

Conférence de Pierre MERLE

Retranscription

Les titres et les tableaux ont été ajoutés. Ils sont issus du support de présentation accompagnant l'exposé de Pierre MERLE.

L'évaluation des élèves: outil d'apprentissage au service de l'efficacité et de l'équité

Bonjour à tous,

Mon propos va donc porter sur l'évaluation des élèves. Le thème de cette journée porte sur l'équité. En l'occurrence, il s'agit d'équité et aussi – je l'ai ajouté dans le titre – d'efficacité. Équité: réduction des écarts de compétences entre les élèves; efficacité: augmentation du niveau moyen.

Souvent en France, peut-être aussi en Belgique, on associe l'équité (la réduction des écarts) à un nivellement vers le bas et une moindre efficacité. Les recherches montrent l'inverse, et pour ce qui est de l'évaluation, il en est de même.

Je voudrais indiquer d'abord la centralité de l'évaluation. C'est une question qui est vraiment transversale, qui va de l'école à l'université, pour un certain nombre de raisons:

- la première, c'est que l'évaluation, et la notation qui va avec, est une obligation institutionnelle qui s'impose à tous les acteurs, tous les professeurs de l'école;
- la seconde, c'est que, évidemment, pour les écoliers, les élèves et les étudiants, la note est d'une importance extrême. C'est à peu près l'équivalent de leur salaire, c'est leur prestige à l'école, c'est aussi leur reconnaissance vis-à-vis des parents. Il s'agit d'un terrain très sensible pour ceux qui reçoivent des notes;
- et puis, enfin, il y a une forte attente des institutions éducatives et des parents à l'égard de l'évaluation des compétences. On le sait, particulièrement en Belgique ou en France où le redoublement est important, c'est l'évaluation qui est un élément majeur de cette décision, et puis on sait – on en a parlé ce matin – que nous sommes dans un système où la réussite scolaire est en quelque sorte la clé de la réussite sociale via la réussite aux examens, donc l'évaluation. On se rend compte à quel point il s'agit d'une question centrale.

Alors, je voudrais, pour vous parler de l'évaluation, vous parler d'abord des élèves. Je pense qu'il faut partir de leur expérience de la notation; puis, vous présenter la conception institutionnelle de la notation et les recherches docimologiques, les recherches qui sont faites depuis longtemps sur la fiabilité de la note, notamment. Après, j'aimerais vous présenter les principaux résultats de la psychologie et de la sociologie de la notation. Ensuite, ce qui me paraît tout à fait important, c'est l'approche interactionniste de la notation. Je pense qu'on a des idées assez inexactes sur la pratique effective de la notation dans la classe. Et enfin, je dirai quelques principes d'une notation au service de l'efficacité et de l'équité.

Vous voyez, un programme chargé... Je ne sais pas si j'arriverai au bout!

1. Les élèves: quelle expérience de la notation?

Je vais vous présenter très rapidement deux enquêtes que j'ai menées auprès de lycéens: seconde, première et terminale, c'est-à-dire équivalant à la 4^e, 5^e et 6^e pour vous. Et puis, même type d'enquête sur les collégiens.

Les enquêtes portaient sur le rapport des élèves aux notes, leur confiance dans l'exactitude de la note, la marge d'erreur selon les disciplines, les notes en cas de double correction, et puis une question ouverte à la fin: "*Désirez-vous ajoutez quelque chose?*".

Cette enquête a permis de montrer l'existence, chez les collégiens et les lycéens, de trois types d'expérience de la notation.

Premier type d'expérience de la notation: ce que l'on peut appeler la sérénité.

Je vais vous inquiéter tout de suite: la sérénité, c'est une expérience peu fréquente, malheureusement, chez les élèves!

Voilà comment on peut résumer cette expérience des écoliers, des collégiens et des lycéens:

- la note est une mesure fiable des compétences des élèves;
- ces élèves adhèrent à un principe d'exactitude de la note, et puis d'impartialité du professeur;
- les évaluations externes dont il a été question tout à l'heure dans le petit film, les épreuves anonymes ne peuvent que confirmer la valeur du candidat. En gros, elles ne servent à rien, on peut faire toute confiance au professeur. Donc, il y a une sorte de principe d'universalité de l'exactitude du jugement.

Cette sérénité, elle est relativement peu fréquente. J'en reparlerai tout à l'heure, pour indiquer des éléments de fréquences.

Deuxième expérience des lycéens: l'insatisfaction.

Cette insatisfaction, elle se traduit dans les commentaires ouverts par des propos de ce type: "*Je pense que mes notes de bac ont été très injustes; je pense qu'au bac, une double correction serait nécessaire afin d'éviter les sautes d'humeur des correcteurs*".

Ces élèves, finalement, sont insatisfaits parce qu'ils considèrent qu'il y a des dysfonctionnements. Par exemple: surcharge du correcteur, impartialité du correcteur, mauvais barèmes... ce qui aboutit à une mauvaise évaluation de leurs compétences.

Mais ces élèves, et c'est l'essentiel, adhèrent au principe d'une notation équitable. La note juste, la note équitable est possible, mais voilà, je n'ai pas eu de chance, l'évaluation s'est mal déroulée.

Une troisième expérience, qui est plus problématique pour le professeur et peut-être encore plus pour l'élève, c'est l'impossible équité.

Voilà un exemple: "*Les notes sont variables d'un professeur à l'autre, d'un établissement à l'autre, et ne sont pas nécessairement représentatives du potentiel d'un élève*".

En fait, ces élèves adressent des critiques fondamentales au système de notation: les notes sont dépendantes d'aléas, et ce sont des aléas qui sont intrinsèquement liés au projet scolaire de mesurer la compétence de l'élève. La note est dépendante d'un contexte d'examineur, d'établissement, de jury.

Ces élèves font, à leur idée en fait, des expériences docimologiques: ils font réaliser leurs devoirs par leurs parents, par des professeurs amis, des grands frères qui sont dans l'enseignement supérieur. Et ils constatent avec inquiétude, et en même temps presque satisfaction, que leur intuition – c'est-à-dire qu'ils sont "*catalogués*", c'est le terme qu'ils emploient – est malheureusement vérifiée, puisque leurs parents, *a priori* plus compétents, ne disposent pas de notes supérieures aux leurs.

Je me suis livré moi-même, en tant qu'agrégé en sciences économiques et sociales, à ce type d'expérience. Je préfère ne pas vous dire la note que j'ai eue!

Le niveau scolaire de l'élève influence le correcteur. Voilà les idées de cette expérience.

Quels sont les déterminants de cette expérience de la notation?

D'abord, la discipline d'enseignement.

En mathématiques, la moitié des élèves pensent que la note est indépendante du correcteur. Ils ont cette idée qu'en maths, c'est bon ou ce n'est pas bon.

En français, le sentiment d'une note fiable est beaucoup moins net: à peine plus d'un élève sur dix partage ce sentiment d'une note stable.

Et puis, ce que l'on constate, c'est que, de la 6^e à la terminale, de la 1^{re} à la 6^e, le pourcentage d'élèves adhérant à la note juste est stable en maths. Il y a des conditions propres à la discipline qui permettent cette stabilité de la confiance des élèves. Mais cette confiance diminue progressivement en français.

Pourquoi? Probablement, je pense, en raison de l'importance du barème qui présente au moins l'avantage pour l'élève, et un avantage très important, de connaître les exercices peu ou pas réussis, où se trouvent les erreurs. Ce qui est quand même assez légitime, pour un élève qui a une mauvaise note, de savoir où se trouvent les erreurs, et ce n'est pas une information dont il dispose toujours dans certaines disciplines.

Un autre élément déterminant, c'est le niveau scolaire de l'élève.

Et la logique est assez simple: plus les élèves ont un bon niveau scolaire, plus ils adhèrent à l'idée d'une notation exacte. On comprend que lorsqu'on vous attribue des notes excellentes, on considère finalement que le professeur ne peut pas se tromper. Et de ce point de vue-là, le système fonctionne bien.

Évidemment, pour ce qui est des élèves faibles, qui partagent le point de vue inverse, on voit qu'on va se retrouver dans une situation difficile.

Il faut indiquer que cette relation est moins vérifiée pour les bons élèves en français qui, malgré leurs bonnes notes ou pas, ont une certitude, comme en mathématiques, sur leur niveau de compétences.

Quelles conclusions peut-on tirer de ces expériences scolaires de la notation?

D'abord, que l'expérience de la note en français et en mathématiques détermine le rapport à la discipline. En fonction d'un barème ou de l'absence de barème, les élèves ont le sentiment d'un lien entre travail et récompense. Et ceci est fondamental dans l'attitude de l'élève. S'il n'a pas l'impression qu'il y a un lien entre son travail et la récompense, on comprend qu'il sera dans une situation de démotivation. D'où l'importance de la justification de la note.

Cette expérience détermine aussi le rapport au maître, à sa légitimité, en fonction du sentiment d'impartialité qu'il inspire.

Je dirais que, globalement, l'expérience de la note va déterminer le niveau d'engagement scolaire de l'élève en fonction des sentiments de justice qu'il peut avoir sur les règles scolaires: récompense, impartialité du professeur.

On sait qu'en Belgique et en France, le niveau d'intégration à l'école est jugé mauvais dans les études internationales. Les élèves n'ont pas une grande confiance dans l'école; ils sont, par contre, en France comme en Belgique, bien intégrés à leurs pairs, mais parfois contre l'école.

2. Conception institutionnelle de la notation et recherches docimologiques

Je dirais que notre système, en gros, fonctionne comme si la note était juste. Et il me semble qu'il s'agit d'une croyance en une note juste puisqu'il y a des sentiments, des doutes sur cette croyance. Pourquoi?

Il y a de bonnes raisons, je dirais:

- les professeurs sont des experts, ils sont diplômés, ils sont lauréats de concours difficiles;
- ils maîtrisent les programmes officiels et les compétences enseignées à leurs élèves *a priori*;
- ils sont donc aptes à évaluer précisément les devoirs liés aux programmes d'enseignement.

Ceci constitue la structure, en quelque sorte, de la croyance en une note juste.

Quand on examine un peu plus avant les hypothèses de cette conception institutionnelle de la notation, on peut avoir des doutes puisque, pour qu'il y ait les mêmes types de notation, il faut:

- la même compréhension de la notation par les professeurs, ce qui ne va pas toujours de soi;
- la même lecture des copies;
- une sensibilité aux erreurs et des qualités: l'orthographe, la syntaxe, la lisibilité, la clarté...

En fait, on se rend compte qu'aucune de ces hypothèses sous-jacentes n'est pleinement vérifiée. Les professeurs ont leur histoire, ils ont un rapport au savoir en partie différent, ils ont des expériences qui sont variées, ils ont des exigences différentes selon le niveau de leurs élèves. D'un côté, il y a les principes. Puis dans les pratiques, on se rend compte que les conditions d'une note juste ne sont pas nécessairement réalisées.

On sait depuis longtemps, par des études de multi-correction, que les notes ne peuvent pas avoir la prétention d'évaluer d'une façon fiable les compétences des élèves. Le principe est à peu près toujours le même dans ces études. Il s'agit de prendre un paquet de copies, par exemple une centaine de copies dans plusieurs disciplines au baccalauréat, et on soumet ces paquets de copies à plusieurs correcteurs. Et après, les notes sont comparées entre elles.

Je vais vous donner quelques résultats. Une très vieille étude, c'est la première: 100 copies de baccalauréat, session de 1930. L'étude est signée LAUGIER et WEINBERG en 1936. Elle concerne 6 correcteurs pour chacune des disciplines. Voilà les résultats:

**Tableau 1: expérience de multi-correction menée sur 100 copies du baccalauréat (session 1930)
LAUGIER et WEINBERG, 1936**

Disciplines	Écart maximum	Moyenne des écarts	Écarts les plus fréquents
Français	13	3,3	6 et 7
Anglais	9	2,2	4
Mathématiques	9	2,1	4
Philosophie	12	3,4	5 et 7
Physique	8	1,9	4

En français, c'est la première ligne:

- écart maximum: 13 points; donc, une copie a eu "3" par un correcteur et "16" par un autre correcteur (C'est beaucoup!);
- moyenne des écarts: 3,3 points;
- écarts les plus fréquents: 6 et 7 points, donc c'est considérable!

Alors, on s'aperçoit que dans des disciplines proches, par exemple la philosophie, on a à peu près le même type de profil:

- écart maximum: 12 points;
- écarts les plus fréquents: 5 et 7 points.

Dans les disciplines dites exactes, on est surpris par l'ampleur des différences. La discipline qui est la plus fiable en termes de notation, c'est la physique, et vous voyez que l'écart maximum est encore de 8 points, la moyenne des écarts de 1,9 point et les écarts les plus fréquents de 4 points.

Les mathématiques ne sont pas mieux loties que la physique, et l'anglais fait plutôt bonne figure.

Donc, on s'aperçoit que les élèves n'ont pas tort quand ils disent que la correction en mathématiques est plus précise qu'en français. Mais ils n'ont pas raison quand ils pensent qu'en maths, c'est bon ou ce n'est pas bon. C'est plus compliqué!

Ces études ont aussi montré qu'on n'a pas d'accord des correcteurs sur les copies des élèves qu'il faut admettre ou non au baccalauréat.

Tableau 2: candidats refusés, admissibles, et refusés ou admissibles au bac selon les notes obtenues par les 6 correcteurs (en %) (session 1930)

Disciplines	Refusés par les 6 correcteurs	Acceptés par les 6 correcteurs	Refusés ou admissibles
Français	21	9	70
Anglais	37	16	47
Mathématiques	44	20	36
Philosophie	9	10	81
Physique	37	13	50

Par exemple, la proportion de candidats refusés par les 6 correcteurs:

- elle est très faible en philosophie: il y a seulement 9% des copies qui seraient refusées de façon unanime par les correcteurs;
- le pourcentage est de 44% en mathématiques.

Plus important, mais tout de même pas si important que ça: acceptés par les 6 correcteurs: 20% en mathématiques.

Et puis "refusés ou admissibles", c'est-à-dire des résultats qui vont dépendre du correcteur:

- 36% en mathématiques: c'est déjà beaucoup, pour une discipline qui se présente comme exacte;
- en philosophie, on s'aperçoit qu'il vaut mieux connaître le correcteur pour savoir si l'élève va être reçu que connaître la copie, puisque vous voyez que 81% des candidats sont refusés ou admissibles.

Alors, ce sont des résultats un peu inquiétants... On peut se dire que ce sont de vieux résultats et que tout a changé depuis, parce qu'on a beaucoup progressé. La réalité n'est malheureusement pas aussi réjouissante! Voilà une étude récente faite par un collègue, Bruno SUCHAUT en 2008, à partir de copies du bac en sciences économiques et sociales. Il y a 30 correcteurs pour chacune des copies.

Tableau 3: expérience de multi-correction de copies du bac en sciences économiques et sociales (années 2006 et 2007- 30 correcteurs) (SUCHAUT, 2008)

	Copie 1	Copie 2	Copie 3	Copie 4	Copie 5	Copie 6
Note au bac	9	11	15	9	15	11
Note minimum	5	5	8	3	8	4
Note maximum	15	16	18	13	17	14
Écart maximum	10	11	10	10	9	10
Moyenne	8,8	9,0	13,0	8,8	12,9	8,0
Écart-type	2,4	2,7	2,5	2,1	2,1	1,8
Mode	7	8	14	9	13	8

Regardons la **copie 1**:

- au bac: 9/20;
- note minimum par un des correcteurs: 5; note maximum: 15 (soit un écart de 10 points, et ce sont des copies du bac récentes!);
- moyenne: 8,8 points;
- écart-type: 2,4 points;
- note la plus fréquente: 7 points.

Donc, on s'aperçoit qu'on retrouve des éléments d'incertitude que l'on constatait déjà dans les études de 1930 sur la correction au baccalauréat.

Bon, on peut quand même avoir une petite satisfaction, si on regarde bien. Pour la copie 1, la moyenne au bac est de 9; la moyenne des copies des correcteurs est de 8,8. Donc, c'est assez proche. Et pour la copie 3: la note c'est 15, et la moyenne 13. On n'est pas si éloignés que ça!

Alors, on en vient à se poser des questions. C'est-à-dire: comment se fait-il que même en mathématiques, il puisse y avoir des divergences aussi grandes? On le sait en regardant précisément. C'est un professeur de mathématiques qui a mené l'expérience, il y a 20 ans exactement: précisément, comment est notée chacune des questions sur une copie?

Tableau 4: utilisation par trois professeurs de mathématiques d'un barème de notation sur une copie de baccalauréat de terminale scientifique (AYMES, 1979) – exemple: copie N°8

	Q.1 1 pt	Q.2 2 pts	Q.3 1 pt	Q.4 2 pts	Q.5 1 pt	Q.6 1 pt	Q.7 1,5 pt	Q.8 2,5 pts	Q.9 0,5 pt	Q.10 0,5 pt	Q.11 0,5 pt	Q.12 1,5 pt	Q.13 1 pt	Q.14 3 pts	TOTAL
Prof 1	1	2	1	0,5	0	1	1,5	2	0	0,5	0,5	0	0	1	12/20
Prof 2	1	2	1	2	0	1	1,5	2	0	0,5	0,5	0,5	0,5	1,5	14/20
Prof 3	1	2	1	2	0,5	1	1,5	2,5	0	0,5	0,5	1,5	0,5	2	16,5/20

Là, en l'occurrence, trois professeurs ont noté la copie 8: la question 1 était sur 1 point, tous les professeurs sont d'accord. Les premières questions sont généralement des questions faciles, et la correction est aussi facile à faire. Question 1, question 2, question 3: chaque fois, les professeurs sont d'accord.

Les questions à la fin sont des questions plus compliquées. On s'aperçoit, sur la question 12 qui est sur 1,5 point, que le professeur 1 a mis 0, le professeur 2 a mis 0,5 et le professeur 3 a mis 1,5. Donc, il y a une énorme différence! Ces différences cumulées les unes aux autres, certains professeurs étant plus sévères et d'autres plus indulgents, aboutissent même en mathématiques, si vous regardez le total: pour le professeur 1, la note est de 12/20; pour le professeur 3: 16,5/20. Soit 4,5 points de différence, avec de gros coefficients en terminale scientifique, ça fait largement la différence entre un reçu et un collé!

Donc, on s'aperçoit que même avec des barèmes, les différences sont moindres, mais elles sont extrêmes.

Alors, on en vient à se poser des questions. Et notamment, cette première question:

Les examens sont-ils des loteries?

Quand on regarde ces résultats, on a tendance à dire oui! En fait, c'est plus compliqué. Heureusement, dans une certaine mesure, parce que le résultat, c'est non. Pourquoi? Parce qu'il y a plusieurs épreuves, que les aléas de notation sont, par définition, répartis de façon aléatoire. Et finalement, puisque les élèves sont soumis à plusieurs correcteurs, à plusieurs disciplines, les mauvaises notes sont neutralisées par les bonnes notes, et on en arrive à un résultat qui est relativement cohérent. C'est-à-dire que les bons élèves sont massivement admis, les faibles sont très généralement collés, et puis l'aléa demeure pour les élèves moyens, mais de toute façon le contrôle continu, on avait déjà des aléas.

3. Psychologie et sociologie de la notation: la découverte des "biais"

On peut se poser d'autres questions: n'y a-t-il que de l'aléa dans la variation des notes d'une même copie? Et c'est ce qui m'amène à ce point 3 sur "la psychologie et la sociologie de la notation: la découverte des «biais»". Jusque là, on était sur des approches docimologiques, c'est-à-dire juste une approche statistique de la notation des comparaisons de notes. La psychologie et la sociologie de l'évaluation vont avoir des optiques un peu différentes. Les résultats principaux ont été établis par deux psychologues: NOIZET et CAVERNI, il y a déjà une trentaine d'années, en fait.

Le principe des expériences est toujours à peu près le même: il consiste, par exemple, à prendre deux copies (copie A et copie B). Sur la copie A, on met une indication, par exemple: "élève dont le père est un journaliste connu", et sur la copie B, "élève fils d'ouvrier". On fait 10 photocopies et on les envoie à 10 correcteurs, qui reçoivent donc les copies A et B. Puis après, on reprend une copie A et une copie B vierges, et sur la copie A où on avait mis "fils d'un journaliste doué", on met "fils d'ouvrier", et sur la copie B on met "fils d'un journaliste doué", on inverse les stéréotypes. On fait 10 photocopies et on envoie ça à 10 autres professeurs. Puis on compare les corrections. C'est le résultat que je vais vous présenter.

On s'aperçoit de quoi? Quand on met sur une copie A "copie d'un très bon élève" et sur une copie B "copie d'un très mauvais élève", on s'aperçoit que le professeur est influencé par le niveau scolaire antérieur annoncé. Donc, il y a un **effet du niveau scolaire antérieur annoncé**.

On s'aperçoit aussi que les correcteurs sont influencés par les notes qu'ils ont déjà données à leurs élèves. C'est un principe d'enquête un peu plus compliqué, mais le résultat est indiscutable.

On remarque aussi **l'effet du statut scolaire**, c'est-à-dire: lorsqu'on met sur la copie A un lycée prestigieux et un lycée de banlieue pour la copie B et puis qu'on inverse, l'effet du statut scolaire joue. Un élève qui vient d'un lycée prestigieux est meilleur pour les correcteurs qu'un élève qui vient d'un lycée de banlieue.

On a aussi un **effet de l'apparence physique**. Là, c'est un peu plus compliqué. Il y a des jurys qui choisissent des photos et puis qui les classent selon des apparences physiques plus ou moins agréables, et puis on met ces photos sur les copies et on s'aperçoit qu'il y a un effet de cette apparence physique.

Effet du genre: les copies du cours de physique associées fictivement à des garçons sont jugées plus rigoureuses et manifestant une meilleure aptitude pour la science.

Effet de l'origine sociale: c'est l'expérience dont je vous ai présenté le principe. Les copies attribuées fictivement à des élèves d'origine aisée sont mieux notées que des copies attribuées fictivement à des élèves d'origine populaire. Ceci vaut pour des dissertations, y compris pour l'orthographe où, *a priori*, il y a des règles d'appréciation des erreurs relativement fixes. En fait, non! À partir du moment où c'est le fils d'un journaliste doué, on voit moins les fautes d'orthographe que pour un fils d'ouvrier.

Ces "biais" sont démontrés par les recherches en psychologie, c'est-à-dire des recherches qui sont fondées sur des expériences. D'un point de vue statistique, c'est incontestable. Ceci dit, ce ne sont pas les types d'enquêtes que l'on mène en sociologie. En sociologie, on s'intéresse à ce qui se passe dans la réalité des classes et des établissements. Ces études ont été faites aussi par des sociologues dans les classes. Ce sont d'ailleurs les psychologues qui en ont donné l'idée.

Que donnent les recherches sociologiques?

C'est un peu le même principe. Simplement, puisqu'on ne fait pas d'expérience, on compare pour les élèves. Cela s'est fait, notamment par Marie DURU-BELLAT et Alain MINGAT, sur la classe de 4^e. On compare les résultats des tests standardisés sur 1.500 élèves. Et on compare leurs résultats à leur moyenne annuelle. Et on voit s'il y a concordance ou pas. Voici le principe de ces recherches.

Qu'est-ce qu'on constate?

Un effet du genre, qui avait déjà été constaté par les psychologues. Les filles sont systématiquement mieux notées que les garçons à compétences égales aux épreuves standardisées. Donc, quand on voit les garçons et les filles qui ont "9" aux épreuves standardisées, on s'aperçoit qu'en classe, les filles ont un peu plus par leurs professeurs, et ainsi à toutes les notes standardisées.

Un effet de l'origine sociale, qui est de nouveau retrouvée dans les études sociologiques: ½ point en plus pour les enfants d'origine aisée, donc tous ceux qui ont 10 ou tous ceux qui ont 13: les enfants d'origine aisée ont un peu plus en classe, à compétences égales, aux tests standardisés. Soit, tout de même, le

quart de la différence moyenne entre enfants de cadres et d'origine populaire. Donc, ce biais d'évaluation est très loin d'être négligeable.

Effet du redoublement: un point de moins par année de retard.

Donc, un élève qui a deux années de retard et qui est enfant d'ouvrier, il fait l'objet d'une notation qui est inférieure à son niveau de compétences.

Quelle interprétation?

L'interprétation est quelque chose d'un peu difficile. On peut faire deux hypothèses, d'abord **un effet de stéréotypie**. Les redoublants, par définition, sont jugés plus faibles, et donc on a tendance, en ayant ce sentiment d'élève faible, à les noter en les sous-évaluant, en sous-évaluant leurs compétences. Et pour les enfants de cadres, il peut y avoir une adhésion plus ou moins explicite par les professeurs à l'idée d'hérédité sociale, voire génétique, des compétences: on s'attend à ce qu'ils soient meilleurs, ils réussissent généralement mieux, et c'est une auto-prophétie réalisatrice.

Bon, il n'y a pas forcément qu'un effet de stéréotypie. Par exemple, les enfants de cadres ont des compétences linguistiques plus conformes aux normes scolaires. Les professeurs les connaissant, ils sont influencés par cette compétence linguistique supérieure en classe. Ce qui influe sur leurs notes pendant l'année. Ou, par exemple pour les filles qui sont mieux notées, on peut penser qu'elles répondent mieux aux attentes des professeurs: plus calmes, plus studieuses... et que ce comportement a des effets en classe.

Il y a d'autres biais que l'on connaît moins. Par exemple, dans les bonnes classes, il y a sous-estimation des compétences des élèves dont les notes sont basses dans les sections linguistiques. C'est ce qu'on appelle la "loi de Posthumus". C'est un mot bien compliqué. Ça vient du fait que, si vous prenez 150 élèves, que vous les faites noter par des correcteurs, vous allez obtenir en gros une courbe de GAUSS: les notes vont aller de 5 à 15 points. Maintenant, vous prenez seulement les copies qui ont eu entre 9 et 11, et puis vous les faites corriger par 30 correcteurs. Vous allez constater que ces copies entre 9 et 11, on va de nouveau les retrouver entre 5 et 15. Si vous prenez les notes entre 13 et 15 ou 17 et que vous faites corriger ça par des professeurs à nouveau, vous allez avoir une courbe de GAUSS. Et donc, quand les élèves sont tous dans des bonnes classes, finalement ceux qui sont notés plus faiblement, ce sont des élèves qui, dans des classes faibles, auraient des bonnes notes. Donc, on a des phénomènes de sous-estimation de compétences dans les bonnes classes.

Effet du niveau des élèves dans l'établissement: lycée prestigieux, niveau élevé au test. Ça, c'est assez étonnant! On s'attendait, quand on a fait des tests de compétences, à ce que, lorsque les élèves ont de bons résultats à des tests de compétences, ils aient, au niveau de leur établissement, plutôt de bonnes notes. C'est l'inverse qui se produit. Dans les lycées prestigieux, la notation est plus sévère, parce que le lycée est généralement plus sélectif. Et puis on sait qu'à de bons élèves, on peut, à la limite, mettre des notes assez moyennes, ils sont suffisamment accrochés à l'école et assurés d'un avenir scolaire pour s'accrocher de toute façon. Alors que, dans des établissements où le niveau à des tests standardisés est faible, on s'aperçoit que les notes sont meilleures au niveau de l'ensemble de l'établissement. Pourquoi? Parce qu'on constate que ce sont des élèves au bord du décrochage scolaire et que le professeur sait très bien que s'il met des "3" et des "4" à ses élèves, il les décourage complètement, il ne les reverra pas l'année suivante ou même le mois prochain. Donc, on a une inversion entre les notes mises dans les établissements par rapport aux notes, aux tests de compétences.

Il y a aussi un **effet de la taille de la classe:** notation plus sévère quand le nombre d'élèves est plus élevé. Alors là, il faudrait entrer dans l'interprétation, mais ça prendrait un peu de temps. On peut, par exemple, considérer que lorsque le nombre d'élèves est plus élevé, il y a des risques de chahut ou des gestions de classe qui sont plus difficiles, et la note faible est, pour le professeur, parfois un moyen de tenir sa classe.

Et puis, il y a plein d'autres effets, par exemple *l'ordre de correction des copies*. On sait que le premier tiers des copies est noté de façon plus indulgente. Le professeur est généralement en cours de rodage de son barème. Il cherche sa bonne note, et il est plus indulgent à ce moment-là. Et puis, il y a des expériences qui sont très intéressantes, ça consiste à mettre, sur un paquet de copies, trois copies excellentes, après tout le reste des copies. Elles sont notées de manière beaucoup plus sévère. Si vous prenez le même paquet de copies et vous mettez trois copies très faibles, aussitôt les notes de toutes les autres copies augmentent. On voit bien que le professeur va adapter son barème à partir de ses premières copies. Et donc, il y a des aléas qui sont liés aux pratiques de notation et de mise au point du barème.

Alors, on peut se demander si les biais sont seulement le résultat de stéréotypes.

4. L'approche interactionniste

La notation peut être considérée comme un arrangement. C'est un terme qui peut surprendre parce qu'*a priori*, la notation n'a rien à voir avec un arrangement. Il y a une copie, il y a éventuellement un barème, il y a des compétences exigées. On est donc plutôt dans le mode du jugement que dans le mode de l'arrangement. On sait qu'il y a application de règles: l'orthographe, la présentation, le contenu. Tout ceci n'a rien à avoir avec l'arrangement.

On peut considérer qu'il y a un arrangement dans la mesure où il y a des compromis qui sont réalisés par le professeur en fonction de contraintes interpersonnelles et/ou institutionnelles. C'est ceci que je voudrais développer. Suite à des entretiens avec des professeurs, il m'a semblé évident, finalement, qu'il y avait des arrangements, des négociations plus ou moins implicites qui vont orienter les pratiques de notations. Une sorte, parfois, de cuisine évaluative.

Alors, sur quoi portent ces arrangements?

Nature des arrangements

En écoutant les professeurs parler de leurs évaluations, on est surpris des variations sur le nombre de contrôles.

Sur la pondération des contrôles: ce contrôle-là sera pondéré sur 1, et puis c'est changé et c'est sur 2.

Et puis, il y a de nouvelles modifications du barème en cours de notation, ou le barème qui était annoncé, finalement, a été modifié parce que les élèves n'ont pas assez bien réussi, parce que pour les questions faciles, les barèmes ont été changés.

Ou des bonus sont accordés si le contrôle a été jugé plus difficile, et le contrôle est compté sur 22.

Quelquefois, pour des élèves qui ont eu de très mauvais résultats, le professeur propose une nouvelle correction sur un devoir qui a été refait par l'élève.

Il peut y avoir des suppressions d'un contrôle raté soit collectivement, soit individuellement. L'élève a le droit de supprimer une note sur l'ensemble des contrôles qu'il a eus.

Il y a des contrôles surprise s'il y a mauvaise conduite et chahut des élèves.

Il y a des moyennes qui sont arrondies d'une façon pas forcément très explicite. Il y a des élèves qui viennent protester, et le professeur s'aperçoit éventuellement qu'il a fait une erreur. Mais il ne va pas dire à l'élève qu'il a fait une erreur, parce qu'il craint que d'autres élèves se ramènent, qu'il y ait un effet boule de neige dans la contestation. Alors il dit: "*Je vais voir...*". L'élève ne sait pas trop ce que ça veut dire! Et le professeur sait qu'il s'est trompé, et il arrondit la moyenne d'1/2 point. Ce qui est assez habile, parce que l'élève n'ose pas, après, aller poser la question, il se dit: "*S'il a arrondi par erreur, il va revenir à ma note. Donc, je ne sais pas s'il a corrigé une injustice ou s'il s'est trompé!*".

Ce sont des astuces qui permettent au professeur de maintenir des relations et une motivation avec les élèves. Il s'agit d'une sorte de bricolage. Et vous n'imaginez pas le nombre de bricolages possibles: des bricolages à l'interne, secrets en fait, secrets partagés avec les élèves pour chaque professeur.

Ces arrangements sont intériorisés par les professeurs, ils ne sont pas présentés et perçus comme tels et résultent, à mon avis, de trois types de contraintes qui sont liées à l'action du professeur.

Je vais détailler ces trois types de contraintes:

D'abord, les contraintes externes

Les contraintes externes sont, notamment:

- **le type d'établissement:** élitiste ou en zone d'éducation prioritaire. Il y a des normes non écrites qui entraînent souvent une notation sévère dans les établissements élitistes ou encourageante dans les autres. Ceci est établi de façon statistique tout à fait claire lorsqu'on est à Louis le Grand en lycée en France. Bon, les professeurs savent qu'à la limite, une notation sévère va entraîner une disparition d'élèves qui vont s'orienter dans d'autres établissements, et on s'assure ainsi, en classe terminale, d'avoir à coup sûr 100% de reçus au baccalauréat scientifique. Ce qui est la norme de l'établissement et ce qui maintient son prestige. Dans les zones d'éducation prioritaire, en scolarisant des élèves difficiles, on est dans une tout autre logique, où le but est de maintenir le plus possible l'élève dans sa scolarité;
- **le type de direction.** En discutant avec des chefs d'établissement, on voit bien – et c'est lié avec le type d'établissement en grande partie, d'ailleurs – qu'il y a des projets élitistes ou d'autres projets qui sont l'accès du plus grand nombre à l'enseignement: ce qu'il faut, c'est permettre à l'élève de poursuivre sa scolarité et non trier les meilleurs élèves. Et le chef d'établissement, notamment en conseil de classe, peut favoriser tel ou tel type de notation par un certain nombre de remarques désobligeantes vis-à-vis des professeurs qui notent sévèrement ou, au contraire, de façon indulgente;
- **La norme de notation dans les disciplines.** Ceci est très établi, finalement. En philosophie, la moyenne tourne autour de 8/20. En sports et arts plastiques, les moyennes sont beaucoup plus élevées; dans le cas contraire, les élèves ne feraient pas grand-chose, c'est la "sucette", en quelque sorte, qui leur est accordée;
- **Des normes disciplinaires** qui sont assez marquées. Je prends un extrait d'entretien, le professeur qui dit: *"Pour un prof de sciences naturelles, j'ai une notation à la limite du pathologique"*. Il comparait ses notes aux notes pratiquées en sports. Il se sentait à côté de la norme de sa discipline.

Donc, tout ceci est assez implicite, en fin de compte. Les professeurs en ont une plus ou moins grande conscience. Et puis, ces normes externes, ces contraintes externes vont orienter les arrangements évaluatifs à l'intérieur de la classe.

Les contraintes internes

Les contraintes internes, les professeurs les connaissent:

- **La dynamique du groupe classe.** On est dans des groupes dont les évolutions ne sont pas évidentes. On est soit dans un climat d'assiduité, de travail, soit de chahut. Il y a des effets de leader positifs ou négatifs. Et il y a nécessité, en fonction de la dynamique de groupe, de se dire: *"Il faut que je retienne l'attention. Je vais peut-être avoir une notation plus sévère, ou inversement"*. Et il existe des stratégies d'évaluation qui ont pour but (c'est le but présenté, pas forcément le but atteint) de sanctionner, d'encourager, de contrôler facilement pour remotiver, ou des stratégies annuelles de valorisation de soi qui consistent à noter bas au début de l'année,

et puis le fait de montrer aux élèves la nécessité de travailler et les progrès nécessaires. On le verra tout à l'heure, ce n'est pas une stratégie favorable aux élèves. En fait, elle met en valeur le professeur, mais ça ne suffit pas.

- **Les négociations particulières.** Et puis, il y a des négociations particulières avec certains élèves qui font preuve d'absentéisme, de bouderie, de larmes, de marchandage, etc. Je vais vous citer un petit extrait de ce type de marchandage. C'est un extrait d'entretien. Je demande au professeur s'il arrive souvent des accidents, en quelque sorte, entre la moyenne donnée à l'élève pendant l'année et puis les notes au bac. Voilà ce que me dit le professeur:

"Ça peut... Oui, oui, ça peut dans les deux sens! Dans l'année, par exemple, je peux avoir mis des notes convenables, et puis pfioutttt... brusque dégonflement, plus rien. Par exemple, je prends le cas de Bruno l'année dernière, je l'avais en classe de première. Voilà le cas d'un élève qui ne maîtrise pas du tout l'écrit, n'arrive pas à structurer sa pensée. Au début de l'année, première note: je lui mets 4/20. Il me fait un résumé: 4. Une semaine après, il vient me voir. J'étais dans une classe. Je me souviens très bien: il passe dans le couloir, la porte était ouverte. Il en profite pour entrer et m'interpelle: "Comment ça se fait que vous m'avez mis 4 à mon devoir? L'année passée, j'avais un professeur de lettres qui me mettait 12!". Je lui ai dit: "C'est fort possible, qui c'est?... Ah, c'est elle!... Ha oui, ha bon, non, non, ça ne m'étonne pas du tout..."

Vous voyez, le professeur ne trouve pas d'autre solution que de limiter les compétences évaluatives de son collègue.

"Bon, tu as eu 12 l'année dernière mais cette année, faudra encore le prouver! Tu as eu 4 au résumé. Tu ne sais pas faire un résumé. Si tu me montres que tu sais faire un résumé, je te mettrai 12 et même plus. Je suis tout à fait disposé, mais c'est à toi de me fournir les preuves. Cette année, l'objectif est de te fournir une notation bac.

N'empêche que Bruno, il s'est offert 4 à l'écrit au bac... Bien, et c'est tout à fait mérité!"

Ça, c'est le début de l'entretien, et puis la suite est tout à fait étonnante.

"Mais pendant l'année, j'ai apprécié chez lui une culture que la majorité des élèves de la classe n'avaient pas: des références à des lectures, même si ce ne sont pas forcément les miennes. C'est quelqu'un qui lisait. Il y a eu donc des affinités qui se sont révélées... C'est quelqu'un qui intervenait facilement à l'oral. En plus, c'est l'exemple même de l'élève capable de faire de l'absentéisme plus ou moins chronique. Donc, pour lui donner envie de rester en cours, j'étais obligé d'envisager une notation thérapeutique. C'est le cas même de celui à qui j'ai offert une notation thérapeutique".

Vous voyez: obligé, offert. Et c'est dans le jeu...

"Mais en étant clair avec lui: je te mets 9, je te mets 10 mais je te préviens, le jour du bac, ce n'est pas ce que tu auras!"

Donc, vous voyez l'argumentation qui est: ce n'est pas tout de mettre 4, c'est facile de mettre 4. Enfin, ce n'est pas si facile que ça, puisque l'élève revient. Et puis, il y a une relation éducative à maintenir, et donc, ça me sert à quoi de mettre 4 si l'élève s'en va? Donc, le but, c'était de maintenir la relation à l'élève. Et grâce à cette note indulgente, cet élève s'est maintenu. Il a eu une très mauvaise note au bac, il a repassé l'année suivante son bac de français, il n'avait rien fait pendant l'année, il est allé voir sa prof de français: *"Est-ce que vous pouvez me signer cette liste d'œuvres que j'ai travaillées (enfin, fictivement)?"* Elle a signé la liste, il a eu son bac finalement. Et plus tard, il est devenu professeur d'histoire.

Comme quoi, si cette prof s'était dit: *"C'est 4, c'est 4! Il n'y a aucune raison de lui mettre autre chose que 4!"*, il est très probable que cet élève aurait abandonné le français, et peut-être pas seulement le français, il aurait quitté l'école. Et finalement, c'était un passage dans sa scolarité puisqu'elle le reconnaît elle-même: c'était quelqu'un qui lisait, qui s'intéressait, qui prenait part aux cours, et il fallait maintenir la relation éducative.

Donc, il faut savoir où sont les enjeux. Est-ce que les enjeux, c'est une note juste et qui correspond finalement à une fiction et à pas grand-chose? Ou est-ce que l'enjeu, c'est maintenir la relation éducative pour que l'élève reste dans la classe et pour que le professeur puisse continuer son travail, c'est-à-dire la transmission d'une connaissance?

Notation thérapeutique: "obligation", "offert"... C'est aussi un choix du professeur.

- **Relation personnelle de séduction ou de répulsion.**

Je vais vous donner un autre extrait tout à fait différent:

« C'est vrai que tu as des élèves (c'est en classe terminale) qui sont épouvantables aussi, ça arrive. J'avais une classe qui avait deux mecs. Un jour, à l'intercours, j'étais seule avec eux parce que les autres étaient sortis et il y en a un qui dit en rigolant à son copain: « on se la coince ». Des mecs comme ça aucune pitié. Je crois qu'il n'a pas mesuré la distance à laquelle je tenais. Pour lui, dans sa tête, ça n'était même pas insultant parce que ça doit être comme ça qu'il doit traiter les filles, par ailleurs. Mais je m'en fous, je n'ai pas une pitié infinie pour ces pauvres petits enfants d'ouvriers ».

Donc, on voit bien qu'il y a des répulsions qui sont liés à des comportements ou à des propos qui ne sont pas admis par le professeur.

Un autre exemple tout à fait dans l'autre sens (en quatrième):

« J'ai de bonnes élèves. J'ai trois petites gamines qui ont une tête d'ange, qui sont adorables, qui sont bosseuses, qui répondent, qui connaissent plein de trucs. Et, c'est sûr que quand j'arrive à leurs copies, j'ai un préjugé positif, je le sens. Alors je me dis quelque fois: « Ah la la ! Est-ce que tu ne l'as pas sur-notée? » Alors je prends la copie d'un élève moyen et je compare. Il m'arrive d'enlever des points: 0,5 point parfois »

Alors la conclusion est assez étonnante, parce qu'on n'est pas dans l'ordre du 0,5 point. On est dans l'ordre de 3 à 4 points. Mais, vous voyez qu'il y a ce souci de l'équité, pour le professeur. Il est forcément pris dans ces dimensions de séduction ou de répulsion. C'est inévitable.

Les contraintes par rapport à soi

Le professeur, c'est aussi un ancien élève qui a vécu sa scolarité sereinement, s'il est un bon élève pour lequel il n'y a pas de problème finalement dans l'école, ou qui l'a vécue avec des difficultés. Et ceci se ressent très nettement, notamment pour les professeurs qui ont eu une scolarité très moyenne. Ils ont plus de réticence à mettre des notes basses. Ils se souviennent de la souffrance qu'ils ont eue dans leur scolarité avec ces notes parfois un peu difficiles.

Et il y a des conceptions professorales de l'école marquées soit par l'élitisme, soit par la réussite de tous. Ce qui va entraîner une conception différente de la note. C'est très intéressant, quand on pose des questions aux professeurs sur les notes minimum. Il y a des professeurs qui disent: "Ma note minimum c'est 0, s'il n'y a rien, il n'y a rien, ou 1 ou 2". Puis il y en a qui disent: "C'est 4 ou 5. De toute façon, ça ne sert à rien de mettre au-dessous". Pour d'autres: "Moi, c'est 6 ou 7. Un élève ne peut jamais ne rien savoir". Donc, on voit bien que le débat de la note minimum montre l'aléa, et surtout les stratégies évaluatives. Ou la note maximum. Il y a des conceptions hiérarchiques, ça m'avait beaucoup frappé. Je dis: "Quelle est ta meilleure note?". Il me dit: "C'est 17/20. 17/20, c'est le très bon élève; 18, c'est moi; 19, c'est l'écrivain (c'est un professeur de français) et 20, c'est Dieu". Donc, on avait à chaque correcteur, à chaque note, cette hiérarchie inévitable qui est: "Je suis au-dessus de l'élève, l'écrivain est au-dessus de moi, et il y a la perfection. Et personne ne peut l'atteindre: c'est 20". Alors que, dans d'autres formes de pratiques pédagogiques avec des objectifs de compétences précises, tous les élèves peuvent avoir 20. Il y a l'objectif: il est atteint, on a 20. Donc, on est dans des mondes totalement différents.

Alors, il y a aussi un **effet du genre du correcteur**. Les notes au bac sont, pour les hommes, plus reliées aux notes pendant l'année. Donc, on peut penser qu'il y a plus de négociations peut-être par les professeurs femmes, comme l'exemple que je vous ai donné tout à l'heure en français.

Il y a un **effet de l'âge** aussi. Curieusement, il y a une diminution de la dispersion des notes. Sans doute, c'est l'effet de l'expérience. C'est probablement se rendre compte que mettre des notes de 3 à 17, ça ne sert à rien. On obtient le même résultat avec des notes de 5 à 15, même de 8 à 12. L'idée d'écarter les notes ne présente aucun intérêt pour les élèves.

5. Principes et pratiques d'une notation au service de l'équité et des progrès scolaires

Alors, j'en viens à quelques conclusions, en quelque sorte, sur les principes et pratiques d'une notation au service de l'équité et des progrès scolaires.

1. Préserver l'anonymat social de l'élève

C'est la conclusion de tous les biais d'évaluation. Je ne sais pas si ça se passe comme ça en Belgique. Ça se passe comme ça en France: au début de l'année, au premier cours, les professeurs font remplir à leurs élèves ce qu'ils appellent une fiche de renseignements dans laquelle ils demandent la profession des parents, le parcours scolaire antérieur, le niveau dans la discipline... Plein d'informations, qui sont potentiellement une source de biais d'évaluation. Parce que toutes ces informations: la profession des parents, est-ce que vous étiez bon avant en mathématiques, en français, en anglais, etc. créent des effets d'attente. Et puis, ces informations ne sont nullement nécessaires à la relation pédagogique et aux apprentissages. Il n'y a aucune raison, pour un professeur, aucune nécessité pédagogique d'avoir ces informations. Et celles-ci sont, de surcroît, mal perçues par les élèves parce qu'ils ont l'impression tout de suite d'être fliqués, en quelque sorte, catalogués. Ce qui n'est pas faux, dans un sens docimologique et sociologique du terme, d'après les études.

2. Préserver l'anonymat scolaire des élèves

Je vous donne un exemple. C'est une étude tout à fait passionnante, qui a été faite en France en 1999. Une thèse que David OGET a soutenue. Il a comparé les épreuves, les résultats au baccalauréat, ce sont des épreuves anonymes en France, avec les résultats obtenus au contrôle continu.

Hé bien, qu'est-ce qu'il a observé? Qu'avec un baccalauréat, c'est-à-dire des épreuves anonymes, il y a plus de garçons admis que si c'est le contrôle continu, plus de redoublants admis, plus d'enfants d'origine populaire admis. Donc, on voit tout l'intérêt d'une épreuve anonyme, et pour ce qui est de la Belgique, l'intérêt d'un certificat d'études secondaires supérieures qui serait organisé de façon anonyme comme le baccalauréat.

Il y a une justice à attendre d'un système où l'anonymat scolaire des élèves est préservé pour les diplômés, et aussi pour les examens.

3. Mutualiser les épreuves d'évaluation

Parce qu'au cours de l'année, c'est aussi un avantage pour l'élève – et pas seulement pour l'élève, pour le professeur aussi – de mettre en place des épreuves communes à plusieurs classes et au moins deux professeurs.

Il y a un certain nombre de professeurs en France qui font ça. Et qui travaillent par équipes, de deux par exemple, qui se mettent d'accord sur une progression dans l'année, on voit par exemple les mêmes textes en français ou en mathématiques. On a exactement le même programme en parallèle, on fait le contrôle en parallèle. Les professeurs font à tour de rôle un contrôle, et puis le collègue regarde si le contrôle convient. Et puis, ils corrigent indifféremment soit les copies de leurs élèves, soit les copies de l'autre classe. Pour les élèves, c'est beaucoup mieux, parce qu'ils ont le sentiment qu'ils ne sont pas catalogués. Et pour les professeurs, c'est aussi beaucoup mieux. Et ils sont tous à considérer qu'il y a un grand enrichissement à comprendre la façon dont l'autre collègue conçoit son cours, le travaille. Et ils

s'apportent respectivement leurs connaissances et leur compréhension du programme. Ils s'aperçoivent aussi qu'ils ont parfois des compréhensions très différentes du programme.

Donc, l'idée de mutualiser les épreuves d'évaluation, c'est tout bénéfique pour les élèves et aussi pour les professeurs. Ça nécessite évidemment un travail en commun sur le programme, ça permet de dégager des exigences disciplinaires minimum.

Bon, c'est aussi une façon de mettre au point collectivement un barème, c'est un outil nécessaire pour les correcteurs et pour les élèves. Quelquefois, les professeurs disent: *"J'ai un barème, mais ce barème n'est pas précisé sur la copie, sur le contrôle"*. Évidemment, c'est très néfaste pour l'élève. L'élève a besoin de savoir, quand il fait son devoir, si la question 1 est notée sur 2 ou sur 4. Il ne va pas y passer le même temps. Donc, donner cette information, qui paraît élémentaire, n'est pas toujours fait dans la classe.

Et puis, la mutualisation permet aussi des comparaisons interclasses qui sont plus solides et plus équitables en matière de redoublants, pour autant qu'il y ait des redoublements. On sait que l'effet est généralement négatif, à part certaines classes très particulières.

4. Préférer une notation encourageante à la croyance de la note vraie

Il faut bien comprendre que l'on distingue souvent l'évaluation sommative qui est de mesurer le niveau, et l'évaluation formative qui est de former, d'aider. En fait, il faut bien comprendre que le niveau des notes exerce en lui-même des effets sur l'image scolaire de l'élève et sur ses progrès, finalement. Donc, on ne peut pas se dire: *"Ça, c'est de l'évaluation sommative"*. L'évaluation sommative, elle a un effet formatif sur l'image scolaire, sur la valorisation de soi.

La recherche a été menée par rapport à des professeurs qui commencent plutôt en début d'année par des notes très basses; on s'aperçoit, sur des élèves qui, au début, ont des niveaux de compétences identiques dans les différentes classes, que lorsqu'on commence par des notes très basses, il y a un effet de découragement, de dévalorisation de soi. Il y a un effet d'auto-prophétie réalisatrice qui fait que l'élève qui commence avec des notes très basses se sent moins bon et travaille moins, et est moins mobilisé.

Je me permets de vous raconter une petite anecdote, en classe terminale (puisque j'ai enseigné quelques années en lycée).

Première note de l'année (j'étais professeur de sciences économiques et sociales): pour un devoir que je ne trouve pas bon du tout, je mets 5/20. J'ajoute: *"Aucun travail"*, et je rends cette copie à l'élève. C'était une fille, je m'en souviens très bien. Elle me regarde, elle se met à avoir les yeux qui s'humectent. Je me dis: *"Elle prend ça à cœur, au moins"*. Elle me dit: *"5/20, d'accord... Mais «Aucun travail»..."*, et elle se met en colère: *"Aucun travail, qu'est-ce que vous en savez? J'ai travaillé tout le weekend sur ce devoir!"*. Qu'est-ce que je pouvais répondre à ça? Je n'étais pas avec elle le weekend, donc si elle me dit ça comme ça, en découvrant sa note, c'est que c'est vrai, elle y a travaillé tout le weekend. Bon! Et je lui ai mis 5/20, passe encore, elle l'accepte. Mais pas *"Aucun travail"*. Donc, il y avait manifestement un problème.

Je vous donne un exemple où on est inspiré, comme professeur. Il y a aussi des ratés. Donc là, il m'a semblé essentiel de rayer *"Aucun travail"*, qui était manifestement une gaffe, une erreur, et puis de lui accorder une réparation symbolique minimum qui était d'enlever l'appréciation fautive: enlever *"Aucun travail"*, rayer son appréciation. Et je lui ai dit: *"Hé bien, si tu as travaillé tout le weekend, en continuant, il n'y a pas de raison que tu n'aies pas de bonnes notes!"*.

Bon, là ça s'est terminé comme ça. La fois suivante, elle a eu 14/20. Vous me direz: *"Oui, mais bon... Vous aviez peur de vous faire engueuler une nouvelle fois!"*. Ce qui n'était peut-être pas faux, d'ailleurs. Peut-être qu'après tout, j'ai gonflé ses notes parce que je craignais ses réactions coléreuses?

Mais au baccalauréat anonyme, elle a eu 14/20. Bon, je lui ai peut-être gonflé ses notes après. C'est possible! Mais ce qui compte, c'est le résultat final, corrigé de façon anonyme par un autre prof.

Moi je pense que l'année de cette élève en économie (gros coefficient) s'est jouée sur ces quelques secondes où je me suis dit qu'il y avait un problème: réparation symbolique, et je lui ai dit: "*Il n'y pas de raison que ça ne marche pas*". Ça, c'est de la négociation!

Quelquefois, on trouve les mots qu'il faut, les gestes qu'il faut. Quelquefois, on ne les trouve pas. Et je suis sûr que si j'avais dit: "*Quand on a 5/20, c'est qu'on n'a pas travaillé!*", elle serait sortie en disant: "*Merde, c'est un con... L'économie, c'est fini!*". Et j'aurais raté mon année scolaire en quelques secondes avec cette élève. C'est ça la négociation! Et c'est ça le principe, en quelque sorte: préférer la note encourageante à la croyance en la note 20. Qu'est-ce que c'était, mon 5/20? Ça n'a pas beaucoup d'intérêt! Ce qui compte, c'est qu'elle continue à travailler. Et là, ça a marché!

Donc, ces moyennes basses pour commencer sont à éviter, parce qu'elles vont favoriser le décrochage. Et c'est ce qu'il faut éviter à tout prix.

5. Contractualiser les pratiques d'évaluation

Éviter un seul contrôle couperet par exemple, ce qui se fait de temps en temps.

Pas d'épreuve croche pied. Il y a des professeurs qui sont spécialisés dans l'épreuve croche pied. Ça ressemble à un contrôle, mais il y a un petit quelque chose qui, en final, change tout. Et alors, l'élève qui a compris son cours, qui peut le refaire, finalement, il ne va rien comprendre. Il va se dire: "*Ça ne ressemblait à rien, je n'ai pas compris!*". Et il se paie une gamelle alors que finalement, si ça avait été à peu près la même chose qu'en classe, il aurait eu une note convenable. Alors, ça ne veut pas dire qu'il ne faut pas d'épreuve difficile ou de question difficile, mais le but, ce n'est pas de faire tomber les élèves! C'est, par rapport à ce qui a été vu en classe, de voir ce que l'élève a pu comprendre.

En ce sens, et un certain nombre de professeurs le disent, faire un exercice et puis dire: "*Dans le contrôle, vous aurez au moins cet exercice-là, qui est sur ce principe-là. Les chiffres seront différents, mais ce sera le même principe*". Il n'y a aucune raison pour encourager les élèves à travailler sur les objectifs qui seront après revus dans le contrôle. Et une partie des exercices doivent être accessibles aux élèves faibles. Il n'y a rien de plus injuste, d'un point de vue scolaire, que de donner des exercices faciles en cours, et puis qui disparaissent au moment des épreuves! Un élève qui a compris quelque chose et qui a travaillé doit avoir un minimum de reconnaissance scolaire par une note minimum qui ne soit pas infamante. Pourquoi travaillerait-il, autrement?

Et puis, varier la nature des épreuves: longues, courtes, sur des points particuliers des cours, orales... et prévenir les élèves des contrôles, il n'y a aucune raison de ne pas le faire!

Voilà, c'est ce que je voulais vous dire. Je vous remercie de votre attention.

Éléments de biographie

- MERLE P., *Les notes. Secrets de fabrication*, Paris, PUF, 2007
- NOIZET G., CAVERNI J.-P., *Psychologie de l'évaluation scolaire*, Paris, PUF, 1978
- SUCHAUT B., *La loterie des notes au bac. Un réexamen de l'arbitraire des notes au bac*, IREDU, 2008

Questions à Pierre MERLE

■ Monsieur MERLE, pourriez vous nous dire quelle appréciation vous avez sur les conseils de classe par rapport aux évaluations?

Pierre MERLE: Je suis assez embarrassé, je dois dire. J'ai considéré que cela faisait partie de mon travail de sociologue de l'école d'être parent représentant des parents d'élèves dans les conseils de classe. Donc, j'ai assisté à des conseils de classe pendant une quinzaine d'années (3 par an, ce qui en fait 45). Je dois vous dire que je trouve que c'est assez souvent attristant et décourageant pour les élèves qui sont là. Même les professeurs...

Il y a souvent une sorte de mise à mort des élèves. J'ai l'impression qu'il n'y a pas véritablement un travail de réflexion mené sur la valeur des notes, sur leur relativité.

Il y a quelquefois des chefs d'établissement ou des conseillers principaux d'éducation qui dirigent le conseil de classe, qui remettent les choses au point, en quelque sorte. Mais par exemple, vous voyez une section linguistique où, au 1^{er} trimestre, les professeurs mettent tous de 15 à 17 de moyenne et disent que c'est une classe excellente, et le professeur de maths qui se réjouit et dit: *"Ah! En maths, j'ai regardé leur dossier, la moyenne de ces élèves était de 16!"*. Et puis vous voyez qu'à la fin de l'année, ils terminent avec 12 en maths, que la moyenne a baissé tout le temps, pareil dans toutes les disciplines. Les professeurs ont oublié petit à petit qu'il s'agissait d'une classe excellente. Et à la fin, ils proposent deux à trois redoublements, la même chose que dans une classe d'élèves beaucoup plus faibles. Bon, il faut que le chef d'établissement intervienne et dise: *"Attendez! Là, vous êtes dans une très bonne classe, vous ne savez pas les problèmes qui se posent à côté!"*. Donc, les professeurs ne relativisent pas par rapport aux autres classes. Alors, il y avait quand même dans cet établissement une évaluation sur toutes les classes, en maths par exemple, en parallèle, qui a bien montré que cette classe était bien meilleure que les autres, mais finalement il n'en est pas tenu compte. Donc, il manque vraiment une réflexion sur la qualité ou la fiabilité des notes mises par les professeurs, qui sont très relatives au niveau de la classe. Et cette réflexion, je pense qu'elle ne peut pas avoir lieu à l'absence d'une formation initiale ou continuée sur les recherches en sociologie et en psychologie de l'évaluation.

Pour l'instant, je pense que le résultat est très peu satisfaisant. D'autant qu'on assiste, en France, et pas seulement en France je suppose, à des redoublements qui sont des redoublements sanctions, parfois étonnants, du style: *"Boh! Il est moyen, il tourne autour de 9,5 mais vraiment, c'est un élève trop casse-pieds. Il faut le faire redoubler, il va murir!"*. Mais ça n'a pas de sens. C'est un élève qui s'ennuie déjà probablement en classe. Ce n'est pas en lui faisant refaire tout à fait la même chose, alors qu'il est tout près de la moyenne, qu'il va apprendre quelque chose!

Et il y a encore une telle autorité des professeurs par rapport aux parents d'origine moyenne et populaire, que les parents n'osent pas contester, ou pas assez, ce diagnostic de redoublement. Pour un redoublement qui ne se justifie pas, parce que plein d'études ont montré qu'un élève faible qui passe finit par mieux réussir qu'un élève qui redouble. On décourage les élèves et on casse un peu leur histoire scolaire.

Donc, moi j'ai un bilan de ce que j'en ai vu: un bilan très négatif, où les parents n'ont pas voix au chapitre. On leur demande de s'installer dans un coin. Moi qui m'installais au milieu des professeurs, j'étais regardé complètement de travers. On se demandait pourquoi je me mettais là, alors qu'il y avait une place au bout de la table avec les élèves. Donc, c'est un monde dans lequel je trouve qu'il y a beaucoup à faire. Voilà. J'espère que mon diagnostic n'est pas trop négatif!

■ **Merci, Monsieur MERLE, pour cet exposé aussi humble que démystifiant. Quelles seraient vos idées en matière d'évaluation de l'école?**

Pierre MERLE: C'est un autre sujet... Tout aussi difficile! Je pense qu'on ne peut pas isoler la question de l'évaluation de l'école, de son recrutement social. La question, ce n'est pas seulement ce qu'on peut faire, en France, par la DEP (Direction de l'évaluation et de la prospective), c'est-à-dire évaluer le niveau de compétences des élèves à un certain niveau et puis comparer les écoles entre elles. Ou ce qu'on fait, par exemple, au niveau du lycée: on calcule des taux bruts de réussite, par exemple au baccalauréat, et puis en fonction du recrutement social, on calcule des taux nets pour savoir si cette école est plus efficace qu'une autre. Bon, ça c'est une approche descriptive, utile mais qui ne permet pas d'envisager des solutions par rapport à la question de l'équité et par rapport à la question de l'efficacité.

On a, en France, un certain nombre d'études (pas seulement en France d'ailleurs, avec les études PISA qui ont une perspective internationale) qui montrent bien qu'il y a un lien entre la mixité sociale de l'école et l'efficacité. Et je pense que le grand défi aujourd'hui, particulièrement en France et en Belgique, mais aussi dans d'autres pays européens, c'est celui d'arriver à instaurer la mixité sociale de l'école. C'est ce que disait François DUBET ce matin. Ce sont des études qui sont maintenant bien étayées. Lorsqu'on a des publics homogènes, on va dire les bons d'origine sociale plutôt aisée d'un côté, et puis les plutôt faibles d'origine populaire de l'autre côté – pour simplifier à l'extrême, mais en gros, c'est ça! –, on observe que les bons progressent plus, alors que les faibles, eux, s'effondrent. Si on les met tous ensemble, ce que perdent les bons, c'est très peu, et ce que gagnent les très faibles, c'est énorme! Donc, on sait aujourd'hui que l'équité et l'efficacité vont ensemble.

Et maintenant, le problème c'est: comment gérer la situation actuelle? On a quand même une ségrégation urbaine qui est forte et qui favorise une ségrégation scolaire qui est forte. Je fais actuellement des études sur la ségrégation scolaire à Paris. C'est une ségrégation qui est considérable et qui est un frein énorme à une évolution de notre système vers plus d'équité et plus d'efficacité. Moi je pense que c'est le grand défi aujourd'hui. En France, on avait une carte scolaire dans le public. On est en train de l'assouplir. Bon! En fait, ça pose un certain nombre de questions. Vous en connaissez d'autres pour aller vers plus de régulation. Nous, on va vers moins de régulation et une régulation différente, notre régulation très étatique n'a pas fonctionné, parce qu'il y avait plein de détournements de ces régulations.

Et je pense qu'on peut poser la question de l'évaluation des établissements, mais il faut aller plus loin sur les contextes de scolarisation. Comment modifier ceux-ci? Je crois que c'est là le vrai défi!

Après, si on veut raisonner, sans parler du contexte, sur les évaluations des établissements, on peut se dire qu'il faut pondérer ces évaluations par rapport aux difficultés propres à l'établissement. C'est-à-dire: quand il y a des établissements avec beaucoup d'enfants d'origine étrangère ne maîtrisant pas suffisamment les compétences linguistiques nécessaires, etc., il faut des compensations financières pour ce type d'établissements.

Alors, dans quelle mesure sera-ce accepté? Dans quelle mesure aussi ces compensations financières peuvent-elles compenser la mixité sociale? Je n'y crois qu'à moitié, à vrai dire. Je pense que finalement, le meilleur système du point de vue un peu idéal, c'est un système dans lequel on aurait de la mixité sociale. Tout ceci est un peu compliqué, parce que quand on regarde un établissement et son recrutement social, on dit que c'est un recrutement social mixte; en fait, quand on regarde les classes dans le détail, on s'aperçoit que la mixité peut être importante au niveau de l'établissement, mais faible au niveau des classes. Puisque, par le jeu des options, on réunit parfois certains élèves. Donc, il nous manque des éléments statistiques pour aller vraiment au cœur de l'unité pédagogique qui est la classe.

Mais vous savez, c'est intéressant, la mixité au niveau de l'école. Parce que les enfants se rencontrent à la cantine, dans la cour de récréation et, on le sait, un des avantages de la mixité sociale, c'est que les enfants ont des copains d'autres origines sociales. Ils se familiarisent avec d'autres univers sociaux,

d'autres univers linguistiques. Ils rencontrent d'autres projets sociaux. Quand on est dans une cité ouvrière, les enfants se disent: *"Je serai ouvrier, ou je serai au chômage"*. Quand on est dans un milieu socialement mixte, ils rencontrent des enfants qui veulent faire des études, qui veulent être médecins et ils se disent: *"Mais moi, j'ai le même âge! Finalement, à l'école, je ne suis pas plus mauvais. Pourquoi pas moi?"*. Donc, cette mixité sociale, elle est importante en termes d'intégration sociale.

Moi je pense qu'à l'école, on doit réussir à trouver une solution au "vivre ensemble" parce que si on ne le fait pas, après c'est dans les banlieues qu'on a des problèmes, en France notamment, de violence forte, des problèmes de banlieues de jeunes, de révolte vis-à-vis du système scolaire, des voitures brûlées... Il ne faut pas croire qu'on va mettre un couvercle sur le problème du "vivre ensemble" et de la mixité sociale!

Donc, on voit bien que de la question de l'évaluation des établissements découlent plein d'autres questions, et on voit aussi comment on arrive à organiser une vie sociale pour faire en sorte qu'il y ait une société, et pas des sociétés qui s'affrontent à l'intérieur d'une nation. C'est ça le grand défi, à mon sens. En tout cas, pour les sociologues, c'est le grand défi d'aujourd'hui! Et ça devrait être aussi le grand défi pour les politiques. ■