les résultats ne sont pas destinés à l'évaluation de validation (au sens défini en 3.3). Il s'agit essentiellement d'évaluations qualitatives destinées à aider les élèves dans leurs apprentissages.

Dans de nombreux pays, selon l'âge des élèves, ou selon le contexte, ces appréciations ou ces notes peuvent être communiquées aux parents, ou à des tiers, mais seulement pour les informer sur l'évolution des apprentissages. Elles n'ont aucun autre rôle externe et ne sont pas cumulées pour appuyer des décisions externes à l'acte d'apprentissage (leur utilisation dans des moyennes peut même être strictement interdite).

Dans d'autres cas, bien qu'éventuellement utilisés à des fins formatives, des travaux sont jugés, appréciés, notés et ces résultats sont utilisés pour décider du niveau global de l'élève, pour contribuer aux décisions conditionnant son futur scolaire (classements, passage de classe, orientation, examens, ...).

La confusion largement entretenue dans notre pays (mais pas seulement), au moins au niveau de l'enseignement secondaire, entre ces deux fonctions de l'évaluation (accompagner et aider les apprentissages – sanctionner les acquis) est, elle encore, source de nombre de malentendus et de blocages. Cet article ne suffira certainement pas à résoudre ces difficultés, mais il souhaite y contribuer.

2.3. Travail fait hors de la classe et évaluation

Souvent, comme partie intégrante de leur travail scolaire normal, les élèves doivent faire des travaux hors de la classe, travaux donnés et contrôlés par l'enseignant. D'un strict point de vue quantitatif, les études internationales mettent en évidence une relation positive entre ce travail des élèves et leurs résultats. Cela, toutefois, jusqu'à une certaine charge de travail ; en effet, il peut être intéressant de noter que, dans notre système, les élèves semblent avoir plutôt davantage de travail que dans des pays comparables, et cela avec des résultats moindres.

Ici encore, ces travaux peuvent être évalués à des fins formatives ou à des fins de validation ; ils peuvent être ou non notés ; ...

Nous laissons ici de côté la question du contrôle qui porte sur l'apprentissage des leçons ou sur l'exécution de travaux divers. L'évaluation ne commence que lorsque l'on s'intéresse à la façon dont l'élève est capable de mobiliser et de réinvestir les contenus qu'il est censé avoir appris.

3. Évaluation de l'apprentissage ou/et évaluation des acquis

3.1. Quelques définitions

Le passage du discours éducatif à travers les filtres successifs de paradigmes plus ou moins contradictoires (comportementalisme – ou béhaviorisme, néo-comportementalisme, constructivisme, socioconstructivisme, ...) a conduit à une babélisation croissante du domaine. Il devient urgent de nous mettre d'accord sur quelques idées-clés et sur la façon d'en parler (à l'évidence, les idées importent plus que la terminologie, mais il est nécessaire de pouvoir échanger dans la clarté et disposer d'un corpus d'expressions le plus monosémiques possible).

Une confusion importante règne dans le domaine de l'évaluation où, bien souvent, plus on parle, moins on se comprend. Je ne prétendrais pas échapper complètement à cette malédiction, mais je vais m'y efforcer en précisant les significations données dans ce texte à quelques-uns des mots ou expressions-clés utilisés.

3.1.1. *Évaluation formative*

Il s'agit de la part de l'évaluation intégrée au processus d'enseignement et destinée à faciliter l'apprentissage, à piloter le processus d'enseignement, ... ; elle n'est pas destinée à cerner les acquis des élèves, ni à faire le point sur ces acquis, ni à placer les élèves les uns par rapport aux autres, ni à décider de leur passage dans la classe supérieure ou de la réussite aux examens. Il s'agit là d'« *évaluation POUR l'apprentissage* » au sens défini dans le texte d'EARLI.

Les didacticiens des mathématiques ont en partie fait l'impasse sur cette question – ou plutôt sur sa désignation. En effet, à partir du moment où l'on s'intéresse aux situations didactiques, c'est-à-dire aux situations susceptibles de provoquer des apprentissages chez les élèves, il devient naturel de s'intéresser au regard porté par l'enseignant sur l'activité de l'élève, sur le jugement qu'il porte sur la valeur de cette activité et sur son adéquation à la situation proposée. Ce regard évaluatif, qui porte davantage sur le processus d'apprentissage que sur la qualité des productions (exécution des tâches) est connu ailleurs sous la dénomination d'évaluation formative. On peut sans doute faire l'économie d'une désignation spécifique, on ne peut pas faire l'économie de la chose.

Dans le monde, le thème de l'évaluation formative est devenu central dans les réflexions et études sur l'enseignement. Nous avons déjà cité l'étude [21] qui peut être prise comme référence.

3.1.2. *Évaluations bilan (ou de validation)*

Dans ce texte, nous utilisons ces expressions pour désigner des évaluations destinées à faire le point sur la réalité des acquis d'un élève dans un domaine donné, et cela à des fins essentielles de classement, de discrimination, d'orientation, de décisions engageant le futur de l'élève. Il s'agit là de l'« *évaluation DE l'apprentissage* » au sens défini dans le texte cité d'EARLI.

Rien ne s'oppose à l'utilisation de l'évaluation bilan à des fins formatives, mais la réciproque est douteuse. L'utilisation de l'évaluation formative à des fins de validation pose de redoutables problèmes qui vont du type de confiance qui peut s'établir (ou se détruire) entre l'élève et l'enseignant à la validité de l'information utilisée. En effet, dans le cadre de l'évaluation formative, l'information recueillie à un instant t , du moins lorsqu'elle traduit un aspect négatif de l'avancement de l'apprentissage visé, est destinée à être démentie à l'instant $t + 1$, justement par l'action de l'enseignant (aidé, tout de même, par l'élève !). Quel sens cela aurait-il pour un chef d'orchestre de faire au moment de la générale la somme de toutes les fausses notes qui auront été émises au cours des répétitions précédentes ? D'autant que l'on sait que ces fausses notes sont, d'une certaine façon, les garantes de la qualité du concert du lendemain (au cours duquel, s'il y a fausse note, ce sera sans doute une fausse note qui n'aura pas été émise ou qui aura été insuffisamment relevée au cours des répétitions).

À propos d'évaluation des acquis, il faut noter que si l'étude de l'IGEN déjà citée [18] fait particulièrement bien le point sur l'état de l'évaluation des acquis dans notre pays, elle est relativement ambiguë sur la place et le rôle de l'évaluation formative.

Centré sur l'évaluation des acquis, lesquels résultent sans doute de l'apprentissage, et non sur l'évaluation du processus d'apprentissage lui-même, le rapport aurait pu faire

l'impasse totale sur la question de l'évaluation formative. Il n'en est rien et le rapport fait même une analyse assez lucide de l'état de la question. Par exemple :

« ... les évaluations diagnostiques et formatives seront d'abord destinées à informer l'élève de ses acquis, erreurs, difficultés ou maladroitness. Cela lui permettra de passer de " je n'y comprends rien " ou " je suis nul en ... ", sources de blocages, à " je calcule mal avec les nombres décimaux mais bien avec les entiers ", par exemple. Des occasions de production individuelle ou collective librement choisie seront ménagées afin de permettre l'expression de compétences diverses. Plus largement, la co-évaluation et l'autoévaluation seront encouragées et soutenues notamment par la mise à disposition d'outils appropriés (grilles à renseigner ou produits informatiques par exemple). »[18]

Ainsi, les auteurs vont même jusqu'à promouvoir l'autoévaluation et la co-évaluation, qui font partie de l'arsenal des procédures à la disposition de l'évaluation formative, mais dont l'utilisation pour l'évaluation des acquis est pour le moins douteuse. Ils auront sans doute hésité à trancher sur le degré de perméabilité possible et souhaitable entre l'évaluation formative et l'évaluation des acquis (ce n'était pas l'objet de leur rapport et il est déjà remarquable qu'ils aient fait une place aussi grande à l'évaluation formative), mais ils ne peuvent s'empêcher de remarquer que : « L'évaluation est certes au cœur de la classe, mais comme un poids plus que comme un stimulant. »

Dans le même ordre d'idée, on lit dans [19] : « La non différenciation entre évaluation formative et évaluation bilan ne permet ni à l'administration de disposer des indicateurs nécessaires ni aux enseignants de bénéficier des outils de régulation pédagogique attendus. »

3.1.3. Évaluation sommative

Nous désignons ainsi la part de l'évaluation qui est destinée à faire le point sur l'état d'une question suffisamment étendue, d'un domaine d'étude (ou si, l'on veut, d'un chapitre de cours). Une évaluation de type sommative peut être conduite à des fins strictement formatives ou à des fins de validation, elle peut, sous certaines réserves, contribuer aux deux types de fonctions.

On notera que cette définition de l'évaluation sommative s'éloigne de celle habituellement en usage : définition qui remonte aux travaux de Scriven (1957 !), puis de Bloom, laquelle ne s'applique pas aux bilans formatifs et qui, de ce fait, ne nous semble plus en accord avec la façon moderne de concevoir le spectre des actions et des fonctions de l'évaluation.

3.1.4. Les notes

Les notes sont, a priori, des nombres utilisés pour indiquer la « valeur » de certains travaux d'élèves (valeur repérée sur une échelle numérique fixée), éventuellement d'un système de codage rendu « isomorphe » à une échelle numérique.

3.2. L'évaluation POUR l'apprentissage

L'évaluation formative est, sans conteste, une belle idée. Toutefois, chacun s'accorde à dire que, si l'expression a pénétré le discours sur l'éducation, la pratique, dans la plupart des pays, est loin d'être en accord avec l'idée.

Pourtant, les constatations et les recommandations, aussi bien aux niveaux locaux ou nationaux qu'au niveau international abondent. L'étude de l'OCDE sur l'évaluation formative [21] conclut sur un certain nombre de recommandations qui pourraient au moins inspirer la réflexion. Le lecteur trouvera en annexe un résumé de ces recommandations.

Il faut cependant se garder des simplifications hâtives. S'il est souhaitable de faire porter l'accent sur une évaluation réellement formative, conçue comme telle et vécue comme telle, en premier chef par les élèves et par les familles, il est non moins nécessaire d'évaluer à des moments clés les acquis aussi bien au niveau individuel qu'au niveau collectif. Ce qui importe, c'est d'assurer les différentes fonctions de l'évaluation sans les confondre ou les écraser les unes sur les autres.

Pour les élèves, l'évaluation se résume souvent aux notes qu'ils reçoivent et les finesses développées dans la littérature pédagogique leur sont, pour l'essentiel, étrangères. Pourtant, un minimum de sensibilisation aux fonctions de l'évaluation pourrait leur permettre non seulement de distinguer les actions d'évaluation formative des actions à visée purement sommative ou certificative, mais aussi de mettre en œuvre de façon réfléchie des stratégies d'autoévaluation, voire d'évaluation métacognitive (mieux savoir comment on fonctionne peut aider à mieux fonctionner). Cela supposerait évidemment que les autres acteurs (les enseignants sans doute, mais pas seulement) soient eux-mêmes au clair avec les différentes fonctions de l'évaluation, les articulations possibles de ces fonctions et les risques qu'il y a à mélanger ces fonctions (« *je vous dis que je fais de l'évaluation formative, donc pour vous aider, mais j'utilise cette évaluation comme sanction...* »).

Mais là encore toute une chaîne, ou plutôt un réseau de relations est à considérer. Si les enseignants sont parfois peu au clair avec les questions d'évaluation, c'est qu'ils ne sont, le plus souvent, ni formés, ni outillés en conséquence, c'est que les cadres intermédiaires ne le sont guère davantage. C'est aussi parce que l'institution met une pression forte sur l'évaluation sans s'être donné les moyens d'une clarification nécessaire (cela est dénoncé même par l'IGEN : [18] et [19]).

C'est aussi parce que dans son ensemble la société (parents, médias, politiques confondus) ne s'intéresse qu'aux résultats « mesurés » sans se soucier de la qualité des outils de « mesure » et des conséquences non souhaitées des actions d'évaluation.

3.3. L'évaluation des acquis individuels

Quittons l'évaluation formative pour nous intéresser à la partie de l'évaluation qui cherche à évaluer de façon globale les acquis individuels des élèves. Il s'agit essentiellement des bilans effectués par les enseignants et des examens.

Dans ce cadre, les notes sont omniprésentes et la notion de mesure est souvent convoquée.

3.3.1. Le contrôle en cours de formation

C'est sans doute là que la difficulté est la plus grande. C'est là que gît la difficulté d'articuler l'évaluation formative et l'évaluation bilan, et c'est là que la pression est la plus grande pour les enseignants.

Les contrôles non strictement formatifs ne sont pas seulement destinés à l'élève, ils sont en quelque sorte commandés par l'extérieur (administration, parents, ...) et, dans notre pays, ressortissent d'une culture de l'évaluation qui fait peu de cas de l'évaluation formative.

Avec toute leur bonne volonté, les enseignants ont bien des difficultés à ne pas voir l'évaluation formative, qu'ils cherchent à pratiquer, se transformer en une évaluation-contrôle permanente.

Cela est assez bien souligné dans [18], repris par [19] : « *La difficulté que rencontre la plupart des enseignants à mettre leurs pratiques d'évaluation en cohérence avec les objectifs qu'ils lui assignent, est à relier avec l'importance qu'enseignants et parents d'élèves disent porter à la note. La pression sociale qui s'exerce sur les enseignants légitime et conforte une conception de l'évaluation avant tout sommative* ».

Le rapport semble oublier, du moins dans cette citation, le rôle de l'administration et en particulier des chefs d'établissements qui, bien souvent, constituent un relais efficace des conceptions des parents.

3.3.2. Les examens

Un peu partout, les examens constituent des dispositifs d'évaluation des acquis des élèves, par rapport à un programme précis. Ils échappent toujours, totalement ou en partie, à l'autorité d'un enseignant unique (même dans les cas où certains examens sont organisés localement, par exemple au niveau d'un établissement).

Les effets en retour des examens sur l'ensemble du système ne sont plus à démontrer. En France, la question des examens est un sujet sensible et relève davantage du politique que du pédagogique. Il y a longtemps déjà, Georges Glaeser [21], repris par d'autres, avait coutume de dire que la seule chose que l'on pouvait dire d'un étudiant qui avait été reçu au baccalauréat était qu'il était arrivé jusque-là, qu'il avait passé les épreuves et qu'il avait été reçu, mais que l'on ne pouvait rien dire de ses connaissances (je cite de mémoire). Autrement dit, la possession du diplôme ne révélait rien en matière d'acquis. Cela est sans doute vrai pour les mathématiques comme pour les autres disciplines. Est-ce encore vrai lorsque l'on connaît la note reçue dans une discipline particulière ? Beaucoup pensent que oui.

Si, par le jeu, bien connu des enseignants, des barèmes et des adaptations des épreuves à l'état des élèves (pour obtenir des résultats socialement acceptables), les examens peinent à révéler quelque chose des acquis d'un élève particulier, ils pourraient facilement renseigner sur les acquis collectifs d'une classe d'âge. Malheureusement, alors qu'il est courant de faire des études sur les notes (études dites docimologiques), il est exceptionnel de faire des études qualitatives prenant en compte le contenu des copies.

En marge d'EVAPM, nous avons proposé, il y a peu, une étude sur cette question, basée sur l'étude d'un échantillon de copies. Officiellement l'institution était d'accord et une quinzaine de collègues du secondaire et du supérieur, praticiens et chercheurs, étaient prêts à contribuer à l'étude (bénévolement bien sûr). Officiellement il n'a manqué que les moyens nécessaires aux déplacements des membres de l'équipe et à l'organisation de séminaires, mais cela a dû faire plaisir à ceux qui pouvaient

s'inquiéter de l'entreprise⁽⁹⁾.

Quoi qu'il en soit, il est indispensable d'aller plus loin que les sondages d'opinion et autres considérations journalistiques en matière d'évaluation, et particulièrement en matière d'examens. Des recherches sérieuses (et donc partiellement indépendantes de l'institution) seraient nécessaires pour y voir plus clair dans l'apport des examens à la connaissance relative aux acquis des élèves et, éventuellement, sur l'effet perturbateur ou facilitateur qu'ont les examens en ce qui concerne ces acquis.

3.4. L'évaluation collective

Dans de nombreux cas, au niveau d'un pays, d'une région, d'un réseau d'établissements scolaires ou au niveau d'établissements particuliers, des évaluations bilan sont conduites dans les classes, dans les établissements, ou, encore, hors des établissements (sous le contrôle des enseignants, de l'administration, ou de personnes extérieures).

Il faut faire une place à part à ces évaluations qui, ne portant pas directement sur les individus, s'attachent à faire le bilan des acquis d'une population (en France, il s'agit des évaluations nationales, des études internationales et aussi, en mathématiques, d'EVAPM, ...).

3.4.1. PISA et les autres évaluations internationales

Les études internationales sont destinées à « mesurer » en particulier en mathématiques, les acquis des élèves à travers le Monde. Elles contribuent à informer les décideurs et, en conséquence, à orienter les politiques éducatives. Elles déclenchent souvent des réactions médiatiques primaires qui ignorent à la fois les complexités méthodologiques de ces études et les limites des interprétations possibles (illustration ici du danger de parler de « mesure »).

Les résultats de l'étude PISA 2006 ont été publiés en décembre 2007. Pour la seconde fois consécutive (2003 et 2006), ils mettent en évidence un recul apparent du niveau français en mathématiques par rapport l'enquête 2000. Une fois de plus il convient de prendre au sérieux ces indicateurs, mais cette fois il semble bien que ce sera le cas au niveau du ministère (avec, espérons le, plus de nuance et de finesse que ce dont il faut s'attendre de la part des médias). En effet, la place globale a reculé⁽¹⁰⁾ ; cependant on doit se demander ce que signifie ce recul, sachant que l'ampleur, comme la signification des écarts entre les pays, sont masqués par un classement sur une échelle volontairement opaque [9].

Il faut aussi prendre en compte le « forcing » réducteur pratiqué dans nombre de pays pour obtenir de bons résultats à l'étude. La réserve, observée à ce jour par les officiels français, n'était peut-être pas sans fondements, mais il y a fort à parier que seulement deux heures consacrées à la thématique PISA dans les classes de seconde et de

(9) Divers partenariats étaient sollicités – ils n'ont manifesté qu'un intérêt distant – et il n'était pas raisonnable de demander à l'APMEP de porter seule le poids financier d'une telle étude.

(10) De 2000 à 2006, en mathématiques, elle est en effet passée successivement du 12ème rang au 13ème, puis au 18ème, sur les 30 pays de l'OCDE concernés (les rangs indiqués ne sont pas nécessairement ceux publiés dans la presse : ils prennent en compte les intervalles de confiance sur les classements).

troisième, ajouté à l'affichage d'un intérêt des responsables et des enseignants, aurait fait monter la France de 5 ou 6 places, si ce n'est plus, dans le classement final ! Bien sûr, il n'aurait pas fallu déduire d'un tel coup de pouce artificiel que le niveau de la France se serait brutalement élevé.

Sans une étude indépendante et approfondie des résultats item par item, avec accès aux copies des élèves, on continuera à pouvoir faire dire ce que l'on voudra à cette étude. Espérons donc que, cette fois, nous ne serons pas obligés de passer par Melbourne et Boston pour obtenir les informations nécessaires⁽¹¹⁾.

De toute façon, il n'était pas besoin d'attendre les résultats de PISA 2006 pour savoir que la France se plaçait fort mal en ce qui concerne les mathématiques pour tous. Sans compter les études EVAPM, l'étude internationale TIMSS 1995 le mettait clairement en lumière. Il est vrai qu'alors il semblait acquis que nous étions parmi les meilleurs du monde en ce qui concerne le niveau mathématique des élèves des classes terminales scientifiques. Il est loin d'être évident que ce soit encore le cas (pour éclaircir ce point, il eut été judicieux de participer à l'étude TIMSS programmée pour 2008, mais il est certainement préférable de faire l'autruche !).

3.4.2. L'évaluation du système

En France, l'évaluation du système est largement dévolue à une direction ministérielle (aujourd'hui la DEPP), direction qui, loin de s'arrêter à l'organisation de la collecte des données, à l'organisation des évaluations elles-mêmes et aux traitements des données (ce qu'elle sait assez bien faire), en fournit aussi une interprétation (souvent pauvre, mais officielle et estampillée). Contrairement à ce qui se passe ailleurs, l'appel fait à des instituts de recherche universitaire est rare et soigneusement encadré. En fait, sauf en ce qui concerne quelques évaluations diagnostiques, qui ne sont pas construites pour l'évaluation du système, la plupart des études soit sont de nature sociologique, soit portent sur de vagues compétences générales, ce qui a l'avantage évident de faire l'impasse sur les disciplines. Ces études ne manquent pas d'intérêt, mais, ainsi que le souligne l'IGEN, elles contribuent à laisser planer le mystère sur les acquis des élèves.

De plus : « ... ces données sont pour le moins sous-utilisées, ou encore détournées de leur fonction initiale (ainsi que le montre l'utilisation fréquente des évaluations diagnostiques comme bilans départementaux, voire académiques ou nationaux) ... » [19]

En ce qui concerne les évaluations diagnostiques de début d'année, nous avons depuis longtemps dénoncé leur détournement. L'idée est fort heureusement reprise par l'IGEN : « Les évaluations CE2/6ème ont eu, dès le début, une fonction ambiguë. » [18]. Servant à la fois à l'évaluation formative et au renseignement d'indicateurs locaux, nationaux et même internationaux (indicateurs à l'évidence fortement biaisés), elles suscitent la méfiance des enseignants et ont peu d'effet sur les pratiques (voir [14]).

(11) Pour l'étude PISA, nous renvoyons en particulier à [7] et [9] qui cherchent à en situer l'intérêt et les limites. En complément des informations officielles à trouver sur les sites du ministère de l'Éducation nationale et de l'OCDE, pour PISA comme pour TIMSS, le lecteur trouvera de nombreux documents sur mon site Internet.

Il serait temps de commanditer, et de financer, des études totalement indépendantes, où les études d'acquis ne conduisent pas à de pauvres indicateurs seulement là pour être corrélés à des variables de nature sociologique (études qui pourraient être confiées à des IREMS ou à d'autres instituts compétents). Dans de telles études, les outils de la didactique trouveraient leur emploi tantôt pour remplacer, tantôt pour compléter les outils issus de l'« *édumétrie* » (voir plus loin).

Cette vision s'oppose cependant à celle de nos collègues qui refusent l'idée même d'études empiriques, qui savent déjà bien tout ce que l'on pourrait savoir et qui connaissent déjà les conclusions à en tirer.

Citons par exemple le texte dit des sept académiciens (dont trois médailles Fields) [3] : « ... *l'utilisation de statistiques brutes qui ne tiennent pas compte de l'évolution du contexte n'a en général aucun sens, et nous n'accordons donc à celles-ci que très peu de valeur, sinon aucune.* »

On peut craindre que ce verdict s'applique à l'ensemble des études quantitatives (données empiriques), et non seulement à la mauvaise utilisation (que l'on sait fréquente) qui peut être faite des statistiques.

3.4.3. *L'expertise de l'inspection*

Voulant placer l'évaluation dans son cadre systémique global, pointer les points où la dissonance se fait sentir et ceux où la cohérence pourrait s'améliorer, il est naturel de faire une place spéciale à l'inspection.

En effet, selon sa lettre de mission pour l'année 2007-2008, « *L'IGEN assure le suivi permanent et l'évaluation de l'enseignement des disciplines* ». Il est clair que cette mission ne peut pas s'exercer sans que les acquis des élèves soient pris en compte.

De plus, les inspecteurs régionaux (IPR-IA) sont invités à davantage s'intéresser à l'évaluation des élèves [18] : « *Au cours des entretiens, les enseignants et les inspecteurs pourraient plus systématiquement placer au centre de leurs échanges ce que savent les élèves, ce qu'ils devaient retenir de la séance et ce qu'ils en auront retenu, leurs progrès de l'année en cours. Dans un temps compté – une heure en classe, une heure d'entretien, une réunion collective – les acquis des élèves occupent trop souvent une place seconde et ne sont évoqués qu'en détour à propos d'autres sujets. On se penche plus volontiers sur les conditions propices aux acquisitions des élèves sans se focaliser sur leurs connaissances et leurs compétences.* »

D'où cette recommandation :

« *Recommandation n° 15 : Associer l'évaluation externe des enseignants et leur autoévaluation. Amener le regard des inspecteurs (généralistes comme territoriaux) à se déplacer des moyens vers les résultats, des professeurs vers les élèves.* » [18]

À condition toutefois de ne pas tomber dans la facilité qui conduirait à rendre les enseignants responsables des manques constatés chez les élèves, boucs émissaires facilement désignés, comme cela est évoqué plus haut. Un peu comme si on rendait le mécanicien responsable des défauts de fabrication de la machine qu'il prend en réparation (les défauts ici ne concernant pas spécialement les élèves, mais plutôt le système, les programmes, la société).

Cette recommandation va cependant dans le sens de ce que nous suggérons plus loin pour améliorer la cohérence du système d'évaluation. Elle suppose toutefois confiance

et respect mutuels, doigté et finesse pour les uns, esprit d'ouverture pour les autres. À défaut, sa mise en œuvre pourrait ne faire qu'accentuer les problèmes.

3.4.4. Les études EVAPM

Depuis plus de 20 ans, l'observatoire EVAPM, développé par l'APMEP, avec l'aide de l'INRP, effectue des études globales qui, de la sixième aux classes terminales, font plus ou moins régulièrement le point sur les acquis des élèves (22 études en 20 ans). Ces études présentent l'avantage d'associer largement les enseignants à leur conception, à leur déroulement, et aux analyses des résultats, tout en conservant une grande indépendance méthodologique par rapport aux possibles groupes de pression, internes comme externes⁽¹²⁾.

À ce jour, la validité des études EVAPM n'a jamais été sérieusement mise en cause et la comparaison en terme de validité peut facilement être soutenue par rapport tant aux évaluations institutionnelles que par rapport aux études internationales (il y a partout des biais difficilement contrôlables, mais la comparaison des indicateurs permet d'en atténuer la portée). De ce fait, ce n'est sans doute que par crainte de donner des idées aux autres disciplines que le rapport de l'IGEN, pourtant piloté par une inspectrice générale de mathématiques, conclut sur le fait que l'institution ne sait à peu près rien des acquis des élèves [18]. À moins que ce ne soit parce que, dans notre système, une évaluation ne peut être prise au sérieux que si elle est totalement entre les mains de l'administration. Cela serait regrettable et enlèverait une partie de la pertinence de l'ensemble de notre propos.

3.5. Notes et « mesure » : un flou à éclaircir

Lorsque l'on intervient auprès des enseignants sur la question de l'évaluation, c'est massivement la question des notes qui émerge. Comme nous l'avons souligné plus haut, les enseignants se disent tiraillés entre le sens qu'ils voudraient donner à leurs notes et celui que élèves, parents, administration, leur font jouer.

Dans leur ouvrage de 1986, peu connu, et à ce jour inégalé, « *Pour une analyse didactique des faits d'évaluation* » [12], Yves Chevillard et Serge Feldmann développent l'idée que la note attribuée à une copie dans le cadre de la classe n'est pas de l'ordre de la mesure, mais qu'elle doit être comprise comme un message. Le 8 attribué à une copie, ne dit pas « *cette copie vaut 8* », mais bien « *je vous mets 8* » et, quand tout se passe bien, je vous dis pourquoi et vous suggère comment vous pouvez exploiter mon message.

Il est à noter que les auteurs se placent là, de fait, dans le cadre de ce que nous avons appelé plus haut l'évaluation contrôle permanente (qui, à l'époque, a dû leur apparaître comme la seule pratique visible des enseignants).

3.5.1. La « vraie note » et la mesure en éducation

Après Chevillard et Feldmann [12], personne ne devrait plus pouvoir donner à la note le sens d'une mesure. Pourtant, la conception « mesure » de l'évaluation continue à se manifester dans le discours et les pratiques des divers acteurs, aidés en cela par les

(12) Le lecteur trouvera de nombreuses informations (épreuves, résultats, analyses, ...) sur le site de l'APMEP. Des brochures par étude sont aussi disponibles, que l'on peut commander sur le même site.

injonctions et textes officiels (lesquels n'hésitent pas à se contredire).

Ainsi, il est pour le moins malencontreux de promouvoir « *la mesure précise des acquis des élèves* » comme on peut le lire dans [18] : « *À tous les niveaux – central, académique, départemental et local – les responsables ont à conduire et harmoniser des politiques pédagogiques, en les fondant sur des objectifs et des résultats évaluables. Dès lors la mesure précise des acquis des élèves devrait permettre de contrôler la pertinence de leur action.* ». Bien que saluant la qualité globale des rapports cités, et l'évolution qu'ils représentent (nombre de propositions jugées naguère iconoclastes y sont reprises officiellement), nous nous permettons de dire : ... *encore un effort !*

Dans le contexte de l'évaluation et de l'éducation en général, il n'y a sans doute pas lieu de bannir l'usage du mot mesure, mais il ne faudrait pas y voir davantage qu'un emploi métaphorique⁽¹³⁾. Dans la plupart des cas, les termes de « *repérage* » des acquis des élèves, d'« *indicateur* », par exemple, sont plus appropriés et moins trompeurs que celui de « *mesure* ».

Ces dernières années, la référence obligée à la psychométrie pour parler de « mesure » en éducation a fait place à l'*édumétrie*, sans que cela traduise toujours un changement de conception. Le scientisme associé à l'édumétrie n'est pas sans danger et la prudence s'impose dans la lecture et l'interprétation des résultats des études menées sous cette bannière. Il ne faudrait cependant pas sous-estimer l'intérêt de certaines de ces études, ni l'intérêt de certaines des techniques qui leur sont associées (analyse des réponses aux items et autres techniques d'analyse des données), en particulier pour la construction et la validation d'instruments d'évaluation.

3.5.2. Notes ou pas notes ?

Là n'est pas vraiment la question, ou plutôt cette question se pose différemment suivant les contextes et les fonctions de l'évaluation. Comme message, comme résumé, comme moyen d'information, elle peut être très utile. De plus, s'il y a des systèmes éducatifs sans notes, du moins jusqu'à un certain âge, la plupart des tentatives pour passer d'un système avec notes à un système sans notes, se sont soldées par des échecs (en France comme ailleurs).

La question qui se pose, c'est la signification de la note par les uns et par les autres, et la difficulté d'avoir un même système de repérage pour des éléments difficilement comparables (apprentissage en cours, acquis, travail, effort, voire, maintenant, respect des règles de la vie scolaire !).

Un certain procès des notes n'est plus à faire. Voici, par exemple, ce qu'écrivit l'Inspection générale toutes disciplines confondues : « *La note est ... relative, peu fidèle, peu explicite. Et pourtant, elle est admise par tous, élèves, parents, enseignants, chefs d'établissement. C'est le support de (presque) tout dialogue sur les acquis des élèves.* » [19]

Pourtant, et cela est important, on sait depuis longtemps que le jugement des enseignants est plus fiable que leurs notes. Cela est réaffirmé dans le rapport de

(13) Pour qu'il y ait mesure au sens strict, il faut pour le moins qu'il y ait une grandeur bien identifiée à mesurer et que l'on ait pu se donner une unité de mesure. On est loin d'être dans ce cas lorsqu'il s'agit des acquis des élèves.

l'IGEN déjà cité :

« ...les enseignants connaissent souvent assez bien les capacités des élèves et leur niveau de maîtrise ; mais la traduction en notes, et surtout en note moyenne par discipline, efface les qualités de leurs analyses. » [19]

Cette simple citation justifierait à elle seule que l'on redonnât plus de poids au jugement des enseignants. Ce jugement pouvant d'ailleurs très bien transiter par des notes, à condition, une fois encore, que les codes soient clairs...

Donner plus de poids au jugement professoral pourrait en effet être un moyen de régler nombre de problèmes d'évaluation, tout en renforçant l'autorité des enseignants. Cela ne devrait sans doute pas se faire sans que soient mises en place des procédures modératrices (comme il en existe dans nombre de pays), mais surtout sans que soient menées des réflexions approfondies à tous les niveaux. Quoi qu'il en soit, les objections qui tiennent à l'idée de justice et d'égalité de tous les élèves devant l'évaluation ne tiennent pas une minute en ce qui concerne l'évaluation formative (et encore moins l'évaluation diagnostique qui n'est qu'une forme de l'évaluation formative).

L'évaluation formative gagne en effet à être largement individualisée (ce qui suppose que les enseignants aient à portée de main des outils d'évaluation validés). Par contre les idées de justice et d'équité tiennent devant l'évaluation visant à la validation des acquis.

Dans tous les cas, qu'il s'agisse de note, d'appréciation, ou de tout autre moyen de repérage, le message doit être aussi dépourvu d'ambiguïté que possible. Les messages, quels qu'ils soient, doivent parler vrai. Si un élève n'a rien assimilé d'un domaine donné, même s'il a, comme on dit, des circonstances atténuantes, il ne faut pas que le message qu'il reçoit, lequel, message éventuellement est transmis à d'autres (famille, administration, ...), puisse être mal interprété. Là encore, cette exigence de vérité sera plus facilement remplie si les enseignants disposent de repères et d'outils externes. Yves Chevallard parle de la fonction « véridictionnelle » de l'évaluation, en regrettant que cette fonction soit souvent mise en défaut sous des pressions diverses. Déjà, des recommandations publiées par certaines inspections pédagogiques régionales déconseillent fortement l'amalgame des notes dans des moyennes et d'une certaine façon prônent une certaine autonomie des enseignants face à l'évaluation chiffrée.

Que penserait-on en effet d'un médecin qui se contenterait de vous mettre sous le nez des résultats d'analyses et qui refuserait de s'engager dans un diagnostic personnalisé. Il s'agit ici de promouvoir un professionnalisme enseignant en matière d'évaluation, professionnalisme qui ne peut se satisfaire de produire des moyennes de notes plus ou moins hétéroclites en ajoutant, comme cela se voit parfois : « *ne mérite pas cette note* », « *a eu – ou n'a pas eu – de la chance* », ou encore, « *vaut mieux (ou moins) que sa moyenne* » !

Mais la cause semble d'ores et déjà entendue : « *Il ne sera pas fait de moyenne entre des notes à statut différent – évaluation formative ou sommative, performance orale ou écrite, travail en classe ou à la maison, participation, travail, effort, etc. – ni entre des notes obtenues dans diverses disciplines, sauf le cas échéant les années*

d'examen. » [18]

Cette position peut être prise comme un appel à une plus grande autonomie des enseignants en ce qui concerne la façon dont, en fin de compte, ils rendent compte des acquis des élèves

3.5.3. Effet Posthumus et « constante macabre ? »

Le phénomène de distribution plus ou moins gaussien des notes est connu depuis longtemps et n'est pas spécial à notre pays. Ce phénomène est connu dans la littérature pédagogique, depuis les années 40, sous le nom de « loi de Posthumus » [22]. Marcel Crahay rend ce phénomène en partie responsable de ce qu'il identifie comme « l'échec scolaire » – ce que, à sa façon, fait aussi André Antibé [1]⁽¹⁴⁾.

Voici maintenant comment Marcel Crahay (chercheur belge qui ne parle pas spécifiquement du cas français – mais qui l'inclut) définit l'effet Posthumus : « ... *Un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver d'année en année, approximativement une même distribution gaussienne de notes. L'évaluation pratiquée par les enseignants présente statistiquement un caractère normatif, c'est-à-dire qu'il s'attache à situer les élèves les uns par rapport aux autres, plutôt que par rapport à des objectifs ou des compétences préétablis.* » [13]

Et encore : « *Pratiquant par tradition une évaluation de type normatif, ils (les enseignants) sont victimes de l'effet Posthumus qui les amènent à produire systématiquement une distribution gaussienne des résultats au sein de laquelle les derniers sont destinés à l'échec. Adaptant le niveau des questions au tiers supérieur, les enseignants conçoivent des épreuves dont le niveau de difficulté varie en fonction de ce groupe. Ceci a pour conséquence que certains élèves échouent dans une classe alors qu'avec le même niveau de compétences, ils pourraient réussir dans une autre ...* »

Rappelons ici que l'évaluation est dite « normative » lorsqu'elle s'attache à comparer les élèves entre eux. La norme en question étant alors celle du groupe-classe ou du groupe de référence, et non une norme externe, comme cela est le cas pour l'évaluation critériée (évaluation qui, justement, s'attache à situer les élèves par rapport à des objectifs préétablis).

Marcel Crahay et André Antibé ont donc quelque raison de voir dans l'évaluation critériée un moyen d'échapper aux effets qu'ils dénoncent.

L'effet Posthumus s'accompagne quasi-logiquement, dans chaque classe, et pour chaque évaluation, d'une proportion d'élèves qualifiés de « faibles », et assez souvent d'élèves « en échec ». André Antibé, estime en moyenne à un tiers de l'effectif des classes, la proportion d'élèves victimes de cet effet !

S'agissant d'un phénomène, ce n'est pas de « besoin » qu'il convient de parler mais plutôt d'effet de système. Il ne faudrait pas oublier alors l'effet symétrique qui conduirait à avoir « besoin » d'une proportion semblable de bonnes notes. D'ailleurs, tandis que d'aucuns dénoncent « la constante macabre », d'autres dénoncent son symétrique.

(14) Sans prétendre à la moindre originalité, mais sans parler d'échec scolaire, j'ai moi-même analysé ces phénomènes dans un texte publié en 1981 par l'IREM de Besançon [5].

D'une façon logique, l'évaluation normative, telle que définie ci-dessus, tend à discriminer pour classer, donc à produire de la dispersion, laquelle, lorsqu'elle est accompagnée de notes, conduit de façon quasi automatique à « l'effet Posthumus », lequel conduit tout aussi automatiquement à « la constante macabre ».

Par ailleurs, André Antibi propose une ingénierie évaluative que l'on peut considérer comme un traitement symptomatique des troubles produits par l'évaluation (suppression de la douleur sans effet sur les causes). On ne peut cependant pas passer les expériences correspondantes sous silence, ne serait-ce que parce qu'un certain nombre d'enseignants s'y trouvent engagés (pas seulement en mathématiques), mais aussi parce que ces idées sont prises au sérieux dans différentes sphères d'influence (politiques, administratives, syndicales, ...). Si la stratégie proposée n'est que symptomatique, comme nous le pensons, elle n'est évidemment pas sans risque⁽¹⁵⁾. La méthode, dite « système d'évaluation par contrat de confiance » vise, comme son nom l'indique à mettre les élèves en confiance et cela en leur communiquant, d'une façon que beaucoup jugent trop précise, les contenus et les conditions de leurs évaluations.

Cette démarche, si elle a l'avantage de rendre les objectifs plus clairs aux yeux des élèves, à moins que ce ne soit seulement la façon dont ces objectifs sont évalués, a aussi l'inconvénient de conforter les élèves dans l'importance qu'ils donnent naturellement à la note. Or, l'importance excessive donnée à la note se fait, généralement, au détriment du sens des apprentissages et donc au détriment des enjeux de savoir.

Cependant, il est délicat de parler d'enjeux de savoir lorsqu'il s'agit d'élèves qui, en matière d'enjeux, reçoivent essentiellement des retours négatifs de leurs efforts et plus encore pour ceux qui, découragés pour des raisons diverses (pas seulement l'école ! pas seulement l'évaluation !), n'accordent plus guère d'intérêt à ces enjeux.

Rappelons encore que l'étude PISA met notre pays au premier rang en ce qui concerne l'anxiété des élèves devant les mathématiques et que cela est directement lié à la façon dont ils vivent l'évaluation.

Finalement, seule une mise à plat du problème de l'évaluation et un recadrage de l'évaluation formative (et donc du didactique) par rapport à l'évaluation de validation permettront d'avancer. Sur cette question comme sur d'autres, des recherches sérieuses s'imposent (et non, simplement, des sondages d'opinion !).

4. Nouveaux enjeux, nouveaux défis

L'évaluation pose depuis longtemps des problèmes difficiles, peu ou mal pris en compte dans notre système et encore peu étudiés par les didactiques disciplinaires (EPS éventuellement mise à part).

À cela viennent s'ajouter des difficultés liées à l'insistance renouvelée sur la question des compétences et des attitudes, à l'instauration par le législateur d'un socle commun de connaissances et de compétences, ainsi qu'à l'obligation, elle aussi renouvelée, de rendre compte des résultats de l'enseignement, en particulier dans le cadre de

(15) Voir, par exemple, les critiques faites par Pascale Pombourq, présidente de l'APMEP, à l'adresse : <http://www.vousnousils.fr/> (mot-clé : macabre).

l'application de la LOLF⁽¹⁶⁾.

4.1. La question des compétences

La notion de compétence souffre de ses filiations multiples : psychologique, managériale, linguistique, ... À cela s'ajoute dans la littérature pédagogique (textes officiels ou non) son intégration dans des discours issus des paradigmes différents évoqués en 3.1. Tout cela pour parvenir à une confusion généralisée qui rend la communication à son propos pour le moins difficile.

« *Des obstacles liés aux injonctions institutionnelles elles-mêmes, qui n'hésitent pas à multiplier les formules et les prescriptions (livret personnel de l'élève, livret scolaire, livret de compétences), à désigner, au gré des textes, une même notion avec des termes différents. Compétences, capacités, savoir-faire, connaissances, ..., l'utilisation de ces termes parfois d'une façon indifférenciée, ou à l'inverse l'utilisation du même terme pour désigner des notions différentes ont contribué à créer une confusion qui ne peut que conforter celle qui règne parfois au sein des disciplines, chacune semblant vouloir utiliser son propre vocabulaire de référence.* » [19]. On ne saurait mieux dire !

Il est de fait que la lecture successive des textes officiels, de la recommandation de la Commission européenne, du décret du BOEN définissant le socle, puis des grilles de référence par pilier, illustre parfaitement ce qui précède !

Passons sur la requalification en compétences de ce qui était naguère appelé *capacités*. La fonction « remplacer » des traitements de texte est en effet particulièrement commode pour mettre au goût du jour des textes venant tout droit de la pédagogie des objectifs. En réalité, même pour les acteurs de bonne volonté qui acceptent de remettre leurs certitudes en question et de réfléchir, la question est ardue. Il faudra pas mal de formations communes aux différentes catégories d'acteurs (enseignants, inspecteurs, personnels de direction) pour faire de cette bouillie quelque chose de consistant (je dis bien formation et non vague commission !).

Sans vouloir entrer dans des débats de spécialistes, et pour faire simple, on peut considérer la question des compétences comme la forme moderne de la tête bien faite opposée à la tête bien pleine chère à Montaigne. Tous les enseignants souhaitent, certes, que leurs élèves acquièrent des connaissances mais aussi, et peut-être surtout, qu'il soient en mesure de mobiliser d'eux-mêmes ces connaissances dans des situations qui prennent du sens pour eux-mêmes et, dans une certaine mesure, pour autrui.

Gérard Vergnaud⁽¹⁷⁾ distingue la *forme prédicative* et la *forme opératoire* de la connaissance. Ces deux formes ne s'emboîtent pas et ne sont que partiellement dépendantes. Un élève peut avoir une connaissance particulière sans être capable de l'exploiter dans des situations qui, pourtant, pourraient relever d'elle. Il peut être capable de traiter certaines situations sans pour autant avoir intégré, du moins sous

(16) LOLF : Loi Organique relative aux Lois de Finance (loi qui entre en vigueur cette année et qui conditionne l'attribution des moyens à une obligation de résultat).

(17) Voir sur le site de Jacques Nimier [23] un petit texte concis et clair de Gérard Vergnaud : *Les compétences, Bravo ! Mais encore ? Réflexions critiques pour avancer.*

leur forme déclarative, les connaissances (procédurales) que pourtant il met en œuvre.

Nous parlons ici d'évaluation et non, directement, d'enseignement, ni d'ailleurs de développement de compétences. Le problème des compétences, pour cet article, est celui de leur identification (ce qui suppose un minimum d'accord au niveau des définitions) et de leur opérationnalisation évaluative (comment reconnaître la situation de telle personne par rapport à telle compétence).

Nous souhaitons, bien sûr, que les élèves soient conduits vers des compétences, et non seulement vers les « *connaissances inertes* » et les « *savoirs morts* » évoqués dans [19], mais nous ne proposons pas de méthode magique pour y parvenir. Le débat actuel sur les compétences porte sur l'approche (de l'enseignement) par les compétences, ou APC, et non sur la question de l'évaluation des compétences. Pour nombre de systèmes éducatifs, l'APC, sous la pression des organisations internationales, est en effet en train de devenir la référence obligatoire en matière d'enseignement ; et cela sans que des validations didactiques aient pu être apportées⁽¹⁸⁾.

À propos de leur évaluation, et avant même toute identification, nombre de conceptions, parfois largement diffusées, méritent d'être revisitées. Parmi ces conceptions notons :

- Les compétences ne pourraient être associées qu'à des situations complexes et ne pourraient être évaluées que dans le cadre de situations complexes (et l'on donne alors au mot « complexe » un sens particulier qui conduit à ne rechercher que des situations en effet très complexes).
- On ne pourrait pas évaluer les compétences sans connaître leurs conditions d'émergence (la formation qui a pu les installer).
- Les compétences ne pourraient qu'être associées à des situations de la vie réelle, voire quotidienne.
- Les compétences s'opposeraient aux connaissances.
- Une compétence serait acquise ou non (il n'y aurait pas de degrés dans l'acquisition d'une compétence).
- Une compétence donnée pourrait toujours être évaluée par une mise en situation unique.

En ce qui concerne la complexité des situations, soulignons qu'elle leur est intrinsèque. Ce qui signifie que la complexité d'une situation ne dépend pas de l'aptitude d'un sujet à l'affronter. En fait, il est courant de confondre la complexité et la difficulté, laquelle dépend en effet des sujets.

Au sens strict, la complexité commence dès qu'il y a deux éléments en présence. Dans cette acception, il est bien rare de rencontrer des questions d'évaluation, au moins en mathématiques, dont la moindre analyse de la tâche ne mette pas en évidence un certain niveau de complexité (dont on peut rendre compte par des graphes associés aux traitements possibles).

Se référer de façon systématique à la vie réelle, à l'authentique, au quotidien, pour

(18) Voir sur le site Educmath un débat sur la question, débat auquel sont associés plusieurs didacticiens des mathématiques.

parler des compétences, et pour les identifier, revient à nier certaines des fonctions essentielles de l'éducation scolaire. En mathématiques, cela conduit à refuser l'abstraction, les symboles, et la charge de la preuve (comme cela est largement le cas, par exemple, dans les études PISA). À moins, évidemment, de considérer que le réel d'un individu est ce qui fait sens pour lui : le réel d'un mathématicien peut alors très bien se situer au niveau de structures abstraites, comme le réel d'un amoureux du grec ancien peut se trouver du côté de la Grèce de Pythagore.

En ce qui concerne les conditions d'émergence des compétences, peu nous importe, au moment de l'évaluation bilan, de savoir par quels méandres d'apprentissages un individu est passé. Dans le cas, par exemple, où ses compétences lui permettent de vous réparer de façon quasi définitive un chauffage central que vous n'aviez jamais réussi à régler correctement, ou dans celui où ses compétences musicales le rendent capable d'interpréter une fugue de Bach d'une façon qui vous laisse sans voix, ou encore dans le cas où ses compétences mathématiques lui permettraient ... de traiter de A à Z le sujet du bac S, la question de savoir comment il est arrivé là ne concerne pas directement l'évaluation des compétences qu'il manifeste. À moins qu'il ne s'agisse d'évaluer d'autres compétences telles celles qui relèvent du savoir apprendre, ou encore si, dépassant les compétences du sujet, on cherche à éclaircir ce qui a bien pu favoriser l'éclosion de telles compétences.

Sur ce dernier point, ainsi qu'à propos du rapport entre connaissances et compétences, citons encore [19] : « *La focalisation sur la notion de compétences permet ... de porter une attention accrue aux processus d'apprentissage, à la façon dont l'élève apprend et utilise ses connaissances, et finalement au fonctionnement cognitif des individus.*

Elle ne disqualifie pas, loin de là, l'absolue nécessité d'ancrer les apprentissages sur l'acquisition rigoureuse, étayée, de connaissances solides sans lesquelles les compétences visées ne seraient que châteaux de sable. Mais elle rappelle l'ardente obligation de donner du sens aux savoirs enseignés à l'école, d'en augmenter la portée au-delà de l'horizon de la seule réussite aux épreuves scolaires, et de mettre au premier rang des missions de l'école la formation de la pensée autonome. »

Mis à part le mot « *focalisation* » qui nous semble trop fort (nous remplacerions volontiers le début de cette citation par « *l'attention portée aux compétences* »), nous ne pouvons qu'approuver.

Sans vouloir ajouter à la confusion, nous proposons une conception qui nous semble avoir l'avantage de laisser toute sa place à la didactique de notre discipline (et à des recherches en didactique), tout en étant compréhensible et acceptable par les autres disciplines. De plus, nous verrons que cette définition est parfaitement opérationnelle :

- Les compétences sont des réseaux organisés de connaissances et de schèmes⁽¹⁹⁾ (certains ajouteront les attitudes) ; chaque compétence étant associée à une classe de situations.

On pourrait dire aussi dire que les compétences sont des connaissances en acte.

Les compétences sont-elles premières ou sont-ce les classes de situations qui sont premières ? En d'autres termes doit-on chercher d'abord à identifier les classes de

(19) Cela au sens de la théorie des champs conceptuels. Voir G. Vergnaud [24].

situations pour ensuite leur associer des compétences, ou doit-on faire le contraire ? Bien sûr, ni l'un ni l'autre : les deux démarches demandent à être menées conjointement et doivent conduire à des itérations, de façon à atteindre un minimum de consistance dans ce qui sera identifié comme une compétence.

Pour cela nous disposons d'outils qui ont fait leur preuve mais qui sont sous-utilisés. Il s'agit en particulier de la théorie des champs conceptuels et de l'analyse implicative. La théorie des champs conceptuels permet de structurer des ébauches de compétences tandis que l'analyse implicative, associée à d'autres méthodes de traitement de données, permet les affinages et les validations nécessaires.

On aura compris qu'il s'agit de développer des recherches et des expérimentations et pas seulement de confronter des opinions, ni d'ailleurs de répondre dans l'urgence à des demandes irréalistes.

Seules de telles recherches peuvent permettre une définition opératoire des compétences. Une telle définition s'accompagnant évidemment de situations d'évaluation susceptibles d'en évaluer le niveau de développement chez les élèves.

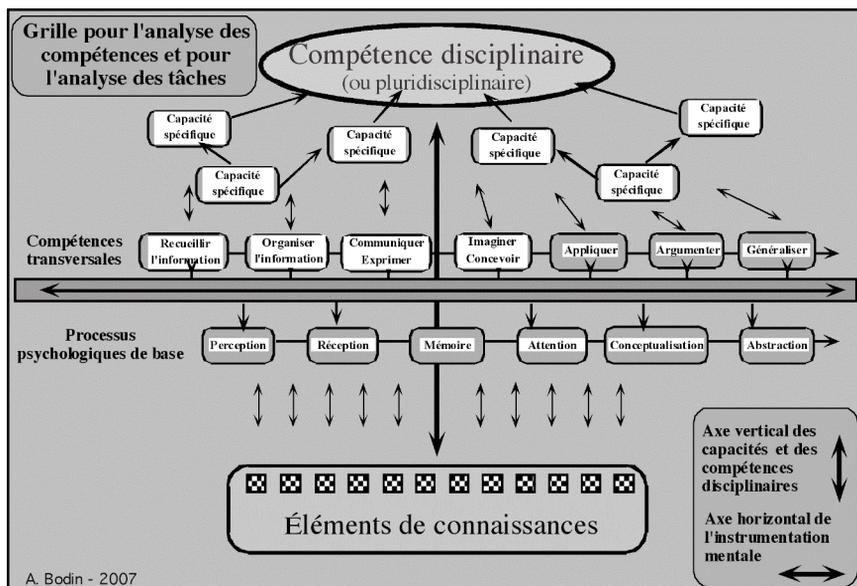
Il ne s'agit pourtant pas, en mathématiques, de partir de rien. Le travail important réalisé naguère à l'APMEP sur la question des problématiques (Régis Gras et les équipes problématiques collège et lycée), comme le travail sur les noyaux thèmes peuvent en effet trouver une deuxième jeunesse dans le cadre des compétences. Les études EVAPM, si elles ne répondent pas directement aux questions posées par la définition et l'évaluation des compétences, recèlent à la fois des questionnements et des données à grande échelle recueillies au fil des années qui pourraient trouver là une nouvelle utilité.

L'existence de documents officiels qui ont été produits dans l'urgence et, comme nous l'avons souligné plus haut, dans une certaine confusion, n'est pas trop inquiétante. Sur la question des compétences, ils sont à peu près vides. On peut les considérer comme des cadres qu'il conviendrait de remplir, quitte à devoir déplacer certaines lignes et, simultanément, à les opérationnaliser en matière d'évaluation.

Dans la mesure où les compétences sont associées à des classes de situations, il est exclu de penser évaluer une compétence particulière par une seule situation particulière. Il faudra plutôt penser à des épreuves qui, centrées sur des compétences particulières, permettraient de repérer l'état de développement de compétences particulières chez les élèves. S'il n'est pas nécessaire pour cela de connaître les conditions dans lesquelles ont eu lieu les apprentissages, il est néanmoins nécessaire d'observer l'activité que les élèves développent en situation. Il faut donc s'intéresser non seulement aux produits (la performance), mais aussi au processus qui, plus que le produit, donne quelque assurance sur la stabilité de la compétence. C'est dire que les ressources dont nous disposons actuellement, qu'il s'agisse des banques d'outils de la DEPP ou, sauf exception, des épreuves mises au point dans le cadre d'EVAPM, ne suffisent pas à répondre à ces exigences.

Pour bien montrer que les questions de vocabulaire, pour importantes qu'elles soient, restent secondaires par rapport aux idées qu'elles recouvrent, nous ne résistons pas à l'envie de rajeunir et de représenter ici un schéma que nous avons publié pour la première fois dans ce même bulletin, il y a 20 ans [8]. Le schéma était déjà titré « *des*

capacités aux compétences »⁽²⁰⁾, c'est dire que le thème n'est pas nouveau, même si on y voit aujourd'hui un peu plus clair qu'il y a 20 ans.



Le schéma met en évidence que les compétences mettent en jeu des connaissances et des schèmes regroupés ici dans ce qui est nommé processus psychologiques de base et compétences transversales. Il légitime en effet la référence aux compétences transversales, qui relèvent de l'axe horizontal et il n'exclut pas qu'une compétence mathématique implique un ensemble d'éléments de connaissances ne se limitant pas aux mathématiques ; de plus il ouvre sur les compétences impliquant plusieurs disciplines. De ce fait, ce schéma est de nature à faciliter le dialogue entre disciplines.

En ce qui concerne le socle commun, le décret précise d'ailleurs que :

« Chaque compétence qui le constitue requiert la contribution de plusieurs disciplines et, réciproquement, une discipline contribue à l'acquisition de plusieurs compétences. ».

Le schéma met encore en évidence le fait que, une fois une compétence identifiée, l'évaluation peut porter, sans déchoir, sur certains éléments constitutifs de la compétence (connaissances particulières, éléments transversaux, ...). En effet, l'idée qu'il faudrait n'évaluer les compétences que dans leur plénitude ne nous semble pas tenir davantage que l'idée qu'il faudrait enseigner, directement, les compétences. C'est dire que les enseignants devraient non pas focaliser l'enseignement sur les compétences, mais, comme nous l'avons écrit plus haut, les garder en tête comme un fil conducteur, tant pour la formation que pour l'évaluation. Pour cela, ils n'ont pas à changer radicalement l'ensemble de leurs pratiques, mais simplement à les repenser et à les réajuster pour mieux intégrer l'idée de compétence.

(20) Le lecteur trouvera le schéma d'origine en annexe.

Répetons toutefois que la bonne volonté ne suffira pas. Il faudra des formations sérieuses, appuyées par des recherches sérieuses produisant les outils nécessaires tant à la formation qu'à l'action ; recherches et productions auxquels les enseignants devront être soigneusement associés.

4.2. La question des attitudes

Il est possible que certaines attitudes soient constitutives de certaines compétences, mais, dans l'ensemble, elles sont plutôt du côté de l'engagement, de la volonté de développer des compétences ou de mettre en œuvre celles qui sont acquises.

« *L'ouverture aux autres, le goût pour la recherche de la vérité, le respect de soi et d'autrui, la curiosité et la créativité* » sont certainement des qualités que l'éducation doit chercher à développer, mais faut-il pour autant, comme cela est fait dans le décret définissant le socle commun, les considérer comme constitutives de compétences ?

Une comparaison rapide des attitudes dont le développement est préconisé, respectivement, dans le cadre des systèmes finlandais et français, suffit à montrer le caractère subjectif et idéologique des attitudes privilégiées.

D'une façon générale, les attitudes voisinent avec les valeurs morales et il convient de se méfier de leur intégration aux compétences.

| Attitudes France versus Finlande | |
|---|--|
| Socle commun - FRANCE | Objectifs officiels - FINLANDE |
| Attitudes : | Attitudes : |
| L'étude des mathématiques permet aux élèves d'appréhender l'existence de lois logiques et développe : | L'école doit faire tout ce qui est possible pour que tous les élèves développent : |
| - la rigueur et la précision ; | - un sens de la curiosité et le désir d'apprendre ; |
| - le respect de la vérité rationnellement établie ; | - leur voie personnelle d'apprentissage ; |
| - le goût du raisonnement fondé sur des arguments dont la validité est à prouver. | - leur confiance dans leurs propres compétences |

Pour rester dans le domaine de l'évaluation, la décision de noter les attitudes comportementales comme cela a été introduit au collège, allant jusqu'à faire compter la « note de vie scolaire » pour l'examen national du brevet, est difficilement acceptable.

Mettre ainsi les compétences en concurrence avec les attitudes est une pratique courante dans les régimes totalitaires ; c'est certainement une erreur que d'avoir, dans notre pays, cédé à cette facilité.

Il ne s'agit pas de dire que les attitudes ne sont pas importantes et que les établissements scolaires n'ont pas à s'organiser pour que les enseignants soient respectés, pour que les élèves se respectent entre eux, pour que la violence soit exclue, pour que les élèves s'investissent dans les activités d'apprentissage qui leur sont proposées, ... Il y a le règlement intérieur pour cela, voire la Loi ; rien n'empêche de renforcer l'un et d'appliquer plus systématiquement l'autre (ce qui d'ailleurs est fait, et bien fait, dans nombre d'établissements) ; ce qui est en cause ici est le fait de mettre sur le même plan, par l'intermédiaire de la note, des choses qui participent de registres très différents. La définition des compétences et l'évaluation des compétences a beaucoup à perdre à ce mélange délétère.

4.3. La question du socle et des livrets de compétences

Les livrets de compétences sont assez fréquents dans divers systèmes éducatifs. Ils sont souvent cause de désagrément, de rancœurs, et finalement de rejet de la part des divers acteurs et partenaires (c'est le cas au Québec et, en partie, en Suisse Romande). Pour les enseignants, ils sont accusés de nécessiter une gestion trop lourde, ou encore de détourner des tâches d'enseignement et de faire porter une attention démesurée sur le contrôle des acquis ... avant même qu'il y ait le moindre acquis à contrôler.

On peut encore leur appliquer ce jugement bien exprimé dans [19] : « *À cela s'ajoute parfois la confusion entre évaluation formative et bilan. L'engagement dans l'évaluation des compétences s'est souvent traduit par la confusion entre le suivi de l'élève dans ses apprentissages et le bilan de ses compétences acquises à certains paliers. On ne sait alors sur quoi renseigne réellement l'évaluation : les progrès de l'élève ou ses acquis.* »

Certains textes officiels semblent favoriser cette dérive (la contradiction des textes officiels a déjà été soulignée) : livret de compétences pour suivre et favoriser les acquis, ou livrets de compétences pour enregistrer les acquis ? L'un et l'autre auraient leur intérêt, mais pas pour les mêmes destinataires ; la dilution de ces deux types de livrets dans un même document ne pourra que brouiller la communication.

La notion de portfolio est parfois évoquée. Le portfolio, un peu comme le « press book » du photographe, permet de faire choix, avec la participation de l'élève, de ses travaux les plus significatifs et de les conserver comme témoignage de sa progression. On voit bien l'avantage qu'il y a à donner ainsi de l'importance aux travaux des élèves et à ne pas en faire des générateurs de notes dont l'intérêt s'efface dès que la note est enregistrée ailleurs. Le portfolio est utilisé avec un certain succès dans plusieurs pays ; il peut être lié à un livret de bilan de compétences très succinct (une page ou deux devrait suffire pour l'ensemble des disciplines).

Sans crainte de nous répéter, insistons sur l'idée que cela suppose qu'il soit fait confiance aux enseignants, que leur professionnalisme soit reconnu et qu'ils soient formés et outillés.

4.4. Le socle commun et le programme PISA... Quelles relations ?

Le socle se réfère directement au programme PISA. Nous savons que ce programme évalue à 15 ans les compétences qui sont supposées nécessaires pour tous les citoyens. Le socle « *ciment de la nation* », selon le décret qui le définit, s'adresse aussi à tous et est facilement présenté comme un minimum vital. La nation se donne pour objectif, non que son élite maîtrise le socle, mais bien qu'il soit maîtrisé par la très grande majorité des jeunes.

Aujourd'hui, à la vue de certaines tentatives d'opérationnalisation, il n'est pas certain que l'ensemble de l'élite (ou soi-disant telle) maîtrise le socle. En tout cas, l'auteur de ces lignes ne le maîtrise certainement pas !

PISA a défini, et décrit, pour chacun de ses volets et de ses champs, des niveaux de compétences allant de 1 à 6. Le zéro n'étant pas politiquement correct, elle a cependant été contrainte à faire apparaître le niveau « inférieur à 1 » !

D'après les tableaux suivants, dans l'ensemble des pays de l'OCDE, 45 % des jeunes de 15 ans sont au niveau 2 ou en dessous pour l'ensemble du volet mathématique ; ils sont 36% en France (données PISA 2003). Dans ces conditions, il serait peu raisonnable et peu opératoire que notre socle suppose un niveau PISA supérieur à 2.

| Répartition des élèves suivant les niveaux de compétence PISA | | | | | | | |
|---|---------------|-----|-------|-------|-------|-------|------|
| Élèves de 15 ans - Données 2003 | | | | | | | |
| Niveau | Inférieur à 1 | 1 | 2 | 3 | 4 | 5 | 6 |
| OCDE | 10% | 15% | 20% | 20% | 20% | 10% | 5% |
| FRANCE | 5,6% | 11% | 20,2% | 25,9% | 22,1% | 11,6% | 3,5% |

Or voici comment est défini le niveau 2 pour les mathématiques :

« Au niveau 2, les élèves peuvent interpréter et reconnaître des situations dans des contextes qui ne demandent pas plus que d'effectuer des inférences directes. Ils n'ont à puiser les informations pertinentes que dans une source d'information unique et peuvent se limiter à un seul mode de représentation. Ils sont capables d'utiliser les algorithmes de base, des formules, des procédures ou des conventions élémentaires. Ils peuvent se livrer à un raisonnement direct et interpréter les résultats de manière littérale. »

Précisons que les niveaux en question ne sont pas construits *a priori*, mais sont produits par l'analyse des données de l'évaluation. Il s'agit là de faits, sans doute regrettables, sans doute surmontables en y mettant le prix, mais de faits dont il faut tenir compte si l'on veut parvenir à une opérationnalisation raisonnable du socle.

5. La dissonance

On peut considérer l'évaluation comme le processus régulateur des relations entre l'enseignement et l'apprentissage.

Au niveau infra, l'évaluation renseigne l'enseignant d'un côté, l'élève de l'autre, sur ce qui a été compris, sur un certain état de savoir, et permet à chacun d'orienter son action dans un sens favorable à l'atteinte de ses objectifs. Évidemment, du côté de l'élève, la régulation en question jouera de façon différente selon que l'objectif premier, sinon unique, de l'élève, est d'avoir une note suffisante et d'éviter les ennuis, ou s'il est d'acquérir un savoir dont il a compris les enjeux⁽²¹⁾. Du côté de l'enseignant, la régulation jouera aussi différemment selon que l'évaluation qu'il mène est largement, ou totalement, orientée vers la satisfaction des demandes administratives, ou si elle est centrée sur le recueil d'informations, sur le processus d'apprentissage en cours et sur les éventuelles difficultés rencontrées.

Au niveau macro, l'évaluation du système contribue à la régulation de la relation enseignement-apprentissage, en particulier par l'influence qu'elle exerce sur les options prises par les décideurs.

Entre ce niveau infra, situé ici au niveau de la relation enseignant-enseigné, et le niveau macro situé au niveau du système, il y a toute la cascade des niveaux qui,

(21) On pourrait parler là de motivation intrinsèque (l'étude pour les enjeux de savoir) versus motivation extrinsèque (travailler pour la note). Dans un enseignement pour tous, obtenir une motivation intrinsèque de tous les élèves peut rester un objectif ... à long terme !

allant de l'élève au système, passent par la classe, l'établissement, la discipline, les dispositifs de validation, de certification, ... et les examens.

Notre thèse est que la cohérence maximale doit être recherchée entre ces différents niveaux d'évaluation et que, à l'évidence, pour le moment, c'est plutôt l'incohérence qui prévaut.

Dans une étude comparative de divers systèmes, dont le système français, Paul Black estime que :

« Ce qu'il y a peut-être de plus remarquable à propos du système (éducatif) français est la pléthore de systèmes d'évaluation et de contrôle des acquis, chacun d'eux étant conçu pour remplir un ensemble très limité de fonctions » [4]⁽²²⁾

Mais, qu'il s'agisse du système français ou d'autres, les différentes études soulignent un manque de cohérence dans l'ensemble des actions d'évaluation (la dissonance). Un peu comme si chacun tirait sa charrette sans s'occuper des charrettes des autres et sans percevoir de finalité commune.

Malgré, en effet, cette pléthore d'évaluations, il semble que l'on sache en général bien peu de choses sur les acquis des élèves (avec EVAPM toutefois, cela est loin d'être exact pour les mathématiques).

« L'opacité des modalités de construction des moyennes, la grande hétérogénéité des modes de communication aux familles et aux élèves, suivant les cycles, les disciplines, les professeurs, les établissements, l'absence de transmission des évaluations d'un niveau à l'autre, tout contribue donc à faire de la connaissance des acquis des élèves un des grands mystères de notre école. » [19]

L'accord se fera facilement sur le fait que cette pléthore d'évaluation est peu ou mal coordonnée.

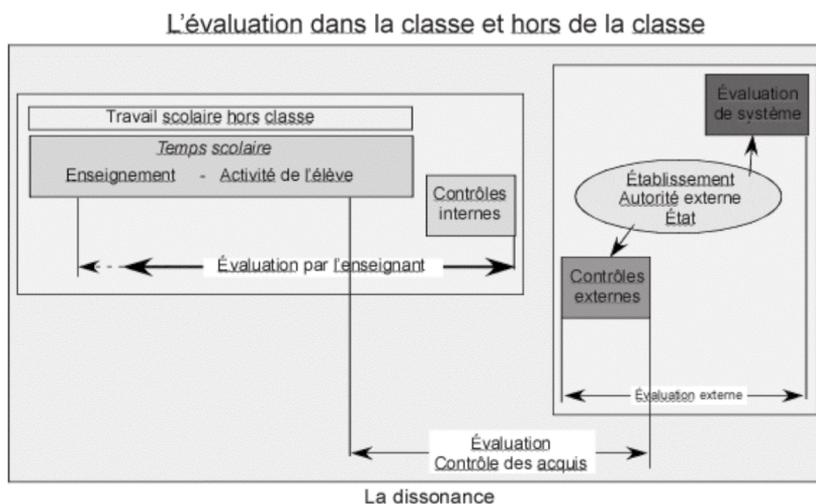
« La généralisation de cette pratique⁽²³⁾ se heurte toutefois à d'importants obstacles, notamment les tensions perçues entre les évaluations formatives intervenant en classe et les tests sommatifs à forte visibilité destinés à amener les établissements à répondre des résultats des élèves, ainsi qu'un manque de cohérence entre les démarches d'évaluation des systèmes, des établissements et des classes. » [17]

Les actions sont juxtaposées, comme essaie de l'illustrer le schéma suivant, sans qu'il y ait d'autre synergie que négative, tout cela accompagné de malentendus et de frustrations.

En regardant ce schéma, qui met en évidence les différents temps et modes d'évaluation portant sur les acquis des élèves ou leurs apprentissages, on peut évidemment penser qu'il y a trop ou même beaucoup trop d'évaluations dans notre système et s'interroger sur le manque de coordination de ces évaluations.

(22) Pour éviter les pièges de la traduction : « *What is perhaps most remarkable about the French system is the plethora of different assessment and evaluation systems, each designed to serve a very limited range of functions* ».

(23) Il s'agit de l'évaluation formative.



6. Quelle cohérence rechercher ?

6.1. Quelle perméabilité entre l'évaluation formative et l'évaluation de validation ?

La question a déjà été évoquée plus haut. Les informations recueillies dans le cadre de l'évaluation formative ont pour vocation d'être rendues caduques par l'action formative qui la suit. Chacun voit donc l'anomalie qu'il y a à utiliser telles qu'elles ces informations pour des jugements qui engagent au delà de l'action formative en cours.

Certains chercheurs ont pu prôner une déconnection complète de l'évaluation formative et de l'évaluation que nous appelons ici « de validation ». Cette position est partiellement tenable dans les classes d'examens ou de concours dans lesquelles l'enseignant peut être vu comme un guide qui se donne pour objectif de faire réussir au maximum l'ensemble de ses élèves ; l'examen ou le concours d'origine externe venant alors, dans une certaine mesure, sanctionner en même temps les acquis des élèves et l'action du professeur. On sait cependant que, même dans ce cas, la déconnection est loin d'être parfaite. Les notes et appréciations des professeurs des classes d'examen continuent, on le sait, à jouer un rôle important dans des décisions dont l'importance dépasse souvent celle de l'examen lui-même.

Dans les autres cas, la non-perméabilité entre l'évaluation formative et l'évaluation de validation ne peut être assurée qu'au prix d'une certaine schizophrénie des enseignants. Chacun connaît des cas où la note trimestrielle, ou annuelle, donnée par un enseignant vient en contradiction avec le jugement porté par ce même enseignant, lequel au demeurant est angoissé par le fait que c'est sa note qui, en fin de compte, sera prise en considération plutôt que son jugement. Dans ce domaine la pratique des moyennes de tout et n'importe quoi, encouragée par le recours systématique à l'informatique, n'arrange rien.

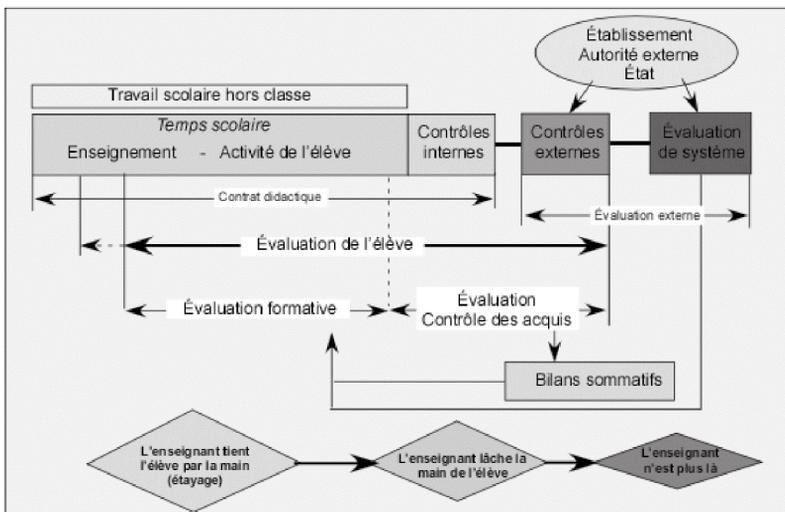
Au cours de la formation, les notes, si elles sont utilisées, ne devraient être considérées que comme des indicateurs commodes, mais partiels, et comme des messages [12] destinés aux élèves (éventuellement à d'autres personnes concernées). Le jugement de l'enseignant s'appuie non sur des notes, mais sur les observations qu'il a pu faire et dont certaines ont pu être résumées par une note. Parmi ces observations, il sait que certaines étaient contingentes, que d'autres ont amené à des corrections réussies, que d'autres encore pouvaient être assurées d'une certaine pérennité. En professionnel, l'enseignant est (doit être ?) en mesure de faire la part des choses et d'émettre un jugement de synthèse qui, à un moment donné, reflète correctement non seulement l'état des acquis, mais aussi la dynamique des apprentissages des élèves.

L'étude de l'OCDE sur l'évaluation formative prône un meilleur « alignement » de l'ensemble des actions d'évaluation. Mais cet alignement, dans certains cas, se réalise par la primeur donnée à l'évaluation du système, au détriment de l'évaluation formative. C'est le triomphe de l'enseignement POUR l'évaluation. Dans d'autres cas, la confiance accordée aux enseignants et les outils dont ils sont pourvus leur permettent de pratiquer une évaluation formative efficace, tandis que l'évaluation bilan est réduite au strict nécessaire, utilise l'expertise des enseignants et s'appuie ponctuellement sur des instruments d'évaluation validés.

Avec un peu d'amusement, nous entendions récemment une jeune collègue déclarer avec véhémence : « *l'évaluation appartient aux enseignants !* ». Ce serait bien mal engager la réflexion sur la mise en cohérence de l'ensemble que de partir de cette idée. Non, l'évaluation n'appartient à personne et certainement pas à ceux qui la font. Elle est au service, d'abord des élèves, et par voie de ricochet, de la société.

La recherche de cohérence que nous proposons peut-être illustrée par le schéma suivant dont il est facile de voir la différence avec le schéma précédent.

L'évaluation dans la classe et hors de la classe



Aligner les actions d'évaluation - réduire la dissonance

La réduction de cette dissonance peut se faire par voie réglementaire, sans concertation (comme cela se fait déjà dans certains pays). C'est sans doute le meilleur moyen d'accroître l'effet « *enseignement POUR l'évaluation* » et de contribuer à la déresponsabilisation des enseignants.

L'autre voie, la seule qui soit acceptable à nos yeux, suppose une négociation continue entre les différentes parties concernées. Elle suppose aussi une bonne connaissance de l'état des pratiques et de l'univers des possibles.

6.2. La formation à l'évaluation

Au cours de leur formation initiale, 37% des enseignants de mathématiques déclarent avoir reçu « une formation sur le thème de l'évaluation des élèves » ; ils sont 21% en formation continue ; 34 % déclarent n'avoir jamais reçu de formation (et même avoir pratiqué la moindre autoformation). On notera que les chiffres correspondants sont respectivement de 71%, 48% et 5% pour les professeurs d'EPS [18].

Le chantier à ouvrir est donc important. Mais les formations devraient être préparées en faisant appel à toutes les ressources (didactique, sciences de l'éducation, mais aussi expertise des enseignants et autres acteurs). Surtout elles devraient éviter au maximum le cloisonnement catégoriel.

Pour une part, les formations devraient être des accompagnements de pratiques. C'est en construisant, utilisant, interprétant, des évaluations, que l'on apprend (un peu) à évaluer, ce n'est pas en écoutant des discours, aussi bien construits soient-ils.

6.3. Propositions pratiques

Il ne s'agit dans ce qui suit que de quelques points clés possibles présentés un peu en vrac... Ces points sont à discuter et à compléter ; ils ne prétendent évidemment pas clore la question. Interroger les conceptions et modifier les pratiques ne peut pas résulter de quelque injonction aussi bien argumentée soit-elle. L'ensemble suppose un réel effort de formation, que je préfère appeler ici de développement professionnel, et qui suppose volonté, moyens, engagement des acteurs, recherche et expérimentation.

6.3.1. Des conceptions qu'il faudrait interroger

- Évaluer ne signifie pas noter ; noter ne signifie pas évaluer !
- Corriger (le devoir d'un élève) ne signifie pas le noter, ni même l'évaluer, ni encore le couvrir de jugements de valeur (juste ! faux ! apprends tes leçons !...). (Corriger – du latin *corriger* : redresser).
- Ne pas confondre les activités d'apprentissage (où il peut être normal d'échouer) et les activités d'évaluation (où, pour celui qui a appris, il devrait être normal de réussir).
- Si notes il y a, il faudrait en finir avec la dictature de la moyenne ; si moyenne il y a, qu'elle soit considérée comme un indicateur, non comme une mesure.
- La note peut garder son intérêt comme message (ou code) ; mais le sens du message doit être explicité. Ce qui suppose en particulier l'explicitation des objectifs.
- L'évaluation au quotidien, évaluation qui ne devrait avoir qu'une fonction formative, n'a aucune raison d'être identique pour tous les élèves d'une classe.

- Par contre l'évaluation sommative gagne à être référée à des « standards » qui échappent partiellement, ou totalement, au contrôle direct de l'enseignant. Ainsi l'évaluation formative devrait être personnalisée tandis qu'une certaine égalité de traitement, dépassant le cadre de la classe et de l'établissement, serait associée à l'évaluation de validation.
- Le point précédent suppose des référentiels communs, des standards, mais aussi des instruments d'évaluation partagés (banques d'épreuves).
- Il y a trop d'évaluation(s) dans notre système éducatif (?).
- Il n'y a pas assez d'évaluation(s) dans notre système éducatif (on ne sait rien sur ce que savent les élèves) (?).

Les deux sentiments précédents peuvent être simultanément ressentis, ainsi qu'en témoigne le rapport de l'IGEN [18].

6.3.2. Pratiques qu'il serait possible de promouvoir

- Mieux distinguer le travail des élèves nécessaire à leurs apprentissages de la part de leur activité qui répond à des demandes d'évaluation (distinction situation d'apprentissage – situation d'évaluation).
- Ne pas confondre enseigner et évaluer. L'évaluation formative est partie intégrante de l'acte d'enseignement, mais le souci de l'évaluation ne doit pas prendre le pas sur le souci d'enseigner (et de guider les apprentissages).
- Distinguer l'évaluation du travail des élèves de l'évaluation de leurs résultats. Distinguer le processus d'avancée dans le savoir (parfois, malheureusement, de recul), donc les progrès, de l'état des acquis. S'intéresser aux processus et pas seulement aux effets.
- Contrôler et limiter soigneusement la perméabilité entre l'évaluation formative et l'évaluation sommative ou certificative pour que le souci de rendre des comptes ne vienne pas s'ériger en obstacle au bon fonctionnement de l'évaluation formative.
- Rechercher la cohérence, voire un « alignement » des actes d'évaluation mis en place par des acteurs différents (évaluations conduites de façon autonome par l'enseignant, évaluation concertée au sein d'un établissement, évaluations externes, ...).
- Cette mise en cohérence suppose un climat de confiance entre élèves, parents, enseignants, encadrement, autorités académiques, ... (climat de confiance qui suppose l'instauration d'une culture de l'évaluation à laquelle il nous faut travailler !)

7. Conclusion

Après plus de 30 ans de réflexion, d'action, et de recherche en matière d'évaluation, d'échanges et de travail avec de nombreux collègues⁽²⁴⁾, j'ai essayé de livrer dans cet article l'état de mes réflexions. Le lecteur voudra bien excuser le caractère à la fois un peu personnel du discours, ses aspects qui pourront sembler parfois trop affirmatifs, trop critiques, voire donneurs de leçons.

(24) auxquels je suis redevable d'une grande partie des idées exprimées dans ce texte.

Il me semble que nous sommes à un tournant : ou bien la communauté mathématique dans laquelle j'inclus évidemment les enseignants de mathématique se saisiront, collectivement de la question, ou bien des solutions et des outils leur seront imposés par des moyens qui nieront de fait leur professionnalisme et qui ne feront qu'accroître la dissonance dénoncée dans cet article.

Références

- [1] Antibi, A. (2003) : La constante Macabre ou comment a-t-on découragé des générations d'élèves ? Éditions Math'Adore.
- [2] APMEP (1987 à 2007) : Rapports EVAPM.
- [3] Balian, R & al. (2004) : Les savoirs fondamentaux au service de l'avenir scientifique et technique. Comment les réenseigner. Les Cahiers du débat. Fondation pour l'innovation politique.
- [4] Black, Paul and Wiliam, Dylan (2005) : Lessons from around the world : how policies, politics and cultures constrain and afford assessment practices. Curriculum Journal, 16 : 2, 249 – 261.
- [5] Bodin A (1982) : Objectifs et évaluation, vol 1 : Irem de Besançon.
- [6] Bodin A. (1997) : L'évaluation du savoir mathématique — Questions et méthodes. Recherches en Didactique des Mathématiques, Éditions La Pensée Sauvage, Grenoble.
- [7] Bodin, A. (2006) : Les mathématiques face aux évaluations nationales et internationales. De la première étude menée en 1960 aux études TIMSS et PISA ... en passant par les études de la DEP et d'EVAPM. Communication séminaire de l'EHESS. Repères IREM, n° 65, octobre 2006.
- [8] Bodin, A. (1989) : L'évaluation du savoir mathématique, Bulletin de l'Association des Professeurs de Mathématiques, 368, p. 195-219, Paris.
- [9] Bodin, A. (2006) : En français : Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français. Communication faite à la conférence Franco-Finlandaise sur PISA. Bulletin de l'APMEP n° 463 (disponible en langue anglaise).
- [10] Bonnet, G. (1997) Country profile from France, Assessment in Education, 4 (2), 295–306.
- [11] Bottani, N. ; Vrignaud, P. (2005) : La France et les évaluations internationales. Haut Conseil de L'Évaluation de l'École.
- [12] Chevallard, Y., Feldmann, S. (1986) : Pour une analyse didactique de l'évaluation. Irem de Marseille.
- [13] Crahay, M. (1996) : Peut-on lutter contre l'échec scolaire, Bruxelles, Éditions De Boeck.
- [14] DEP-MEN (Braxmeyer & al.) (2004) : Les pratiques d'évaluation des enseignants en collège. Dossier DEP.
- [15] EARLI (European Association for Learning and Instruction) (2005) : A Learning Integrated Assessment System. Position paper (texte destiné aux responsables, aux chercheurs, aux formateurs, ... et qui exprime des recommandations concernant l'évolution des politiques évaluatives).

- [16] Gauthier, R.F. (2006) : Évaluation des acquis des élèves, évaluation de système, où en est-on en France en 2006 ?
- [17] Glaeser G. (1995) : Fondements de l'évaluation en mathématiques. APMEP, PARIS.
- [18] IGEN (Bardi, A. M. & al.) (2005) : Les acquis des élèves, pierre de touche de la valeur de l'école ? Inspection Générale de l'Éducation Nationale. Rapport à Monsieur le Ministre de l'Éducation Nationale.
- [19] IGEN (Houchot, A., Robine, F. & al.) (2007) : Les livrets de compétences : nouveaux outils pour l'évaluation des acquis. Inspection Générale de l'Éducation Nationale. Rapport à Monsieur le Ministre de l'Éducation Nationale.
- [20] Mercier, A. & Buty, C. (2004) : Évaluer et comprendre les effets de l'enseignement sur les apprentissages des élèves : problématiques et méthodes en didactique des mathématiques et des sciences. Revue Française de Pédagogie, n° 148, p. 47-59. INRP- Paris
- [21] OCDE (2005) : L'évaluation formative. Pour un meilleur apprentissage dans les classes secondaires (CERI – Centre de Recherche pour l'Innovation dans l'Enseignement).
- [22] Posthumus, K. (1947). Levensghell en school. La Haye.
- [23] Site de Jacques Nimier : <http://perso.orange.fr/jacques.nimier/>
- [24] Vergnaud G. (1990) : La théorie des champs conceptuels. Recherches en Didactique des Mathématiques Vol 10/ 2.3. Éd. La pensée sauvage, Grenoble.

Contact et site internet

Antoine Bodin : antoinebodin@mac.com

Site Internet : <http://web.mac.com/antoinebodin/>

Le lecteur trouvera sur ce site, aux pages « ÉVALUATION », « COMPÉTENCES » ET « SOCLE » de nombreux documents et diaporamas complémentaires de cet article.

Annexes

1. Résumé des recommandations issues de l'étude OCDE sur l'évaluation [21]

- Importance de sécuriser les élèves pour que ceux-ci osent prendre des risques et faire des erreurs en classe...
- Attention des élèves sur la maîtrise des tâches plutôt que sur la compétition avec les pairs.
- Développement de la motivation intrinsèque.
- Renforcement de l'estime de soi.
- Confiance des élèves dans leurs capacités scolaires.
- ...
- Les résultats s'améliorent lorsque les élèves travaillent sur des objectifs de processus plutôt que sur des objectifs de produits.

- Les notes peuvent saper l'aide positive du feedback ciblé sur les tâches.
- L'une des missions cruciales des enseignants est d'aider les élèves à prendre confiance en eux et à élaborer diverses stratégies d'apprentissage.
- Instauration d'une culture de classe qui encourage l'interaction et l'utilisation d'outils d'évaluation.
- Définition d'objectifs d'apprentissage et suivi des progrès individuels des élèves vers ces objectifs.
- Utilisation de méthodes d'enseignement variées pour répondre aux besoins diversifiés des élèves.
- Recours à des méthodes diversifiées pour évaluer les acquis des élèves.
- « Feedback » sur les performances de l'élève et adaptation de l'enseignement pour répondre aux besoins identifiés.
- Implication active des élèves dans le processus d'apprentissage.

2. Six principes d'action sont proposés :

- Rester ciblé sur l'enseignement et sur l'apprentissage.
- Aligner les démarches d'évaluation formative et sommative (appelée ici « de validation »).
- Veiller à lier les données recueillies dans les classes, dans les établissements et dans les systèmes et à les utiliser de manière formative pour apporter des améliorations à chaque niveau.
- Investir en formation et en soutien à l'évaluation formative.
- Encourager l'innovation.
- Renforcer les liens entre la recherche, l'action publique et la pratique.

3. Schéma du 4.1 sous sa forme originale (1987)

