

8

Production des notes : les déformations du jugement

L'ignorance
n'arrange
rien

Les phénomènes qui entravent la recherche de l'objectivité dans une démarche de correction de copies ou d'autres travaux font partie de ce dont chacun a entendu parler, et parfois croit connaître, parce qu'ils sont systématiquement évoqués dans tout écrit qui traite de la notation, y compris les articles de journaux. Ils méritent cependant une attention plus soutenue, parce que la plupart d'entre eux ont pour caractéristique commune d'agir à l'insu de l'enseignant concerné. La connaissance des principaux effets connus ne vaut certes pas garantie d'y échapper, mais leur ignorance en facilite évidemment les ravages.

De nombreux éléments susceptibles de perturber l'objectivité d'une évaluation ont été repérés par plusieurs auteurs. Nous en reprenons ici une liste assez classique, sachant que les appellations peuvent varier d'un auteur à l'autre. Nous nous attachons ici à fournir les moyens d'identifier avec sûreté l'un ou l'autre des effets. Il nous paraît utile de classer les effets de distorsion en trois catégories :

- les effets qui sont liés au contexte, et à la situation concrète et matérielle de correction : il s'agit des effets de fatigue, de succession, de relativisation et de stéréotypie ;
- les effets qui relèvent de paramètres individuels du correcteur, et de son style docimologique : les effets de biais personnel et les effets de tendance centrale ;
- les effets qui relèvent en quelque sorte de l'interaction de la personne avec la situation, et en particulier de l'existence d'un barème de correction établi à l'avance, de sa qualité et de la manière de l'utiliser : les effets de flou, de contamination et de halo.

I. LES EFFETS LIÉS AU CONTEXTE

Il n'entre pas dans notre propos de traiter ici de questions largement étudiées par les spécialistes de l'évaluation, comme la note « négociée », la note « instrument de paix sociale » ou la note « transaction », bref tout ce qui relève de *l'arrangement évaluatif* au sens de Merle (1996).

L'arrangement évaluatif

« En classe de seconde, un professeur de mathématiques a l'habitude de ne comptabiliser sur les n contrôles attribués au cours d'un trimestre que $(n - 1)$ contrôles. Il s'agit d'un accord, d'un arrangement entre le maître et les élèves de cette classe de seconde. Cet arrangement professoral qui permet aux lycéens de faire disparaître leur plus mauvaise performance contribue fortement à renforcer la légitimité de l'enseignant, et celui-ci peut d'ailleurs en tirer une certaine popularité auprès de ses élèves. L'approbation des élèves est bien évidemment totale lorsque le professeur dans son rôle d'entraîneur sait assouplir la rigueur du professeur-arbitre : les lycéens apprécient d'avoir en quelque sorte un droit à l'erreur et de bénéficier d'un contrôle « qui ne compte pas » alors même que le professeur en assume la correction.

A contrario, lorsque le professeur décide de ne pas comptabiliser dans la moyenne trimestrielle un contrôle qui s'est avéré « trop facile », en quelque sorte trop bien réussi par une majorité d'élèves, la contestation lycéenne est générale : les règles du jeu évaluatif n'autorisent pas qu'un tel contrôle ne contribue pas à mesurer la valeur scolaire des élèves. L'autorité du professeur en sort fortement affaiblie. L'arrangement doit être perçu par les élèves comme une chance, une possibilité de se racheter, voire de sortir « gagnant ». Ou du moins, l'élève doit avoir le sentiment d'avoir obtenu une compensation eu égard à sa situation antérieure. Le principe de l'arrangement tient à ce que le professeur peut entretenir avec ses élèves une plus ou moins grande connivence sur la difficulté du prochain contrôle, les notes qu'il pourrait comptabiliser ou non (le cas des devoirs à la maison par exemple), les pondérations qu'il envisage, la liberté qu'il pourrait prendre éventuellement à l'égard de son barème... »

Le but de cette citation d'une longueur inusuelle est de situer clairement la frontière que nous nous sommes assignée entre les aspects techniques de la notation et ses aspects psychologiques ou sociologiques, que nous n'ignorons pourtant pas. Nous cherchons à faire œuvre de docimologie pratique, et nous nous arrêtons aux portes de l'analyse de la relation pédagogique. Aussi les effets décrits ici le sont sous l'hypothèse de la recherche d'une certaine objectivité dans la notation,

pour autant qu'elle soit possible, et sans préjudice de l'enjeu qu'elle peut constituer ensuite dans les relations entre personnes et entre groupes.

EFFET DE FATIGUE

On l'appelle aussi l'effet d'ennui, et peut-être que *lassitude* serait un méta-mot commode pour désigner ces deux variantes liées à la durée et au caractère répétitif des tâches de correction. Cet effet, qui semble relever de la compréhension immédiate, est cependant assez complexe, car les mouvements de l'humeur du correcteur (espoir, euphorie, déception, découragement) mais aussi les variations de sa condition physique : éveil, hypoglycémie (le creux de onze heures), torpeur post-prandiale (également appelée syndrome de Pickwick, d'après un personnage de Dickens, c'est la tendance à un léger assoupissement après le repas), mélancolie crépusculaire (la tristesse inexplicable quand le jour baisse et qu'on n'a pas encore allumé les lampes), épuisement, déshydratation (toujours avoir une bouteille d'eau avec soi)... peuvent avoir des effets de sens différents. Les troubles de la vigilance, du jugement et de la neutralité affective vis-à-vis des copies n'augmentent pas de manière linéaire avec le temps de correction. L'analyse d'un exemple réel, dans le prochain chapitre, montrera qu'il existe vraisemblablement une durée idéale de correction, celle-ci nécessitant sans doute d'être étalonnée selon la discipline et le correcteur. Quels que soient le correcteur et sa résistance à l'épreuve, il est clair que des durées exagérées sont très dommageables à la qualité du travail de correction. Malheureusement, si un enseignant dans sa ou ses classes peut programmer de manière judicieuse sa charge de correction, les grandes cérémonies docimologiques que sont le baccalauréat et les écrits à l'université accablent systématiquement les correcteurs de quantités énormes de copies à corriger dans des délais absurdes.

EFFET DE SUCCESSION

L'ordre dans lequel les copies sont corrigées ne dépend généralement pas du correcteur, à moins qu'il ne les trie préalablement en « tas » de qualité contrastée après une première lecture très rapide. Il s'expose dans ce dernier cas au risque de l'effet de relativisation (voir ci-après). S'il s'agit de copies d'examen, elles sont généralement disposées en piles *lifo* (*last in, first out* : le dernier élément à entrer est le premier à sortir), mais cela n'a probablement pas de conséquences en matière

de fluctuation de la qualité, car la durée de composition n'est pas un indicateur fiable du niveau de la copie. On se place donc dans le cas général, où l'ordre des copies n'a pas été organisé ni préparé. C'est dans ce contexte que peut apparaître, au hasard de cet ordre, l'effet de succession, dit aussi de *contraste* : la qualité de la copie précédente sert de référence dans l'évaluation de la copie en cours de correction : après une mauvaise copie, une copie médiocre peut être surnotée, après une très bonne copie, la même copie médiocre peut être sous-notée. L'ordre des copies devient crucial. J.-J. Bonniol (1972) consacre à l'étude de ces phénomènes sa thèse de troisième cycle. Cette étude est résumée de la manière suivante dans Bonniol et Vial (1997) :

« Il proposa à des groupes d'enseignants de corriger des séries de copies identiques, mais dans des ordres différents. En répétant cette expérience un nombre de fois suffisant, il constata deux phénomènes :

- les correcteurs notent par contraste, c'est-à-dire que la note d'une copie dépend en partie de la ou des copies précédentes et leur est en quelque sorte opposée ;
- d'une manière générale, les correcteurs sont plus sévères à la fin d'une série qu'au début...

À partir de cet effet de contraste, J.-J. Bonniol utilise la notion d'ancre (référence à l'instrument de marine). Dans une série de copies sélectionnées après plusieurs corrections, pour leurs notes moyennes, le chercheur introduit une ou plusieurs très bonnes (ancres hautes) ou très mauvaises (ancres basses) pour constater leurs effets sur les copies suivantes. Il montre, par exemple, qu'en anglais, l'effet de contraste est important surtout après les ancres basses ; alors qu'en mathématiques, ce sont davantage les ancres hautes et lourdes (plusieurs copies excellentes) qui entraînent la sous-estimation. »

À noter que la seconde remarque générale dans cette citation, « les correcteurs sont plus sévères à la fin », peut bien sûr être attribuée à l'effet de fatigue plutôt qu'à l'effet de succession.

EFFET DE RELATIVISATION

Plutôt que de juger une copie en elle-même, le correcteur l'évalue relativement au « paquet de copies » dont elle fait partie. Cela implique que le niveau de performance attendu n'est pas défini *a priori*, mais qu'il est induit des performances du groupe : c'est la performance collective qui définit la norme. Concrètement, cela peut aller jusqu'à une notation provisoire (au crayon, par exemple), qui sera ensuite révisée

Énormes
paquets
de copies

Le dessus
du panier

de manière à fournir une distribution conforme à ce qui est attendu : le classement des individus demeure généralement inchangé dans ces réajustements, mais la moyenne générale, et éventuellement la dispersion dans les cas les plus sophistiqués, sont déterminés *ex post*. On peut fournir des justifications pédagogiques à ces pratiques, comme le souci de maintenir une certaine pression sur le groupe-classe en ne « récompensant pas trop », ou au contraire d'éviter la catastrophe collective, qui entame finalement autant le crédit du professeur que celui des élèves. Il n'en reste pas moins que cela signifie clairement que le correcteur ne dispose d'aucun moyen objectif d'estimer par avance la difficulté du sujet proposé, et qu'il se fie à la réussite concrète d'un groupe d'élèves; ce faisant il se place dans la situation du baron de Münchhausen, qui, enlisé avec son cheval dans un marais, sauve sa vie en se soulevant lui-même par les cheveux, emmenant entre ses cuisses serrées son cheval pour faire bonne mesure. La relativisation devient ici une méthode systématique au service de *l'arrangement évaluatif*.

Relativisation et marécage

EFFET DE STÉRÉOTYPIE

Le professeur maintient un jugement immuable sur les productions d'un élève, quelles que soient les variations effectives de valeur de ces productions. Cet effet ne concerne pas les examens à correction unique et anonyme, mais est souvent décelé dans les classes du second degré. Il est en effet nécessaire, pour que cet effet apparaisse, que le correcteur connaisse l'auteur de la copie, et qu'il ait pu avec le temps se forger une représentation consistante de ses capacités. Tout se passe alors comme si, à l'objet à corriger, se superposaient la physionomie de l'élève et les jugements habituels qui le concernent. La pratique des devoirs communs à correction anonymée dans le second degré (collèges et lycées) est de nature à rompre salutairement de telles représentations, de même que l'usage de tests étalonnés sur une vaste population (à l'école et au collège). Cependant, en dehors de ces moments d'objectivation collective orchestrés par l'institution, rien n'interdit à un enseignant d'organiser de temps à autre une épreuve à correction anonyme, à condition bien sûr d'occulter les caractéristiques personnelles les plus évidentes (devoir sur traitement de texte plutôt que manuscrit).

Bienfaits de l'anonymat

II. LE STYLE DOCIMOLOGIQUE DU CORRECTEUR

Les effets de cette catégorie sont liés à la personnalité et aux façons d'être individuelles du correcteur. Certaines proviennent de l'inexpérience ou de l'appréhension devant les responsabilités qui sont celles de l'évaluateur, d'autres au contraire ont plutôt tendance à s'invétérer pour composer avec le temps un style personnel.

EFFET DE BIAIS PERSONNEL

Certains correcteurs sont réputés pour se montrer systématiquement sévères ou indulgents dans leur notation. L'interprétation de ces comportements en termes de générosité ou de méchanceté est éventuellement imprudente : une pratique d'indulgence systématique et excessive ne manifeste pas nécessairement une grande bonté d'âme; elle peut aussi dénoter une indifférence cynique vis-à-vis de la valeur de la discipline enseignée et de ceux qui font l'effort de l'étudier, et le simple souci d'éviter tout contentieux, voire de diminuer le nombre de copies de la seconde session. Dans le cas où la discipline est en concurrence avec d'autres dans un système d'options, il peut s'agir d'une forme de *dumping* docimologique visant à garantir un flux de clientèle et la permanence confortable d'une offre de formation qu'il n'est pas utile de faire évoluer, puisque la réputation de facilité amène chaque année un nombre suffisant de demandeurs. On est là dans les cas extrêmes de *l'arrangement évaluatif*. De telles situations, déontologiquement très regrettables, sont parfois difficiles à déceler, car personne ne se plaint. Seule la vigilance docimologique d'un président de jury ou un incident imprévu, comme des copies appartenant normalement à un autre enseignant et corrigées — généreusement comme toujours — sans être lues, peuvent révéler ce type de pratiques et permettre d'y mettre fin. À l'inverse, une tendance individuelle à noter bas est souvent liée à une comparaison implicite entre les travaux réels qu'on corrige et le souvenir *imaginaire* des performances qu'on avait au même âge. La compétition entre l'idéal et la réalité tournant souvent au désavantage de celle-ci, cette attitude engendre parfois des postures du genre « le niveau baisse », imperméables aux études démontrant que la réalité est bien plus nuancée. Cette distorsion n'est pas nécessairement le fait d'enseignants âgés et aigris, on la rencontre aussi chez des débutants, anciens *bons élèves* projetés dans le réel d'une classe mélangée, et sans autre étalon que leur image d'eux-

La l'él

mêmes. Désignée dans ces cas comme une tendance à l'*hypercorrection*, c'est-à-dire à l'élévation injustifiée des critères de niveau, elle est assez fréquente chez les enseignants débutant à l'Université pour que la mise en garde contre cette tendance fasse partie des conseils rituellement distribués par les collègues plus anciens.

EFFET DE TENDANCE CENTRALE

De crainte de surévaluer ou sous-évaluer, et en l'absence de règles solides et un peu objectives d'évaluation des productions, le correcteur groupe ses notes vers le centre de la distribution : il note par exemple entre 8 et 12. Cela peut se présenter chez des enseignants débutants, mais de telles timidités systématiques peuvent devenir un élément structurel du style docimologique. Cette pratique, qui peut sembler découler du *primum non nocere* (en premier, ne pas nuire) hippocratique, équivaut à un refus de prise de responsabilité dans les démarches d'évaluation. Même inspirée par le souci de ne pas causer de tort, elle peut en causer indirectement, par exemple en dévalorisant les efforts des élèves dans la discipline concernée. On verra dans un autre chapitre que dans le cas d'une note destinée à être additionnée à d'autres, avec ou sans coefficient, cette pratique a pour effet de diminuer le poids de cette note dans le total général. Elle équivaut alors à une sorte d'autocensure, voire à une castration docimologique.

Timidité
et auto-
censure

III. LES DÉFAILLANCES DE BARÈME

Les distorsions de cette dernière catégorie sont celles que l'usage effectif d'un barème bien construit pourrait rendre plus rares, sans rêver pour autant à la correction parfaite, dont on sait bien qu'elle est hors d'atteinte, et que la croyance qu'on y est parvenu serait peut-être la plus dangereuse des dérives.

EFFET DE FLOU

Les compétences recherchées et les critères de notation ne sont pas définis avec précision, de sorte que d'une copie à l'autre ce ne sont pas les mêmes éléments qui sont pris en compte, ni les mêmes critères appliqués. Une copie sera appréciée sur la rigueur de sa construction, une autre sur l'agrément de son style, une autre sur la qualité de restitution des connaissances apportées par le cours, une autre

sur l'astuce argumentaire, et que sais-je encore... Cette situation, qui relève de l'appréciation impressionniste, se produit évidemment en l'absence de tout barème, mais cette absence n'est elle-même que le symptôme d'une incertitude sur ce qu'il convient de mesurer, non seulement en termes de degré d'accomplissement (niveau), mais en termes de performances attendues. Dans des situations comparables, qui constituent la caricature de ce qu'on a pu reprocher à un certain type de correction des dissertations littéraires et philosophiques, ou à des épreuves de culture générale, le critère ineffable de l'*élégance* est parfois mis en avant par les correcteurs, et le critère secret d'affinité avec les choix esthétiques ou idéologiques du correcteur soupçonné, à tort ou à raison, par les auteurs des copies. Dans tous les cas, aucune communication pédagogique utile ne peut s'appuyer sur les résultats de ce type de correction : nul ne sait vraiment *pourquoi* il obtient cette note-là, ni *comment* il peut faire pour en obtenir une meilleure, ou une aussi bonne.

EFFET DE CONTAMINATION

Un barème peut avoir été établi, mais l'étanchéité des boîtes mal contrôlée : la richesse des connaissances restituées est minorée parce que l'organisation du texte est défaillante ; une conclusion brillante sauve un développement confus... Les notes partielles attribuées séparément aux différents aspects d'un même travail s'influencent mutuellement : les couleurs se mélangent et une dimension pertinente de l'évaluation prend une importance exagérée en oblitérant d'autres dimensions.

Une interprétation discutable

Les travaux de Caverni (1975) que Bonniol et Vial (1997) présentent comme représentatifs de l'effet de contamination ne nous semblent pas du tout concerner cet effet :

J.-P. Caverni (1975) propose à deux groupes d'enseignants de sciences naturelles de noter quatre copies différentes. Il donne à chaque correcteur un faux dossier scolaire par copie : le premier groupe de correcteurs dispose d'appréciations antécédentes élevées et stables, le deuxième groupe se fonde sur des dossiers aux notes faibles et dispersées...

Il nous semble qu'il s'agit plutôt d'un effet extra-docimologique, celui de l'influence d'un jugement externe et préétabli, voire du jugement sous sa forme bureaucratiquement légitimée, le dossier. En d'autres termes, il ne s'agit pas

réellement d'une expérience de notation, mais d'une variante négative d'une dimension de l'effet *Pygmalion* selon Rosenthal et Jacobson (1968), celle qui concerne l'altération du jugement des maîtres sous l'effet de données fabriquées, et non – on l'espère – celle qui concerne l'évolution concomitante des performances des élèves. Rappelons brièvement le propos central de l'expérience de Rosenthal et Jacobson : au moyen d'un test et de résultats truqués, les expérimentateurs font croire à des enseignants que certains de leurs élèves (en réalité tirés au sort) ont manifesté dans le test les prémices d'une prochaine phase importante de développement cognitif. Après plusieurs mois, une nouvelle mesure des capacités des élèves, au moyen de véritables tests usuels cette fois, montre que les enfants désignés comme « devant progresser » ont effectivement progressé nettement plus que les enfants pour lesquels cette prophétie n'a pas été faite. L'interprétation classique de l'expérience est que ce sont les *attentes positives* des enseignants qui *modèlent* un meilleur apprentissage des élèves désignés.

EFFET DE HALO

Le correcteur est influencé par des caractéristiques périphériques du travail : présentation, soin, écriture, orthographe, et surestime ou sous-estime *inconsciemment* la valeur du travail. À la différence de l'effet de contamination, où les éléments influenceurs sont légitimement critériés, mais avec une importance exagérée, l'effet de halo met en jeu des critères qui, formellement, n'ont pas vocation à être pris en compte, parce qu'ils n'ont pas été introduits dans le barème. L'effet de halo est lié à des éléments externes, alors que la contamination est associée à des éléments internes. L'effet de halo est parfois sous-tendu par un modèle implicite, à base culturelle, des critères auxquels *un travail de ce niveau-là* doit satisfaire pour être pris en compte, sans pour autant apporter de « rémunération » notée ; ce modèle est tout à fait recevable, à condition d'être public et explicite. D'une part, si on décide d'employer un barème (et nul n'y est contraint), il paraît nécessaire de le rendre *complet* et de ne pas en sortir. D'autre part, si l'évolution des populations qui sont soumises à certaines épreuves fait que certains sub-critères de correction langagière ne sont plus satisfaits, de manière significative, alors qu'on souhaiterait qu'ils le fussent, il conviendrait d'en prendre acte et de leur faire une place visible dans les barèmes de correction. Pas plus qu'il ne peut résulter de l'assemblage de boîtes à l'étanchéité incertaine, un barème ne peut comporter de critères clandestins, ou alors ce n'est plus un barème. Même quand on en est averti,

il n'est pas facile d'échapper à l'effet de halo, ne serait-ce que parce que les compétences officiellement évaluées et les compétences qui en principe ne le sont pas vont parfois de pair, et que c'est une caractéristique professionnelle de l'enseignant de pouvoir prélever rapidement des indices de qualité d'un travail sur des éléments de surface.

L'expérience de Nimier

Jacques Nimier (2002), dans son site personnel sur Internet, rapporte avoir, avec des enseignants de l'IREM de Reims, fait corriger vingt copies de BEPC par trente professeurs de mathématiques après établissement d'un barème. Parallèlement, Nimier a noté les copies sans les lire, avec un simple coup d'œil à l'aspect global de la copie (écriture et présentation). Le maximum de l'écart pour une même copie était de 11 points sur 20, et la note de Nimier n'était jamais à une extrémité de la distribution pour une même copie. En d'autres termes, la valeur prédictive d'un coup d'œil esthétique-pifométrique n'était pas pire que celles des corrections sérieuses d'un groupe de spécialistes.

Le véritable danger de l'effet de halo correspond donc aux cas où la performance quant à ce qui est en principe à évaluer entre en contradiction avec les signes superficiels : le jugement du premier coup d'œil peut étendre subrepticement son influence à l'ensemble de l'appréciation.

De quel effet s'agit-il ?

Saurez-vous identifier quel (et parfois quels) effet(s) de distorsion sont présents dans les saynètes ci-dessous, où ce qui est transcrit est la voix « off » dans la tête du correcteur ?

1. Est-ce qu'on peut se rendre compte de l'effet que ça fait de lire cent vingt-neuf fois les mêmes niaiseries ? Ce paquet de copies du bac est trop épais, et les délais sont trop courts. Il est bientôt deux heures du matin, et j'ai les yeux comme un lapin russe. Basta, 4/20, c'est bien payé !
2. Je n'ai pas encore entamé mon paquet de copies, mais je suis prêt à parier à cinq contre un que l'élève Vandembrouck Marcelline aura entre 11 et 13, que l'élève Lopez Arthur ne fera guère mieux, et que Jérémie Dupont atteindra difficilement 9. En revanche, je serais très déçu si Nora Benali faisait moins de 16.
3. Je note d'abord toutes mes copies au crayon, et après je calcule ma moyenne, mon écart-type, tout, quoi. Je décide alors si je dois enlever ou rajouter des points à tout le monde pour arriver à un résultat convenable. Je reprends mes copies, je gomme et je mets la nouvelle note.

4. Je trouve ça dégoûtant de mettre des mauvaises notes. D'ailleurs j'ai horreur d'écrire en rouge sur le travail des autres. Alors je m'arrange pour noter au-dessus de 10. Si c'est vraiment nul, je mets 8 ou 9, pour envoyer un message à l'élève.

5. C'est vrai que les résultats à la fin des exercices ont l'air d'être exacts, mais l'encre qui commence noir et qui finit gris-bleu, le coin de la feuille chiffonnée, et puis, franchement, « ipoténuze » !!!! Ça ne peut pas valoir plus de 12. Allez, hop, suivant.

6. La fin du développement est un peu faible et la conclusion quasi absente. D'après mon barème, ça va faire dans les 12/20. Oui, mais l'introduction était excellente, et le vocabulaire étudié en cours a été réemployé avec bonheur. Je remonte à 14.

7. Comment cet élève ne parvient-il pas à transcrire clairement la démonstration dans le deuxième exercice ? Ce ne doit pourtant pas être hors de portée, c'était correct dans toutes les copies précédentes ! Je dois avoir affaire à un attardé. Allez, zéro pour l'exo.

8. Toujours aussi médiocre, Benjamin Ducrocq ! Jamais fichu d'aligner une phrase bien construite en commentaire de texte. Ah, là, ça a l'air moins idiot que les autres fois ! Ouais, bof, sans doute une illusion d'optique. Pas que ça à faire, 8/20 comme d'habitude.

9. Wouah ! Après le tombereau de sottises que je déblaie depuis des heures, enfin des phrases avec un sujet, un verbe et un complément ! Niveau philosophique c'est pas extra, mais au moins ça se lit ! 18/20, soyons fous !

10. Certes, cette copie n'est pas entièrement mauvaise, et on pourrait même dire d'un certain point de vue que tous les points requis ont été traités correctement. Oui, mais de mon temps, on était autrement brillant que ça, et je ne suis pas de ces démagogues soixante-huitards qui donnent des points comme s'il en pleuvait. 14/20.

Les réponses

1. Évidemment *fatigue* (lassitude, ennui).
2. *Stéréotypie* : le jugement sur l'élève est constant, la performance réelle est secondaire.
3. Un cas très caricatural de *relativisation* systématique, avec peut-être une trace d'intoxication par la loi normale (cf. chapitre 9).
4. Un *biais personnel* de type « généreux », lié en fait à la crainte du rapport d'évaluation lui-même (mais nous sortons du sujet...).
5. Un très bel effet de *halo* : la performance réelle est oblitérée par des aspects latéraux qui ne font pas normalement l'objet de l'évaluation.
6. Dans ce cas, l'effet le plus évident est l'effet de *contamination* : divers aspects, qui auraient dû demeurer distincts, s'influencent ; une trace de *flou* n'est pas à exclure.

7. C'est un effet de *succession/contraste*.

8. À nouveau, un effet de *stéréotypie*. Quoi qu'il fasse, Benjamin Ducrocq n'aura jamais une bonne note.

9. Un autre effet de *succession/contraste*, sans doute à la faveur d'une baisse de vigilance due à la *fatigue* ; mais on décèle aussi un peu de *contamination*, ou de *halo*, selon que la qualité syntaxique fait ou non partie des éléments à évaluer.

10. Un effet de *biais personnel* « vache », nourri par les faux souvenirs de la belle jeunesse du correcteur et ses présupposés idéologiques.

Résumons-nous

► Dans ce chapitre consacré aux distorsions du jugement de l'enseignant dans les tâches de correction, nous avons d'abord exclu du champ de l'étude ce qui relève de l'arrangement évaluatif, pour concentrer l'attention sur les effets actifs hors de toute stratégie relationnelle et avec le présupposé d'honnêteté et de bonne volonté docimologique.

Nous avons décrit les effets suivants : fatigue (ennui, lassitude), succession (contraste), relativisation, stéréotypie, biais personnel, tendance centrale, flou, contamination, halo.

► Le principal conseil qu'on puisse adresser à l'enseignant actuel ou futur est évidemment la *vigilance*. Toutefois, celle-ci peut difficilement s'exercer simultanément sur tous les aspects, et si l'on souhaite exercer de temps à autre une certaine méfiance docimologique à l'égard de sa propre pratique de correction, si on en a une, on recommandera, après une série de corrections, de reconsidérer l'ensemble du travail sous l'angle d'une et d'une seule des distorsions connues, en se posant la question pour chaque copie : tel effet a-t-il pu jouer dans l'élaboration de la note attribuée à ce travail-là. C'est en effet par l'analyse attentive d'objets concrets, et non par la récitation des principes, que les pratiques docimologiques peuvent s'améliorer.