

DIVERSITE D'UTILISATION DE L'ECHELLE DE NOTES PAR LES ENSEIGNANTS DU SECOND DEGRE

Yvan Abernot

Aix Marseille université
UMR P3 ADEF
3, Place Victor Hugo
13 331 Marseille Cedex 03
yvan.abernot@univ-provence.fr

Mots-clés : *docimologie, pédagogie, psychologie*

Résumé. *Les travaux de docimologie ont établi des connaissances utiles à l'amélioration des outils d'évaluation scolaire. A la suite de ces apports différenciant les correcteurs, nous prenons en compte la variable niveau dans le cursus (collège, lycée) et la variable sexe. Par ailleurs, nous voudrions approfondir la psychologie du correcteur, en mettant au jour des particularités du « profil docimologique », éventuellement corrélées à des options pédagogiques ou à des rapports différents à la mission éducative. A cet effet, nous avons recueilli 30 520 notes distribuées par près d'une centaine d'enseignants du secondaire. Un cumul général montre que la courbe n'est ni lisse ni strictement gaussienne ! La courbe de distribution de chaque enseignant visualise des phénomènes communs ou singuliers comme l'utilisation de certaines notes ou de certaines parties de l'échelle, par excès ou par défaut.*

1. Une problématique ancienne : la difficulté d'évaluer

Il est connu, sans doute depuis que la notation existe, mais plus scientifiquement, en France, depuis les travaux de docimologie initiés par Pieron (1963) dans les années vingt, que des problèmes de justesse, voire de justice, se posent quant à la correspondance entre les notes scolaire et la « valeur » des copies. De nombreux auteurs (Abernot, 1998, pour une brève revue de travaux), ont établi que la note attribuée à une copie est sous l'influence de nombreuses variables : les autres copies, le niveau de la classe, l'établissement, le pays, etc., et surtout le correcteur !

1.1 Une question vive : apprendre à évaluer

Les travaux sur la plus ou moins grande « sévérité » des correcteurs ne sont pas à reprendre, celle de la plus ou moins grande concentration des notes sur la zone centrale de l'échelle de notes, non plus. Les outils d'évaluation les plus normés (dictées, exercices, etc.), c'est-à-dire, ceux qui disposent d'une unité et font action de mesure, ont des courbes de distribution plus étendues que la dissertation, par exemple, qui procède davantage de l'appréciation. Par ailleurs, les moyennes sont toujours plus concentrées que les notes attribuées aux prestations elles-mêmes car pour avoir zéro de moyenne, il faudrait n'avoir eu que des zéros (fait quand même rare) et ainsi du vingt.

Au XXI^e siècle encore, certains chercheurs (Antibi, 2003) rappellent la force de la normalité statistique, montrant, après Karl Posthumus (1947), et bien d'autres, que les notes se retrouvent toujours nombreuses au milieu de l'échelle et rares aux extrémités. Les élèves faibles des classes très performantes sont donc éventuellement meilleurs qu'il n'y paraît ! De là à oublier que l'idéal de tout outil d'évaluation est de positionner tous les élèves d'un groupe sur l'échelle disponible, il n'y a qu'un pas ... à ne pas franchir ! Que dirait-on d'une évaluation standardisée au niveau

national à laquelle toute une classe aurait zéro (ou 20) toute l'année ? Evidemment, les parents d'un élève faible aime entendre que celui-ci est peut-être fort malgré ses notes... mais rares sont les parents qui ignorent que leur enfant est dans une classe d'excellence ! En revanche, il faudrait éventuellement s'intéresser aux élèves forts des classes faibles dont la déconvenue risque d'être marquante.

Raillerie à part, il serait très judicieux d'effectuer un test national à spectre large par année mais les organisateurs du baccalauréat sont suffisamment obnubilés par ce sacrosaint examen pour ne pas leur en imposer un à chaque fin d'année du cursus. Sans compter avec le bachotage associé, car même les plus intéressés d'entre nous par l'évaluation ne veulent pas confondre éducation et préparation aux examens, tout en accordant un effet motivant à l'évaluation. Encore faut-il s'interroger sur la nature des apprentissages motivés par l'évaluation et sur leur perdurée (Lieury & Fenouillet, 1996). Tout enseignant devrait être initié à ces phénomènes et s'interroger régulièrement sur sa propre manière de noter.

Ce n'est donc pas pour remettre en question le système de notation de 0 à 20 que nous avons recueilli et traité un grand nombre de notes attribuées par une centaine d'enseignants du second degré, mais pour en découvrir d'éventuels usages particuliers.

Hypothèse : tous les évaluateurs ne font pas le même usage de l'échelle de notes.

1.2 Des objets de recherche peu travaillés

De nombreuses études portent sur les aspects techniques de l'évaluation, comme le choix du type d'échelle (numérique, ordinale) ou la *sensibilité* de l'évaluation (de 0 à 10, de 0 à 20). Les problèmes de *validité* et de *fidélité* sont largement mis au jour depuis les travaux de G. De Landsheere (1972). Les difficultés inhérentes aux types d'outils (des plus fermés, comme le Q.C.M., aux plus ouverts, comme la dissertation) ont également donné lieu à de très riches apports depuis G. Noizet et J.P. Caverni (1978). Quant aux aspects éthiques laissant au hasard et à la subjectivité du correcteur une part de pouvoir sur le sort des élèves, ils sont régulièrement soulevés par toutes les parties prenantes de l'école. En revanche, les particularités des utilisateurs sont moins étudiées. Il est souvent question de « fidélité » au sens de la docimologie classique, c'est-à-dire, de l'influence de la stabilité de la note face aux répétitions de l'épreuve et face à la multi-correction. Les désaccords entre correcteurs font encore régulièrement les titres des journaux. On y conteste l'objectivité mais la subjectivité du correcteur est généralement traitée comme une sensibilité particulière à des éléments de fond, de forme ou de personne (Didierjean & Abernot, 2005). Ainsi, certains enseignants ne prennent pas en compte officiellement l'orthographe mais retirent des points lorsqu'elle est trop mal respectée, ce qu'Abernot (1998) appelle une « variable de débordement ». Tel enseignant de mathématiques sortira de ses gonds (et de son barème) en voyant apparaître une flèche dans une copie, ce qui déclenchera l'attribution immédiate d'un zéro ! Ces « variables choc » peuvent également être positives. De même, l'influence sur les correcteurs de certaines caractéristiques de l'élève (milieu social, sexe, etc.) est bien établi (Gilly, 1980). Mais l'usage personnel de l'outil d'évaluation est peu étudié car il se situe à l'interface de deux domaines : la docimologie et la psychologie clinique. Ce point aveugle pourrait aussi être expliqué par le fait que sa mise en lumière risque de ranimer la contestation des évalués et des tenants de l'abandon de l'évaluation chiffrée.

1.3 Des variables contextuelles et des attitudes personnelles

Comme il y a des genres et des styles d'enseignants (Clot & Faïta, 2000), il y a des genres et des styles d'évaluateurs. Nous n'avons à ce jour que des données cumulées sur les critères de niveau, de sexe et de matière mais déjà, certains phénomènes y sont nettement corrélés. Au-delà de ces aspects catégoriels, un regard sociologique pourrait permettre d'explorer d'autres macro-facteurs éventuellement actifs comme l'établissement, la région ou le pays. Ainsi, rien n'indique que les enseignants français évaluent comme leurs homologues européens ou ressortissants d'autres continents. La culture vient peut-être ici marquer les pratiques. Un regard plus psychologique,

pourrait permettre d'établir d'éventuelles caractéristiques individuelles des enseignants correcteurs. L'analyse de ces dimensions, tant sociologiques que psychologiques, relève de la recherche fondamentale mais déjà dans notre analyse, située à l'interface, certains facteurs laissent entrevoir des répercussions sur la formation des enseignants et un gain de compréhension du sens de leurs actions.

2. Investigation

2.1 Population

Nous avons recueilli les notes de 86 enseignants dont 34 de divers collèges et 52 de divers lycées. La figure 1 montre un exemple de profil de correcteur qui distribue beaucoup de notes élevées, notamment 18 mais presque jamais 11 ! Il ne donne presque aucune note en dessous de 3 sauf 0 !

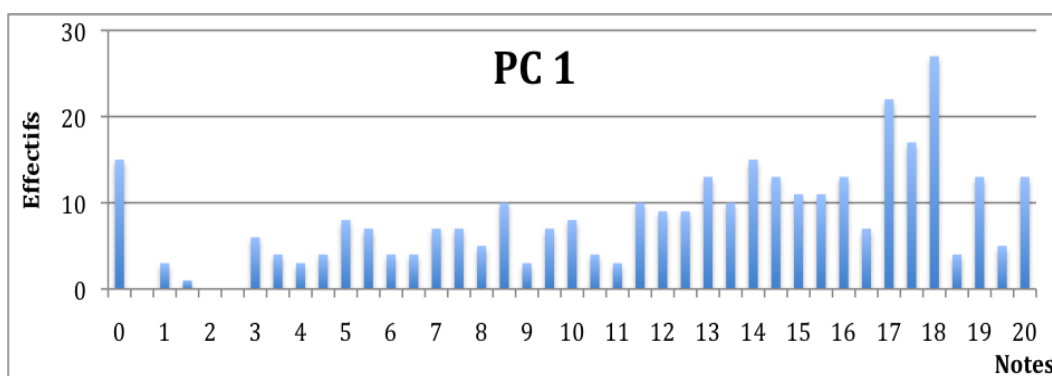


Figure 1 : Distribution statistique des notes du Professeur de Collège PC1.

Les courbes de distribution des 86 enseignants vont donc être analysées individuellement de manière à y repérer les aspects communs et les particularités. Les données seront également cumulées pour pouvoir être comparées sur le critère de niveaux dans le cursus (collège - lycée) et celui de sexe et de matière. Les données sont anonymes tant du point de vue des enseignants que de celui des élèves.

2.2 Données

Les notes considérées sont toutes attribuées directement à une prestation (écrite, orale ou autre comme en EPS, Arts Plastiques ou Musique). Les moyennes ne sont pas prises en compte car elles masquent les particularités. Ainsi sélectionnées, nous disposons de 30 520 notes distribuées par 87 enseignants dont 35 professeurs de collège et 52 de lycée.

3. Résultats

3.1 Données cumulées

Le cumul de toutes les données (30 520 notes) permet de visualiser la distribution générale. Il y a, certes, plus de notes au milieu qu'aux extrémités de l'échelle mais un Khi deux indique que la distribution n'est pas gaussienne !

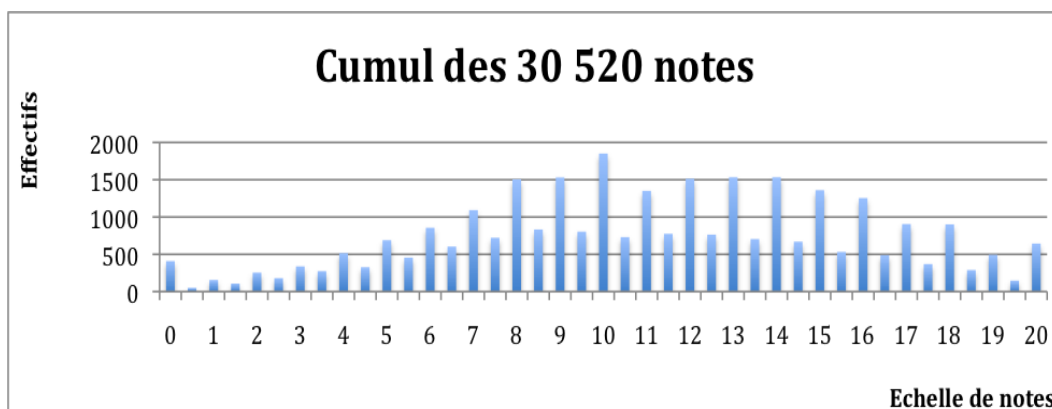


Figure 2 : 30 520 notes attribuées par 86 enseignants du second degré.
N = 30 520 Moyenne = 11,09 Ecart-type = 5,32

Epistémologiquement, il existe plusieurs manières de repérer une note distribuée de manière remarquable : comparaison de sa fréquence par rapport à la normalité gaussienne (à l'aide d'un seuil) ou à la courbe cumulée des enseignants, rapport aux notes adjacentes (par exemple, plus du double ou moins de la moitié de la moyenne des adjacentes). Plus simplement, il peut suffire de repérer les fréquences inférieures à la plus basse (ou supérieure à la plus haute) des fréquences adjacentes.

Sur cette courbe de la figure 2, il est notoire que les notes extrêmes (0 et 20) ne participent pas à l'extinction des effectifs. De même, la modalité médiane de l'échelle (10) est, elle aussi, très utilisée par rapport au lissage de la courbe. Nous verrons ci-dessous que de très nombreux enseignants en font leur mode (note la plus distribuée). En revanche, l'effectif de la note 11 est inférieur à celui de la moyenne entre les notes 10 et 12 (et même, inférieur à celui de la note 12) !

3.1.1 Les demi-points

Une nette distinction entre la courbe des notes entières et celle des demi-points apparaît immédiatement sur la figure 2. Les seconds sont beaucoup moins distribués que les premiers (9 830 contre 20 690). L'hypothèse d'un effet de cumul serait plausible si les enseignants étaient rares à distribuer des demis, or ce n'est pas le cas. Presque tous utilisent les demis et tous les distribuent deux à trois fois moins que les entiers. Ce n'est donc pas du cumul que vient cette différence de fréquences mais d'un usage particulier de l'échelle. Ceci nous rappelle un constat déjà ancien (Abernot, 1987) sur les comportements de vote sur les échelles ordinales. Lorsque 5 modalités sont proposées, les cases 2 et 4 sont sous-utilisées. Il semble donc que nous soyons face à un phénomène qui n'a pas seulement à voir avec une *sensibilité* insuffisante de l'échelle mais aussi avec une tendance à « préférer » les entiers ou, pour le dire autrement, à n'utiliser les demis qu'en cas d'indécision entre 2 entiers. Si les demi-points avaient le même statut que les entiers ils s'inscriraient dans le lissage de la courbe.

3.1.2 L'écart-type

Le graphique et le calcul permettent de constater que l'écart-type est élevé (5,32 : plus d'un quart de l'échelle). Contrairement à la déduction précédente, il s'agit d'un effet de cumul car très rares sont les enseignants qui, pris individuellement, offrent une telle dispersion.

3.2 Comparaison collègue – Lycée

Les courbes montrent de notoires différences attestées par les statistiques ($P < .001$).

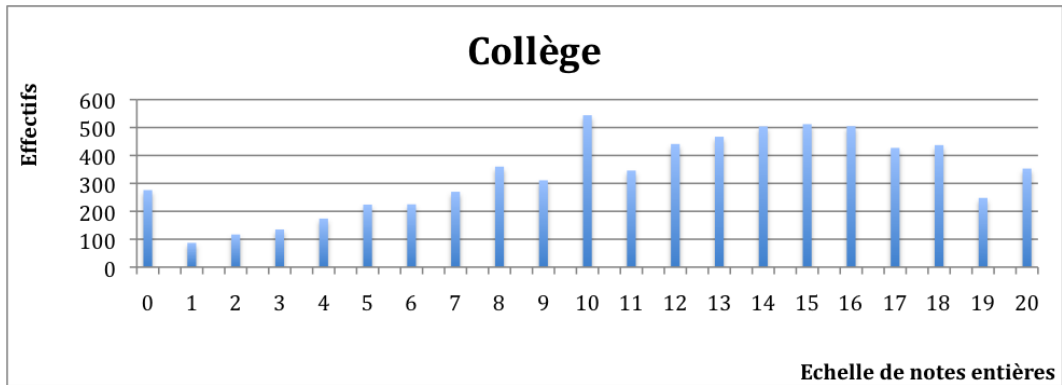


Figure 3 : Collège : moyenne 11,68 Ecart-type. : (5,08)

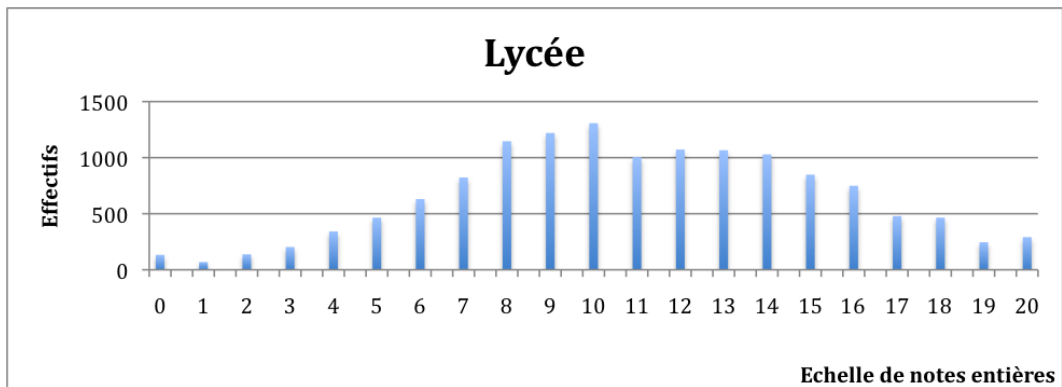


Figure 4 : Lycée : 10,82 (4,26)

On remarque qu'au lycée les notes présentent une moyenne inférieure et sont plus serrées autour de leur moyenne.

3.3 Comparaison hommes – femmes

Les courbes rassemblant les notes des enseignantes de collège montre un nette tendance à distribuer des notes du haut de l'échelle (jusqu'à 20).

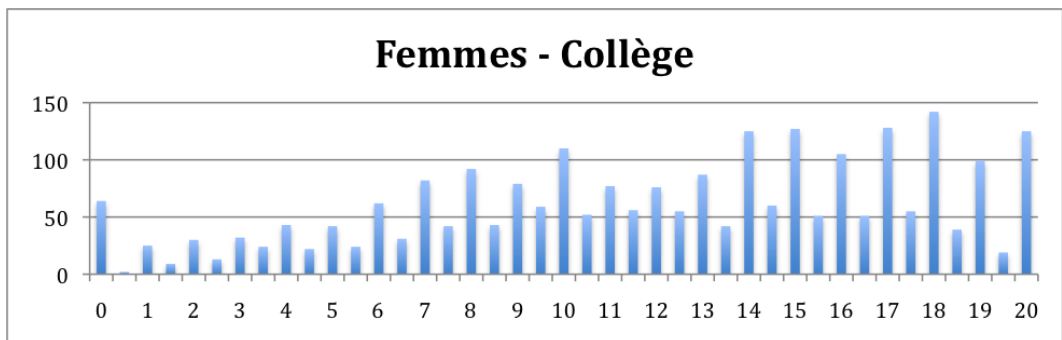


Figure 5 : Enseignantes de collège.

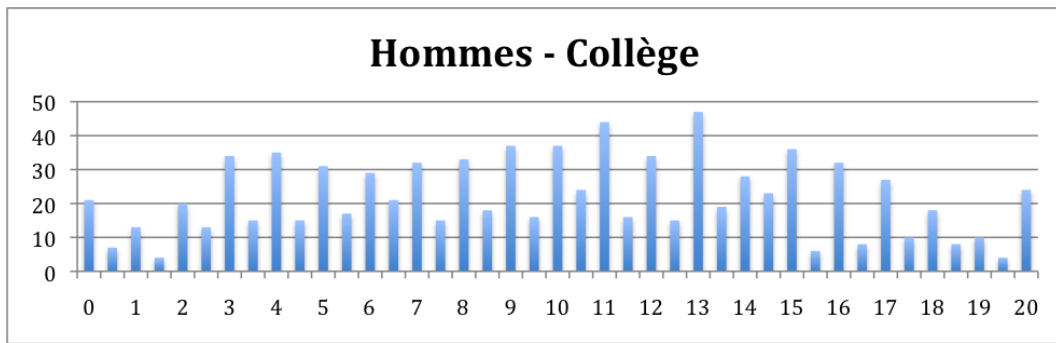


Figure 6 : Enseignants de collèges.

Au collège les hommes sont plus « sévères » que les femmes : H : 9,89 (5,20) F : 12,07 (5,29) (P<.001)

Au lycée la tendance s'inverse (sans être significative) H : 11,21 (3,65) F : 10,57 (3,72).

Retenons que les enseignants du collège sont plus « généreux » que ceux du lycée. Ce sont notamment les femmes qui marquent cette différence en devenant plus « sévères » au lycée. Ces constats devront servir de base à des investigations ultérieures, plus cliniques.

3.4 Analyse de cas particuliers

Les phénomènes relevés sur la courbe générale (figure 2), c'est-à-dire, les statuts particuliers des notes 0, 10, 11 et 20, sont également visibles chez certains enseignants. En revanche, certains d'entre eux montrent des particularités invisibles sur la courbe cumulée.

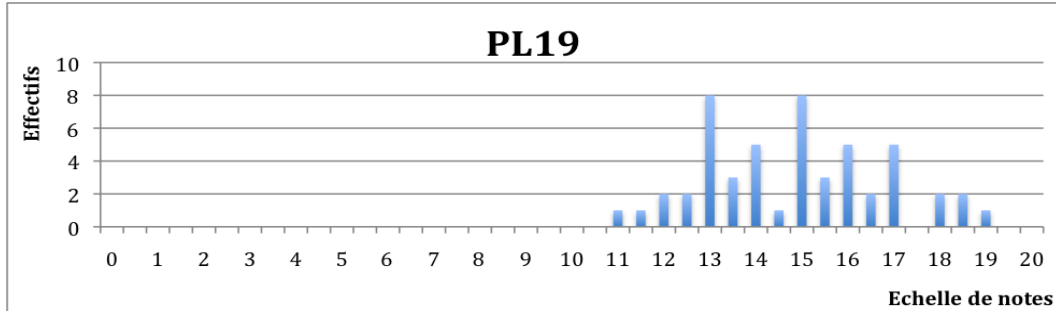


Figure 7 : Enseignant « généreux »

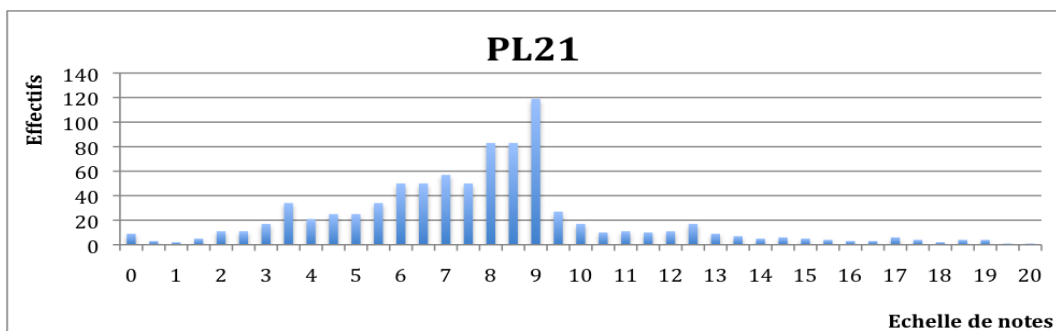


Figure 8 : Enseignant « sévère »

Les figures 6 et 7 montrent des enseignants n'utilisant qu'une moitié de l'échelle. Certains n'utilisent que la partie centrale, d'autres montrent une courbe plate sur toute l'étendue disponible. Un phénomène assez riche de développements ultérieurs consiste en une utilisation ordinale de l'échelle, c'est-à-dire, fondée sur quelques notes pivots, correspondant aux appréciations « insuffisant », « acceptable » et « satisfaisant ».

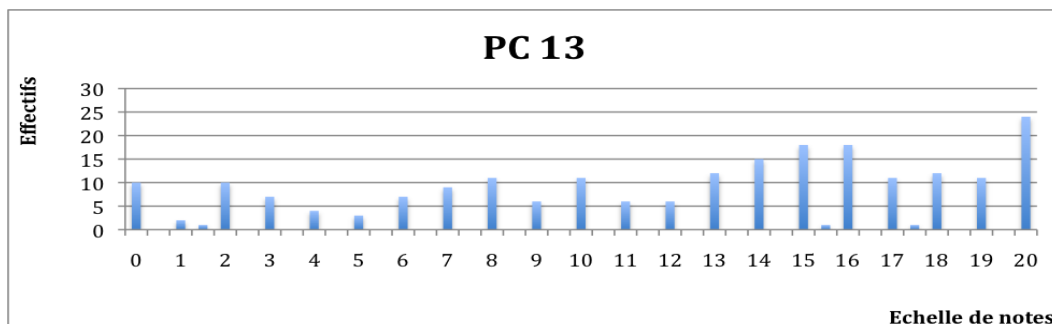


Figure 9 : Usage ordinal de l'échelle.

Chaque courbe est intéressante par ses pics et ses dépressions, inexplicables par l'observateur. Au-delà du repérage de phénomènes généraux, nous voudrions initier une prise de conscience individuelle. Cette recherche aura donc une suite consistant à présenter l'image de sa courbe de notes à chaque enseignant qui voudra la connaître et la commenter. Ainsi, tel enseignant qui sur près de mille notes ne distribue pas une seule fois 4 ! Ainsi, tel autre qui utilise la note 17 deux fois plus que 16 et trois fois plus que 18 ! etc.

4. Discussion

Dans l'état du traitement de nos données, nous ne pouvons guère proposer d'explication. En revanche, au-delà des phénomènes de docimologie bien connus, nous pouvons déjà considérer deux catégories d'explications. Celle qui tient de la psychologie consciente du correcteur, comme l'usage des demis et des quarts, l'usage d'une partie de l'échelle, etc. et celle qui tient de sa psychologie inconsciente, comme l'usage de certaines notes et l'éviction d'autres. Il sera donc très intéressant de faire s'exprimer le correcteur sur la correspondance qu'il établit entre sa conception de la pédagogie (voire sa mission et ses valeurs) et la distribution de ses notes. Encore plus passionnant (mais délicat à appréhender) la partie inconsciente qui cause la faveur ou la désaffection de certaines notes.

Quoi qu'il en soit, la posture d'enseignant est suffisamment complexe pour qu'on s'intéresse à tout apport censé améliorer la prise de conscience des attitudes. A l'instar d'Ada Abraham (1972, 1984), nous pensons que la conscience de soi est très importante quant à l'effet-maître (Bressoux, 2003). Ce n'est, évidemment qu'une entrée parmi d'autres dans la problématique mais elle a l'avantage de comporter des aspects concrets, visualisables, susceptibles de susciter des réactions. Le travail sur soi passe d'abord par la connaissance de soi !

5. Conclusion

Notre hypothèse d'utilisations particulières de l'échelle de notes est avérée. Les critères de niveau (collège, lycée), et de sexe distinguent les correcteurs. Quant aux disciplines évaluées, leur inégalité de représentation par la note est connue depuis longtemps. Mais plus précisément, les enseignants s'approprient l'échelle de manière originale en montrant des comportements remarquables vis-à-vis d'une partie de l'échelle ou de certaines notes.

A partir de ces constats, nous pouvons imaginer deux types de collaborations susceptibles de faire progresser la compréhension des phénomènes mis au jour. Par exemple, avec un mathématicien,

nous allons tenter de montrer que certaines notes sont des attracteurs de sous distributions, comme si les modes statistiques de la courbe correspondaient aux modalités d'une échelle ordinale. Par ailleurs, en collaboration avec un psychologue clinicien, nous allons approfondir notre compréhension de l'acte d'évaluer, par exemple, en enregistrant un enseignant qui accepterait de partager ses réflexions à haute voix tout au long de son travail de correction ou en menant des entretiens d'explicitations (Vermersch, 1997) ou encore en utilisant les techniques de confrontation simples et croisées ou d'instruction au sosie (Clot, 2005).

D'une manière générale, cet apport s'inscrit dans la perspective d'une prise de conscience de soi favorable au bien-être professionnel et existentiel de l'enseignant et partant, susceptible de faire partie de sa formation initiale et continue, profitable aux élèves.

6. Bibliographie

- Abernot, Y. (1987). Approche qualitative multidimensionnelle de l'enseignant. Perception et attentes des étudiants, *Recherche et formation*, 2, 21-36.
- Abernot Y. (1998). *Les méthodes d'évaluation scolaire*. Paris : Dunod.
- Abraham, A. (1972). *Le monde intérieur des enseignants*. Paris : Epi.
- Abraham, A. (1984). *L'enseignant est une personne*. Paris : ESF.
- Bressoux P. & Pansu, P. (2003) *Quand les enseignants jugent leurs élèves*. Paris : PUF
- Clot, Y. , Faïta, D. (2000). Genre et style en analyse du travail. Concepts et méthodes. *Travailler*, 4, 7-42.
- Clot, Y. (2005). L'autoconfrontation croisée en analyse du travail : l'apport de la théorie bakhtinienne du dialogue. In : L. Fillietaz, & J.-P. Bronckart (Eds.), *L'analyse des actions et des discours en situation de travail. Concepts, méthodes, applications* (pp. 37-55). Louvain-la-Neuve : Peeters.
- de LANDSHEERE, G. (1992), *Evaluation continue et examens, Précis de docimologie*, Bruxelles : Labor, 6e édition.
- Dubus, A. (2006). *La notation des élèves*. Paris : Armand Colin.
- Gilly, M. (1980). *Maître-élève : rôles institutionnels et représentations*. Paris : PUF.
- Lieury, A. & Fenouillet, F. (1996). *Motivation et réussite scolaire*. Paris : Dunod.
- Noizet, G. & Caverni, J.P. (1978). *Psychologie de l'évaluation scolaire*. Paris : PUF.
- Pieron, H. (1963). *Examens et docimologie*. Paris : PUF.
- Posthumus (1947) in : De Landsheere G. (1980). *Examens et évaluation continue. Précis de docimologie*. Paris : Nathan.
- Vermersch, P. & Maurel, M. (1997). *Pratiques de l'entretien d'explicitation*. Paris : ESF.