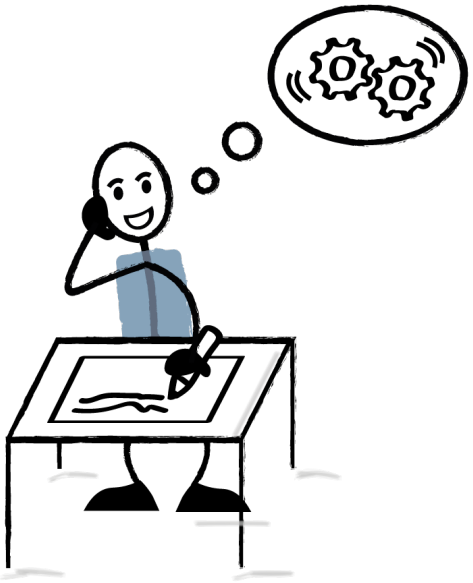
ACADÉMIE
DE GRENOBLE
*Liberté*
*Égalité*
*Fraternité*

# MEMENTO DOCIMOLOGIE

## Les principaux biais et effets

Mis à jour en mai 2025



# MEMENTO : Les principaux biais docimologiques dans l'évaluation scolaire

L'évaluation est une tâche complexe et essentielle dans notre métier au quotidien, mais il est important de reconnaître que, malgré nos meilleures intentions, nos jugements peuvent être influencés par divers biais. Ces biais sont largement étudiés par les chercheurs, et sont inhérents à toute activité d'évaluation. Ils doivent être explicites (explicitement communiqués/précisés) afin de nous permettre de construire progressivement des stratégies pour éviter ces écueils. En effet, en prenant conscience de leur existence, nous faisons déjà un pas vers une évaluation plus équitable et rigoureuse.

Dans les sections suivantes, nous présenterons les principaux biais identifiés par la recherche. Cette prise de conscience est cruciale pour améliorer nos pratiques et garantir une évaluation plus juste et plus robuste pour tous nos élèves.

## Objectif

Identifier les biais susceptibles d'altérer l'objectivité et la fiabilité des évaluations afin de promouvoir une évaluation plus juste et plus robuste.

## Définition

Dans les années 1920, Piéron et Laugier fondent la docimologie : c'est la science de l'évaluation. Elle analyse les conditions de l'évaluation scolaire, en identifiant les facteurs qui peuvent altérer la validité, la fidélité et l'équité de la notation.

## Conséquences pédagogiques

- Baisse de confiance des élèves en l'équité scolaire
- Autoévaluation biaisée
- Renforcement des inégalités scolaires
- Difficultés à adapter les pratiques pédagogiques

## Le saviez-vous ?

Laugier et Weinberg ont calculé que, pour obtenir une « note vraie » aux épreuves du baccalauréat à la fin des années 19201, il faudrait recourir à la moyenne des notes de 13 correcteurs en mathématiques, 78 en composition française et 127 en philosophie.

1 la « note vraie » correspond à la valeur scolaire d'une copie avec une marge d'erreur statistique faible

# Synthèse

**Deux classes de biais docimologiques**

Biais cognitifs :

- Biais de contraste ou d'ordre
- Effet d'assimilation
- Biais de position
- Effet Posthumus
- Biais de tendance centrale
- Effet de sévérité et l'effet d'indulgence
- Biais de légitimation

Biais sociaux :

- Biais de Halo ou de contamination
- Effet de stéréotype
- Effet de classe sociale

**Conséquences pédagogiques**

- Baisse de confiance des élèves en l'équité scolaire
- Evaluation biaisée
- Renforcement des inégalités scolaires
- Difficultés à adapter les pratiques pédagogiques

**Synthèse des préconisations en fonction des biais.**

Pour réduire les biais, il est essentiel de privilégier les évaluations critériées. Ces évaluations utilisent des descripteurs clairs, lisibles et expliqués en amont, avec des attendus explicites pour chaque seuil. Cela permet d'éviter l'effet d'une évaluation globale et générale, souvent sujette aux biais. En outre, l'évaluation par compétences est particulièrement bénéfique, car elle est plus précise et informe les élèves sur ce qu'ils maîtrisent et ce qu'ils doivent encore améliorer pour atteindre les objectifs fixés par l'évaluateur.

Dans le cadre des examens, l'anonymat (la double correction pour les concours) permet de contourner certains biais.

| Biais de halo ou de contamination | Description |
| --- | --- |
| Piéron, 1975<br>Richard E. Nisbett et Timothy D. Wilson, 1977 | L'effet de halo, effet de notoriété ou encore effet de contamination, est un biais cognitif qui affecte la perception des gens ou de marques. C'est une interprétation et une perception sélective d'informations allant dans le sens d'une première impression (« il ne voit que ce qu'il veut bien voir »).<br>La contamination par la présentation - La qualité de l'écriture, la propreté de la copie ou la mise en page peuvent influencer inconsciemment la notation du contenu. Une copie soignée peut recevoir une meilleure note qu'une copie au contenu identique mais à la présentation négligée.<br>La contamination par des évaluations antérieures - Les performances passées d'un élève peuvent influencer la perception de son travail actuel. Un élève habituellement d'un bon niveau académique sera évalué plus favorablement même en cas de baisse de qualité.<br>La contamination par l'impression générale - Une première impression positive ou négative peut "contaminer" l'évaluation de toutes les parties de la copie, créant un effet de halo.<br>La contamination par des facteurs personnels - Des éléments comme le nom, le genre, l'origine ethnique ou sociale de l'élève peuvent inconsciemment influencer l'évaluation.<br>La contamination par le langage ou le style - Le style d'expression peut influencer la perception du contenu, indépendamment de sa qualité réelle.<br>Il est important de relever que la simple connaissance de cet effet ne permet pas de s'en prémunir. |

| Effets sociaux | Description |
| --- | --- |
| Jussim & Harber, 2005<br>Bonefeld & Dickhäuser, 2018<br>Rangvid, 2015<br>Sprietsma, 2013 | Il existe un effet très proche, l'effet de stéréotype : les attentes liées au genre, à l'origine ou au statut social influencent l'évaluation. (Jussim & Harber, 2005). Plusieurs études récentes montrent que les évaluateurs attribuent des notes différentes à des élèves de statuts sociaux ou d'origines différentes, même à performance égale. Par exemple, Autin et al. (2019) ont montré que les enseignants trouvaient plus d'erreurs dans la copie d'un élève de classe sociale défavorisée que dans celle d'un élève favorisé, à performance identique. Ce biais disparaît lorsque l'évaluation est formative plutôt que normative. Des résultats similaires sont observés pour l'orientation scolaire (Batruch et al., 2019). D'autres travaux (Bonefeld & Dickhäuser, 2018 ; Rangvid, 2015 ; Sprietsma, 2013) confirment que les élèves d'origine immigrée ou de statut socio-économique faible sont plus souvent sous-évalués, même à niveau égal avec leurs pairs. |
| **Stratégies pour éviter ces effets.** | • Anonymat.<br>• Centrer l'évaluation sur des aspects factuels, clarifiés en amont.<br>• Evaluation « en bosse ». Stratégie d'interrogation orale où l'enseignant cherche à faire cranter l'élève vers le haut en s'appuyant sur des critères explicites. |

| Biais de contraste ou effet d'ordre de corre | Description |
|---|---|
| Bonniol,1965 | L'évaluation est influencée par la performance des copies précédentes.<br>L'influence des copies précédentes sur l'actuelle, surtout s'il y a un grand contraste, par exemple une copie sera jugée bonne si la précédente était médiocre ; ou bien les copies corrigées en premier obtiennent des notes moins bonnes que les dernières. |
| Effet d'assimilation | Description |
| Sherif, 1958 | L'effet d'assimilation est étroitement lié à l'effet de contraste, mais fonctionne dans la direction opposée. Ces deux effets montrent comment nos jugements sont influencés par les points de référence (ou "ancres") que nous utilisons, souvent inconsciemment. Dans le cas d'ancrage similaire, le correcteur peine à discriminer les copies entre elles. |
| **Stratégies pour éviter ces biais** | • L'utilisation de grilles d'évaluation précises qui explicitent clairement les critères et les attendus pour chaque seuil ou niveau de maîtrise.<br>• La correction "à l'aveugle" (sans connaître l'identité de l'élève).<br>• Le mélange aléatoire des copies avant correction.<br>• La double correction par des évaluateurs différents dans un ordre inversé dans le cadre de concours de recrutement.<br>• La relecture des premières copies corrigées. |

| Biais de position | Description |
| --- | --- |
| Fraisse & Piéron, 1963<br><br>Scribner & Cole, 1973 | Les copies en début ou fin de lot sont notées différemment de celles du milieu. Ce biais existe également pour les évaluations orales. Ce biais est également à associer à l'effet de fatigue. Les notes peuvent devenir plus sévères ou au contraire plus laxistes à mesure que l'évaluateur avance dans sa tâche et se fatigue. |
| **Stratégies pour éviter ces biais** | • Analyse statistique et attention particulière à des évaluations en dehors de l'espace de confiance.<br>• Réfléchir au rythme de correction ou d'évaluation pour instaurer une pause cognitive entre deux copies ou candidats.<br>• Les évaluations utilisant des descripteurs d'évaluation lisibles, clairs, et des attendus explicites par degré d'acquisition. |

| Effet posthumus ou «constante macabre» | Description |
|---|---|
| Antibi, 2003<br><br>Crahay, 2019 | Loi de Posthumus (1947) les évaluateurs ont tendance à vouloir toujours retrouver la même distribution de notes autour d'une même moyenne et cela quel que soit le niveau de la classe concernée (les publics les plus performants sont jugés plus sévèrement et vice versa) : « quelle que soit la distribution des compétences au début de l'année scolaire, la distribution des notes en fin d'année épouse grosso modo une forme gaussienne ». Cette loi est aussi connue sous le nom de « constante macabre » (Antibi,2003) : la proportion de mauvaises notes est similaire quel que soit le sujet de l'examen et quel que soit le correcteur, indépendamment de la qualité véritable des réponses données par ceux qui passent l'examen. |
| **Stratégies pour éviter ces biais** | • L'évaluation par contrat de confiance (EPCC) - Méthode proposée par Antibi lui-même, où l'enseignant annonce clairement à l'avance ce qui sera évalué et prépare les élèves spécifiquement à ces contenus.<br>• L'analyse statistique des résultats - Suivre l'évolution des distributions de notes pour détecter les patterns récurrents et ajuster les pratiques d'évaluation.<br>• Evaluation croisée avec d'autres enseignants. |

| Biais de tendance centrale | Description |
|---|---|

| | |
|---|---|
| Perrenoud, 1998 | L'enseignant privilégie une notation moyenne par crainte de surévaluer ou de sous-évaluer. |
| **Stratégies pour éviter ces biais** | • Utiliser des grilles avec des descripteurs très précis.<br>• Eviter les grilles avec un nombre impair de degrés.<br>• Utiliser des grilles analytiques plutôt que globales. |

| Effet de sévérité et l'effet d'indulgence | Description |
|---|---|
| Merle, 1996 | Tendance à noter trop sévèrement ou trop généreusement.<br><br>L'effet de sévérité s'oppose à l'effet d'indulgence (où l'évaluateur attribue systématiquement des notes plus élevées) et s'inscrit dans ce que Merle appelle les "arrangements évaluatifs" - des pratiques de notation qui ne reflètent pas uniquement la valeur objective du travail évalué mais sont influencées par divers facteurs psychologiques et sociaux. Par exemple, l'image que veut se donner l'enseignant dans la communauté éducative ou auprès de ses élèves |
| **Stratégies pour éviter ces biais** | • Utiliser des grilles d'évaluation critériées très précises avec des indicateurs clairs pour chaque niveau de performance.<br>• Pratiquer l'évaluation en double correction dans le cadre de concours ou en interne.<br>• Organiser des sessions d'entente où les évaluateurs comparent leurs façons de noter et harmonisent leurs critères.<br>• Mettre en place des systèmes d'anonymisation des copies pour limiter les biais liés à l'identité des élèves.<br>• Former les enseignants à reconnaître leurs propres biais d'évaluation à travers l'analyse statistique de leurs notations sur le long terme.<br>• Tenir un journal de notation pour prendre conscience de ses tendances évaluatives et les réajuster. |
| Biais de légitimation | Description |
| Lefrançois, 2000 | Le biais de légitimation décrit par Lefrançois est un phénomène en docimologie où l'évaluateur tend à justifier ses |

| | |
|---|---|
| | notations ou décisions d'évaluation en s'appuyant sur des critères qui n'étaient pas nécessairement prévus initialement, afin de donner une apparence objective et légitime à son jugement.<br><br>Ce phénomène se retrouve souvent lorsque l'évaluateur prend conscience d'un effet de Hibou (Pasquini, 2017). C'est- à-dire que la correction de l'évaluation fait prendre conscience au correcteur que les choix pré-établis pour sa stratégie de correction n'étaient pas pertinents. |
| **Stratégies pour éviter ces biais** | • Tester les outils de correction sur de petites échelles avant le déploiement plus large.<br>• Prendre conscience que toute correction/évaluation a une partie d'approximation et d'erreur. |

# Synthèse des préconisations en fonction des biais

## Préconisations face aux biais docimologiques

### Biais de halo

- Anonymat
- Criteres et attendus explicites
- Évaluation « en bosse »

### Biais de contraste

- Mélange des copies
- Grilles d'évaluation
- Double correction croisée

### Biais de position

- Faire des pauses
- Tests statistiques
- Travail par compéttence

### Constante macabre

- Évaluation par contrat de confiance (EPCC)
- Suivi de l'évolution des notes
- Évaluation croisée

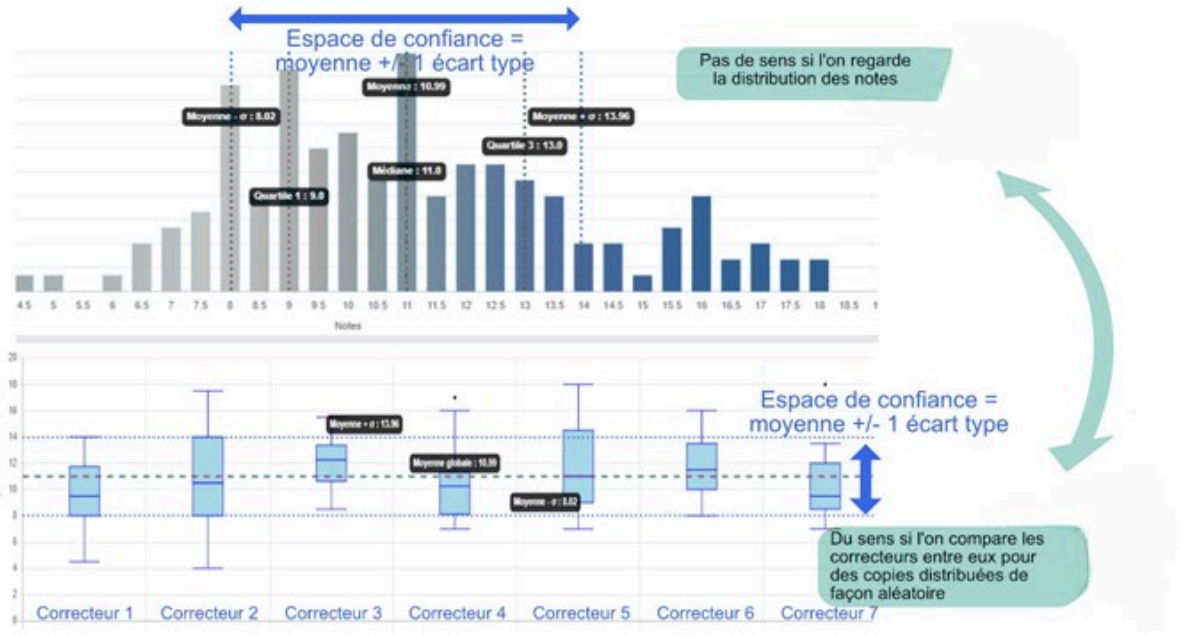
### Autres biais

- Grilles détaillées
- Correction croisée
- Formation sur les biais
- Analyse statistique des notes

### Formation des personnels de l'Éducation nationale

- Identifier, analyser et réguler ses biais

## La notion d'espace de confiance



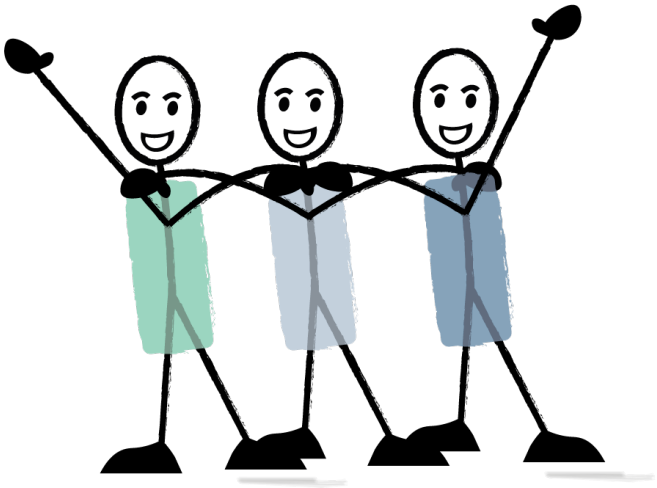

## Remarques sur la notion de critères d'évaluation précis

Il est nécessaire de ne pas confondre la précision des critères avec leur démultiplication. Afin d'obtenir une évaluation précise, une stratégie pourrait consister à multiplier les critères d'évaluation dans une démarche visant la recherche d'une grande finesse. Nous savons que notre culture scolaire répugne encore à attribuer la note maximale, encore plus de manière répétée (donc si beaucoup de critères sont évalués). Ce fractionnement des critères génère souvent une diminution des notes. Nous conseillons une approche globalisée de l'évaluation par compétences et non une vision pointilliste. Les dernières grilles du DNB, du grand Oral ou encore des épreuves anticipées de français vont dans le sens de cette globalisation. Elles sont construites avec des descripteurs précis, mais la compétence globale est évaluée au moyen d'un nombre restreint de critères.

# Références bibliographiques


- Antibi, A. (2003). La constante macabre ou comment a-t-on découragé des générations d'élèves ? Math'adore.
- Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. Journal of Educational Psychology, 111(4), 717-735.
- Batruch, A., Autin, F., & Butera, F. (2019). The paradoxical role of meritocratic selection in the perpetuation of social inequalities at school. Dans J. Jetten & K. Peters (Eds.), The Social Psychology of Inequality (pp. 141-157). Springer.
- Bonefeld, M., & Dickhäuser, O. (2018). (Biased) grading of students' performance: Students' names, performance level, and implicit attitudes. Frontiers in Psychology, 9, 481.
- Bonniol, J.-J. (1965). Les divergences de notation tenant aux effets d'ordre de la correction. Cahiers de Psychologie, 8, 181–188.
- Crahay, M. (2019). Peut-on lutter contre l'échec scolaire ? De Boeck.
- Fraisse, P., & Piéron, H. (1963). Traité de psychologie expérimentale. Presses Universitaires de France.
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and selffulfilling prophecies. Journal of Personality and Social Psychology, 93(1), 67–76. https://doi.org/10.1037/0022-3514.93.1.67
- Laugier, H., & Weinberg, D. (1927). Le facteur subjectif dans les notes d'examen. L'Année Psychologique, 28, 236–244. https://doi.org/ 10.3406/psy.1927.4042
- Lefrançois, D. (2000). L'évaluation des apprentissages. De Boeck.
- Merle, P. (1996). L'évaluation des élèves. Enquête sur le jugement professoral. Presses Universitaires de France.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. Journal of Personality and Social Psychology, 35(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250
- Pasquini, R. (2017). Évaluation : Première question vive de la pratique. L'Éducateur, 4, 38–39.
- Perrenoud, P. (1998). L'évaluation des élèves : De la fabrication de l'excellence à la régulation des apprentissages. De Boeck Université.
- Piéron, M. (1963). Examens et docimologie. Presses Universitaires de France.
- Piéron, M. (1975). La docimologie : science des examens. Presses Universitaires de France.

- Rangvid, B. S. (2015). Systematic differences in test scores across native and immigrant students: Quasi-experimental evidence. Economics of Education Review, 46, 112-126.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. Holt, Rinehart & Winston.
- Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. Journal of Experimental Psychology, 55(2), 150–155. https://doi.org/10.1037/h004878
- Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. Empirical Economics, 45(1), 523-538.
- Sprietsma, M. (2013). Discrimination in grading: Experimental evidence

**Collège des inspecteurs**

**Académie de Grenoble**

Sous la direction de Loïc Sénéjoux, IA IPR EPS et pilote du groupe Evaluation de l'académie de Grenoble.