



# Fiche Outil

## Comment évaluer les **compétences** et les **performances** ?



« On compte, on calcule, on compare, on additionne... tout ça pour une note. Mais une note ça sert à quoi ? »

**Marie-Camille Coudert** - Enseignante de physique-chimie en collège et lycée, spécialisée dans le domaine de l'évaluation des apprentissages

On a tellement l'habitude de la note à l'école que questionner son utilité nous paraît impossible. Pourtant, c'est bien parce qu'elle nous demande tant de temps qu'il faut réfléchir à cette pratique. Une note ça sert à qui ? **Spoiler : pas aux élèves.**

**Objectif :** prendre conscience de certains biais qui empêchent la note d'être précise.

### Conseils d'utilisation de la fiche

- 12 questions pour faire le tour de la question de la note.
- 12 questions pour tester vos représentations et les déconstruire si besoin

### Faire un diagnostic de sa pratique

Q1. La note a été inventée pour classer les candidats au concours de sortie de Polytechnique

VRAI  FAUX

Au début du 19e siècle, l'école Polytechnique, fraîchement créée, éprouve le besoin de classer ses élèves à la sortie. Ils mettent en place une notation sur 20 points, évitant ainsi les ex-aequo. La note est donc un outil qui permet de classer les apprenants, elle a été créée pour cela.

Q2. Au 19e siècle, le baccalauréat était évalué avec des smileys.

VRAI  FAUX

Faux, mais pas tant que ça. Au 19e siècle, le baccalauréat est un examen oral, le candidat est évalué par un jury de 3 personnes. Chaque membre du jury évalue à l'aide d'une boule de couleur (les smileys de l'époque...). Boule rouge : admis, boule blanche : douteux, boule noire : mauvais. Le



candidat pouvait prétendre à une mention s'il n'obtenait que des boules rouges. La note sur 20 arrive au baccalauréat en 1890 puis par la suite se répand dans tout le secondaire puis le primaire.

Q3. La note varie d'un correcteur à un autre. L'écart peut aller jusqu'à 9 points pour une copie notée sur 20.

VRAI  FAUX

Il s'agit de l'expérience fondatrice de la docimologie (la science de l'évaluation en tant que mesure). En 1930, le professeur Laugier refait corriger 166 copies du concours de l'agrégation d'histoire. Les résultats furent étonnants : la moyenne des notes du premier correcteur dépassait de près de deux points celle du second. Le candidat classé avant dernier par l'un était classé second par l'autre. Les écarts de notes allaient jusqu'à 9 points. La moitié des candidats reçus par un correcteur était refusée par l'autre.

Q4. L'établissement d'un barème précis permet de diminuer voire d'annuler les disparités entre correcteurs.

VRAI  FAUX

Je vais illustrer ce point par une expérience réalisée en 1975. Un groupe de correcteurs a travaillé pendant près de 15h à faire un barème très précis (au 6e de point près !) du BEPC. Malgré cet effort, on retrouve des copies dont la note varie entre 4 et 13. Pire, même après confrontation des correcteurs, il a été impossible de donner raison à l'un ou à l'autre. Pour la même réponse, certains voyaient «une réponse peu claire», d'autres «un bon raisonnement» et encore d'autres «un raisonnement faux».

Il est admis aujourd'hui qu'une même copie corrigée avec un barème au point (chaque question est divisée en sous-questions pour atteindre le barème : une question = un point) génère plus de différences qu'un barème large (les grandes questions sont notées avec un barème à 4 ou 5 points). La raison est liée au fait qu'avec un barème au point nous sommes forcés d'attribuer 0 ou 1 à une question. Avec un barème large, nous hésitons plus à attribuer 0 ou 5. Or une somme de vingt 0 est égale à 0...

Q5. Une double correction (par deux correcteurs différents) permet d'obtenir une note exacte.

VRAI  FAUX

Par note exacte, on entend une moyenne de notes telle que l'adjonction d'une autre note ne modifie pas sensiblement cette moyenne. Les expériences de docimologie montrent que le nombre de correcteurs pour obtenir cette note exacte dépend de la matière, il est de 127 correcteurs en philosophie, 78 en composition française, 28 en anglais, 19 en version latine, 16 en physique, 13 en mathématiques. La double correction est donc un processus illusoire. Il n'améliore que peu la précision de la notation.

Q6. Une copie notée 15 est meilleure qu'une copie notée 13.

VRAI  FAUX

Si cette copie est corrigée par des évaluateurs différents, nous avons vu que la dispersion des notes peut atteindre 9 points pour une même copie.

Pour un seul correcteur, les résultats ne sont pas meilleurs. Des chercheurs ont ainsi fait l'expérience suivante : à sa demande un professeur de physiologie de la Faculté des Sciences accepta 37 copies - dactylographiées et anonymes - qu'il avait corrigées trois ans et demi auparavant. Le degré d'accord de ce professeur avec lui-même ne fut pas plus élevé qu'avec deux de ses collègues chargés de la même tâche : les coefficients de corrélation atteignant respectivement 0,58, 0,59 et 0,56.



L'admissibilité, avec ses nouvelles notations, aurait été modifiée ; la moitié des précédents admissibles aurait été refusée et la moitié des refusés, déclarée admissible. Dans 7 cas seulement, il remit la même note au même devoir. Dans les 30 autres cas, il y eut des divergences comprises entre 1 et 10 points.

**Q7. Les bons élèves sont surnotés.**

VRAI  FAUX

C'est ce qu'on appelle «l'effet de halo», celui qui fait que les bons élèves (= ceux que l'enseignant juge bons) performant mieux et sont mieux notés, juste à cause du regard de l'évaluateur (et non de leurs performances propres).

Cet effet a été démontré en 1968 par Rosenthal et Jacobs. Ils ont confié 6 rats à deux équipes d'étudiants. Les étudiants devaient apprendre aux rats à traverser un labyrinthe. A une équipe d'étudiants, ils ont dit que les rats avaient été rigoureusement sélectionnés, c'étaient des rats "top-niveau". Inversement, l'autre groupe a été informé que leurs rats avaient des difficultés, voire des dégénération génétiques.

Bien évidemment, tous les rats étaient issus de la même lignée et associés au hasard. Et pourtant, les rats "brillants" ont tous traversé le labyrinthe alors que certains rats déficients n'ont même pas franchi la ligne de départ. Pour en savoir plus, [cliquez ici](#).

**Q8. Les bons élèves résistent mieux à la pression que les mauvais élèves.**

VRAI  FAUX

Ce résultat paradoxal a été démontré en 2010 par Sian Beilock, chercheuse américaine ayant travaillé sur l'effet du stress sur les conditions d'examen ou de compétition sportive. Plus généralement, la performance d'un élève lors d'un examen dépend de beaucoup de conditions extérieures, le stress bien sûr mais aussi la fatigue.

Pour en savoir plus, vous pouvez aller lire le livre de Sian Beilock [«why we choke under pressure»](#).

**Q9. Les enseignants jeunes notent plus sévèrement que les enseignants plus expérimentés.**

VRAI  FAUX

Eh oui, des études montrent qu'en moyenne, nous avons tendance à devenir de moins en moins sévères avec l'âge.

**Q10. Les filles sont mieux notées que les garçons.**

Attention, plusieurs réponses sont possibles

VRAI  FAUX

En fait, cela dépend des matières. Effectivement, dans les langues ou les humanités, se sont les filles qui seront sur-notées. Alors que dans les disciplines scientifiques, l'évaluateur met des meilleures notes aux garçons à performance égale.

Plus généralement, la note dépend de ce que l'évaluateur sait des conditions sociales ou autres des candidats. C'est ce qu'on appelle l'effet de contexte.

**Q11. Pour avoir une meilleure note, il vaut mieux que ma copie soit en fin de paquet, derrière une mauvaise copie.**

VRAI  FAUX

Ces résultats ont été démontrés de façon répétée par des expériences de recherche. L'évaluateur est plus sévère en début de séance de correction. De même, il est influencé par la copie précédente. Il note plus sévèrement après une copie réussie et à tendance à sur-noter après une mauvaise copie. C'est notamment pour cela qu'on conseille de corriger les copies d'examen exercice par exercice et non copie par copie.



Q12. Un bon évaluateur obtient une courbe de Gauss sur ces notes.

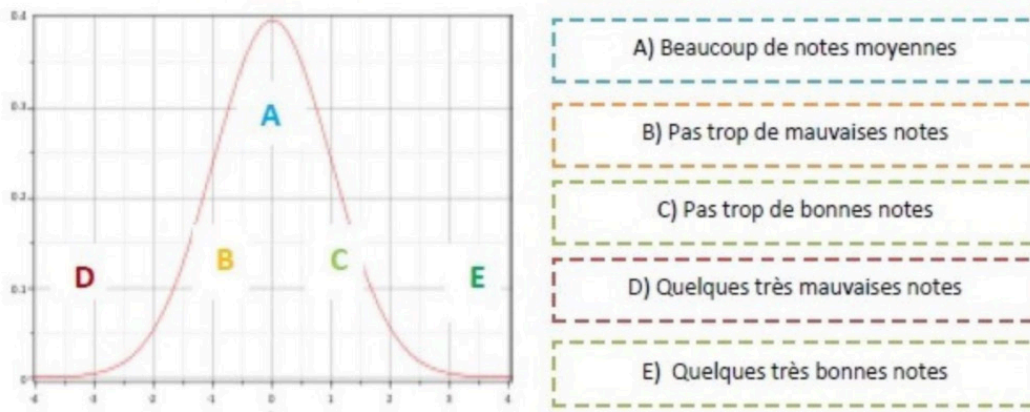
VRAI  FAUX

L'échantillon d'une classe est beaucoup trop petit pour qu'il soit normal d'obtenir cette fameuse courbe de Gauss, vous savez, celle en cloche. Il faudrait des échantillons de l'ordre d'un millier d'étudiants, pris au hasard (et non au sein d'une même classe).

Et pourtant, on a tendance inconsciemment à reproduire cette courbe de Gauss. Ce qui fait qu'on a toujours une partie des élèves « pas bons », un ventre mou et des élèves de « tête de classe ». Autrement dit, quelle que soit la classe, on a tendance à adapter la difficulté du sujet et la sévérité de la correction pour obtenir un nombre de bons et de mauvais résultats qu'on juge acceptable.

C'est ce qu'on appelle la loi de Posthumus. Le chercheur A. Antibi montre que cela exclut de façon automatique les étudiants en queue de classe, c'est la constante macabre. En effet, si je prends les 35 meilleurs pâtissiers de France et que je leur pose un examen, je vais me débrouiller pour qu'une partie échoue. Et donc dans toutes les classes, une partie des étudiants ne réussira pas, d'où le nom de constante macabre. Ce mécanisme est inconscient ce qui explique qu'il soit si difficile à éradiquer.

*Courbe des notes lors d'une évaluation illustrant la notion de constante macabre*



Pour en savoir plus, [cliquez ici](#).

### Analyser

Vous voilà informé-e de ce qui se joue dans la note et ce qu'on y projette.