



# Le mythe de l'évaluation juste et équitable



Dachet Dylan

Maître en Conférences et Chercheur

# Qu'est-ce que l'évaluation ?



« *Évaluer* consiste à **définir des critères et des indicateurs** dans le but de prendre des décisions, donc à choisir des éléments considérés comme pertinents dans **le référent** (ce qui devrait être, ce que l'on projette) et à déterminer durant l'action et à l'issue de celle-ci, si ces éléments sont bien **présents dans le référé** (ce qui est) » (Raynal & Rieunier, 2014, p. 215)

# Qu'est-ce que l'évaluation ?

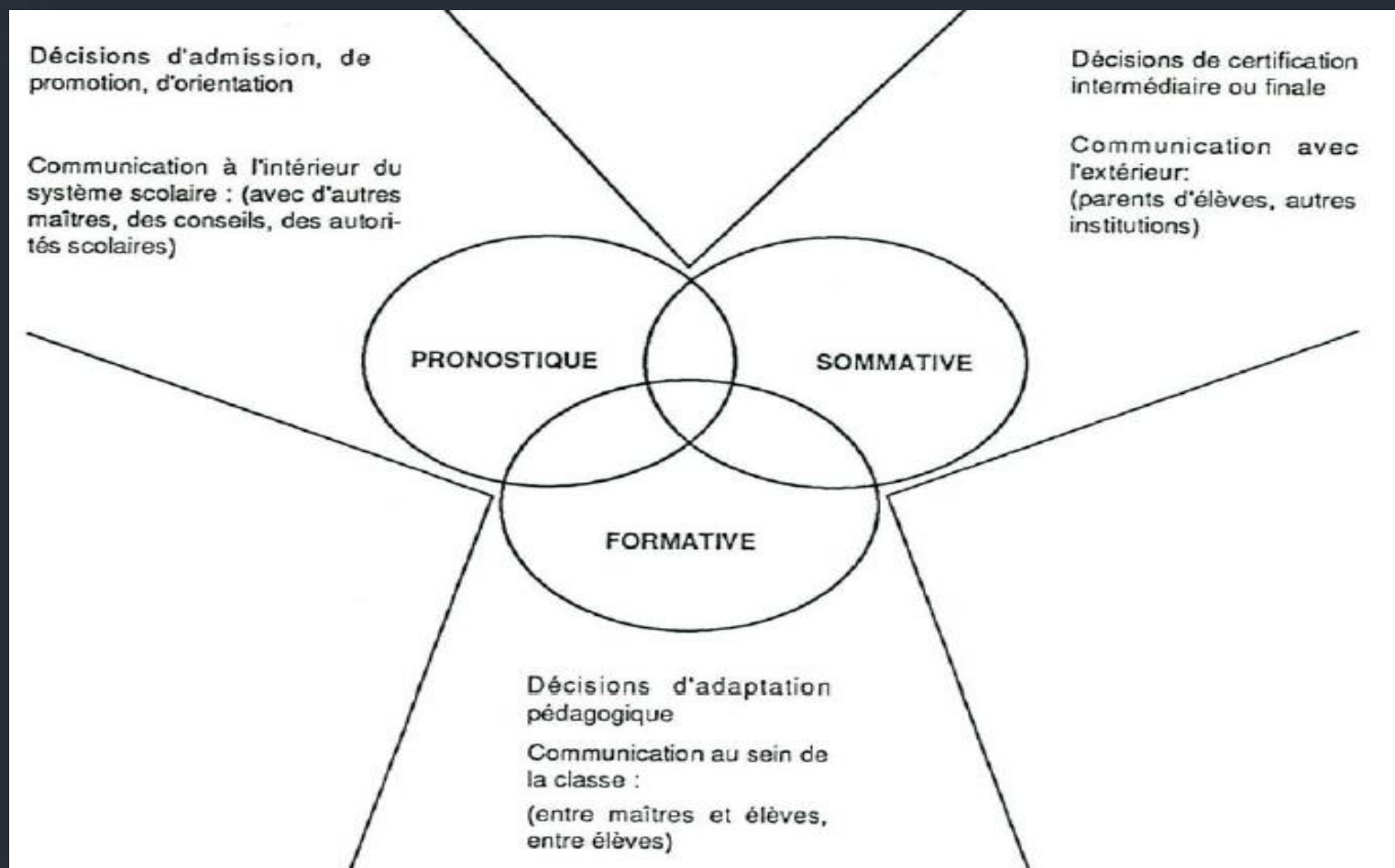
(Crahay, 2019)

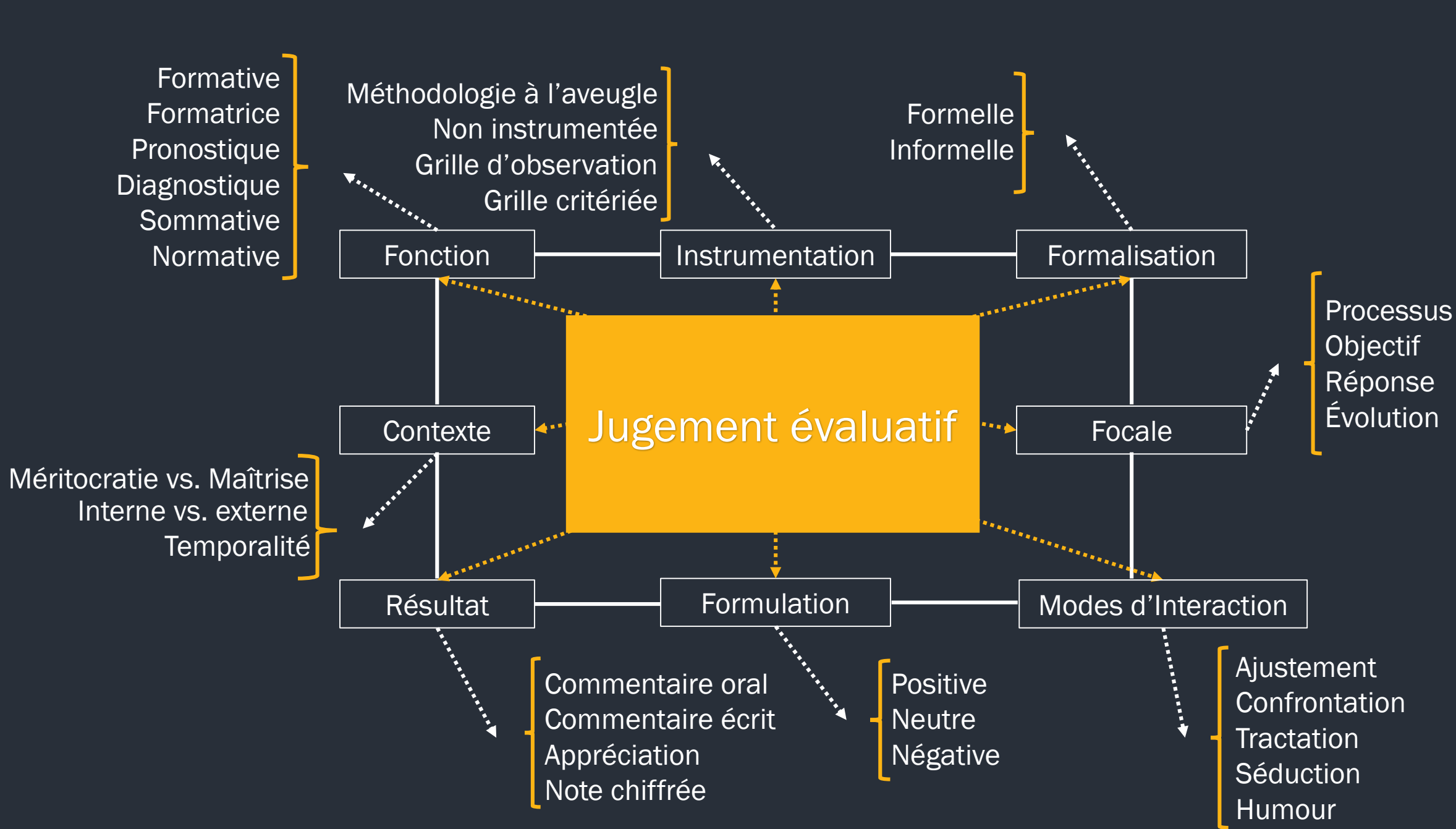


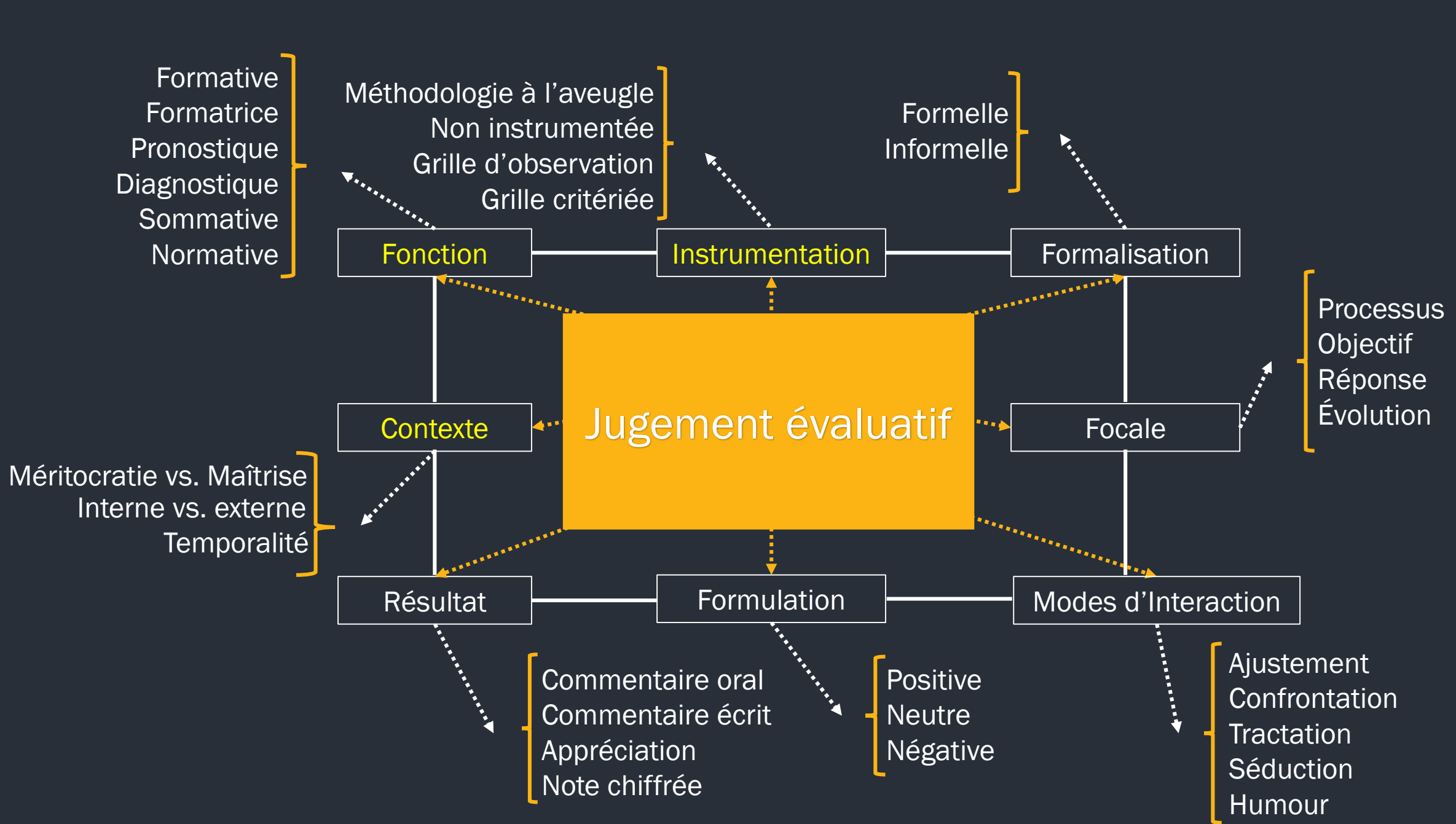
Fonction	Définition (Récolte d'informations permettant de ...)	Visée
<b>Formative</b>	Adapter les activités d' <u>enseignement</u> et d' <u>apprentissage</u> afin de <u>favoriser la progression</u> des apprenants vers les objectifs de formation	Régulation (enseignement ou apprentissage)
<b>Sommative</b> (Certificative)	Établir un <u>bilan certifiant</u> les compétences et/ou les connaissances acquises par l'apprenant à la <u>fin d'une étape</u> d'un cycle de formation.	Régulation (atteinte des objectifs)
<b>Pronostique</b>	Guide les <u>décisions d'admission</u> ou d'orientation de l'apprenant vers une nouvelle étape ou un nouveau cursus de formation.	Régulation (parcours de formation)

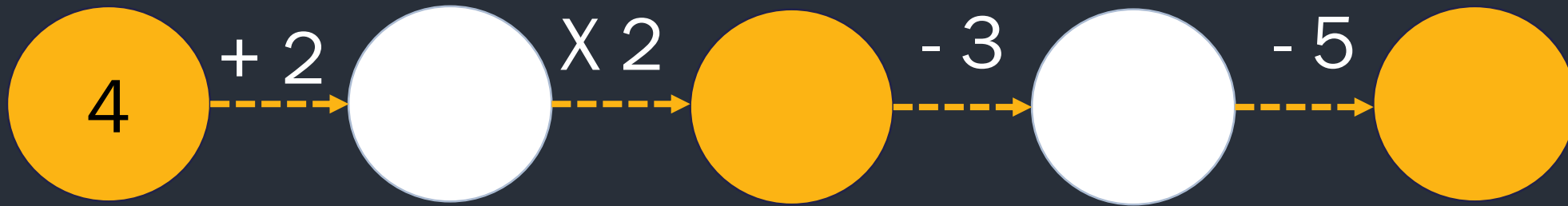
# Qu'est-ce que l'évaluation ?

(Crahay, 2019)









Réponse de Thomas



Quelle note ? /2

# Le concept de biais



Monde idéal

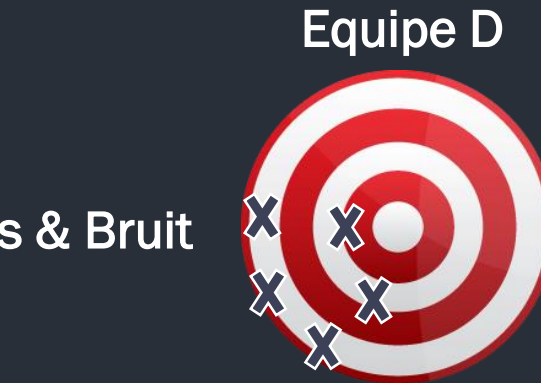


Biais

Le biais et le bruit – l'écart  
systématique et la  
dispersion aléatoire – sont  
deux composantes  
différentes de l'erreur



Bruit



Biais & Bruit

(Kahneman et al., 2021, p. 9-11)

Que sait-on des biais qui impactent le jugement évaluatif ?

# Le concept de biais

Les biais cognitifs = les **erreurs systématiques** que notre mental a tendance à commettre lorsqu'il réfléchit (Thalemann, 2021, p. 25)

Il s'agit d'un schéma de pensée trompeur et faussement logique  
(Usabilis, 2018)

→ Altère le jugement

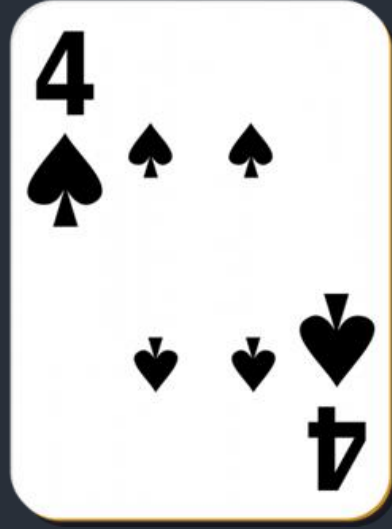
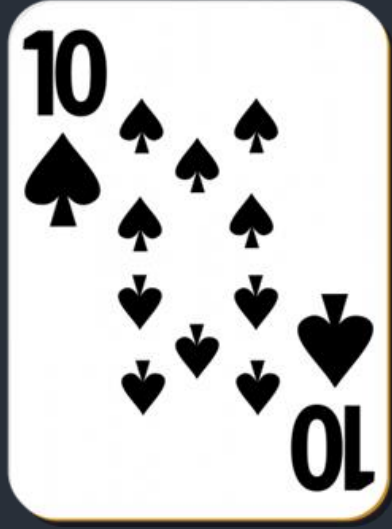
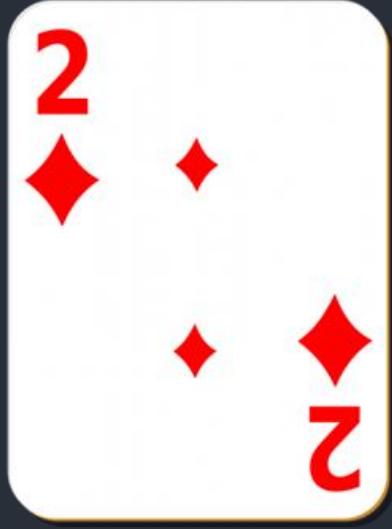
→ Facilite/accélère la prise de décision

Deux vitesses de la pensée (Kahneman, 2012) :

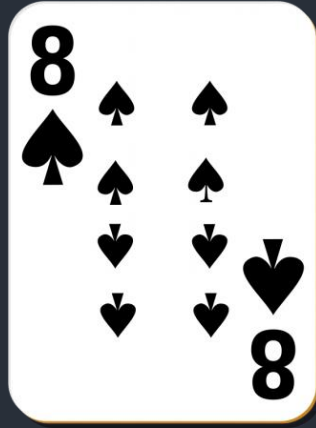
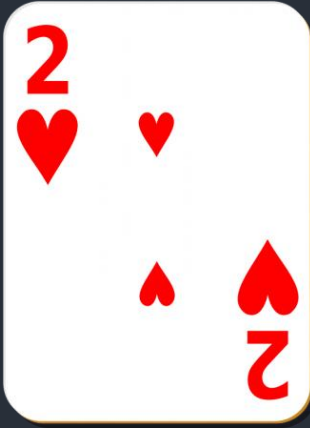
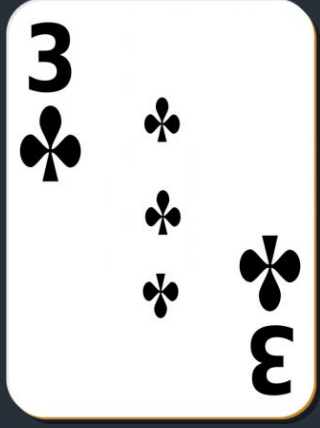
(1) Système 1 : automatique et incontrôlable

(2) Système 2 : calculateur









# Le biais de confirmation

= tendance à privilégier les informations et/ou exemples qui confirment nos croyances au détriment de celles qui pourraient les remettre en question (Durand, 2019, 121-135 ; Thalemann, 2021, p. 25-26)

## Exemples :

- (1) Syndrome de l'étudiant en médecine
- (2) Étude de Lord et al. (1979) sur la peine de mort
- (3) Étude de Sunstein et al. (2017) sur le réchauffement climatique
- (4) ... Avortement, homosexualité, ...

Explication : psychologie évolutionniste (Boutang & De Lara, 2009)

→ Intérêt adaptatif (Thalemann, 2021)



Source : [www.comicscience.net](http://www.comicscience.net)

# Les biais de l'évaluation

**Effet de halo** : l'évaluation est influencée par des caractéristiques de l'élève (présentation, attitude en classe...) et/ou de son travail (soin, orthographe, élégance du style...) même si on précise que ces éléments ne doivent pas être pris en compte.

**Effet de contagion (ou contamination)** : les notes attribuées successivement aux différentes questions s'influencent mutuellement. Un élève qui aura très bien répondu aux quelques premières questions d'un test verra son résultat surestimé pour une question suivante qui n'est que moyennement (ou pas du tout) réussie, contrairement à l'élève qui aurait mal commencé le test.

**Effet d'ordre ou de contraste** : l'évaluation est influencée par les copies antérieures. Une copie d'élève qui vient dans l'ordre après une copie très brillante peut être sous-estimée et vice versa.

# Les designs de recherche



La recherche corrélacionnelle

(cité par Rot et Butas, 1959)

# L'effet *Posthumus* et l'expérience fameuse de Gjorgjevski

## Étape 1

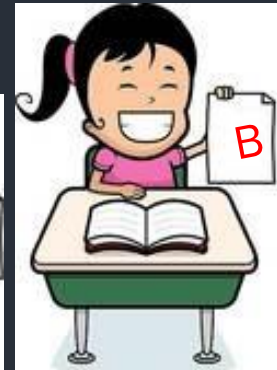
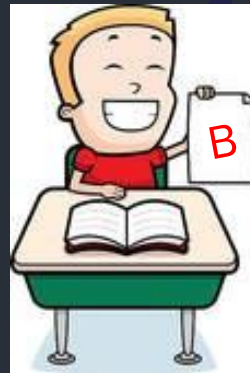
5 correcteurs



## Étape 2

15 copies notées « Bien »

4 autres correcteurs

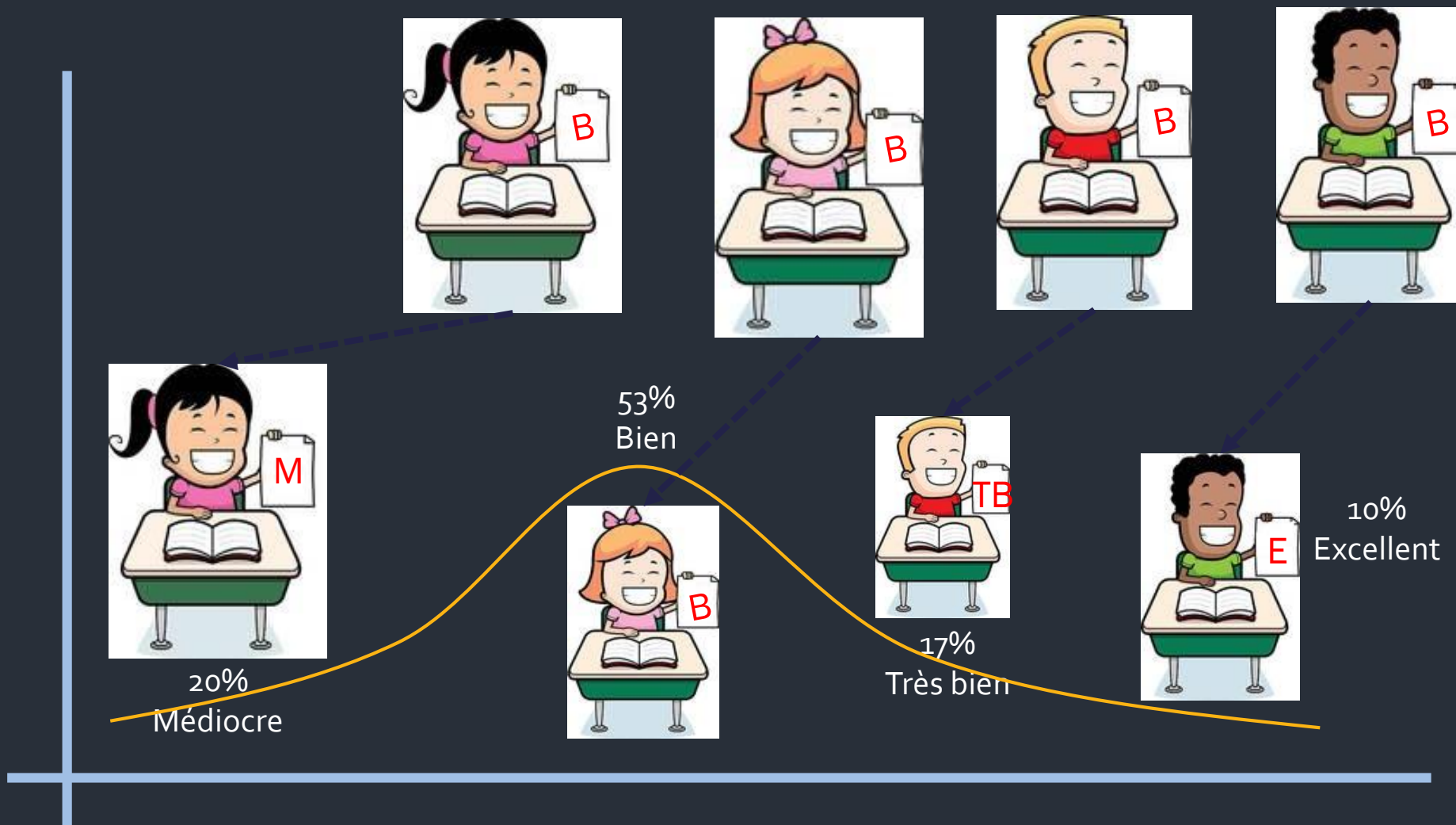


100 épreuves

(cité par Rot et Butas, 1959)

# L'effet *Posthumus* et l'expérience fameuse de Gjorgjevski

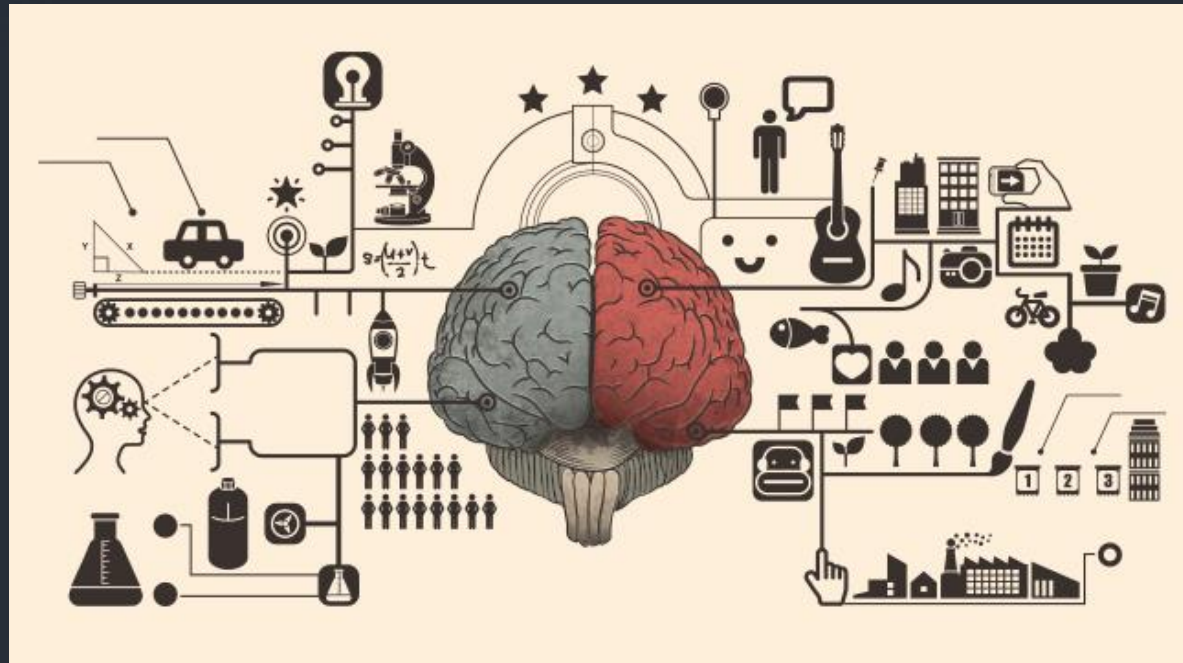
## Résultats



(cité par Rot et Butas, 1959)

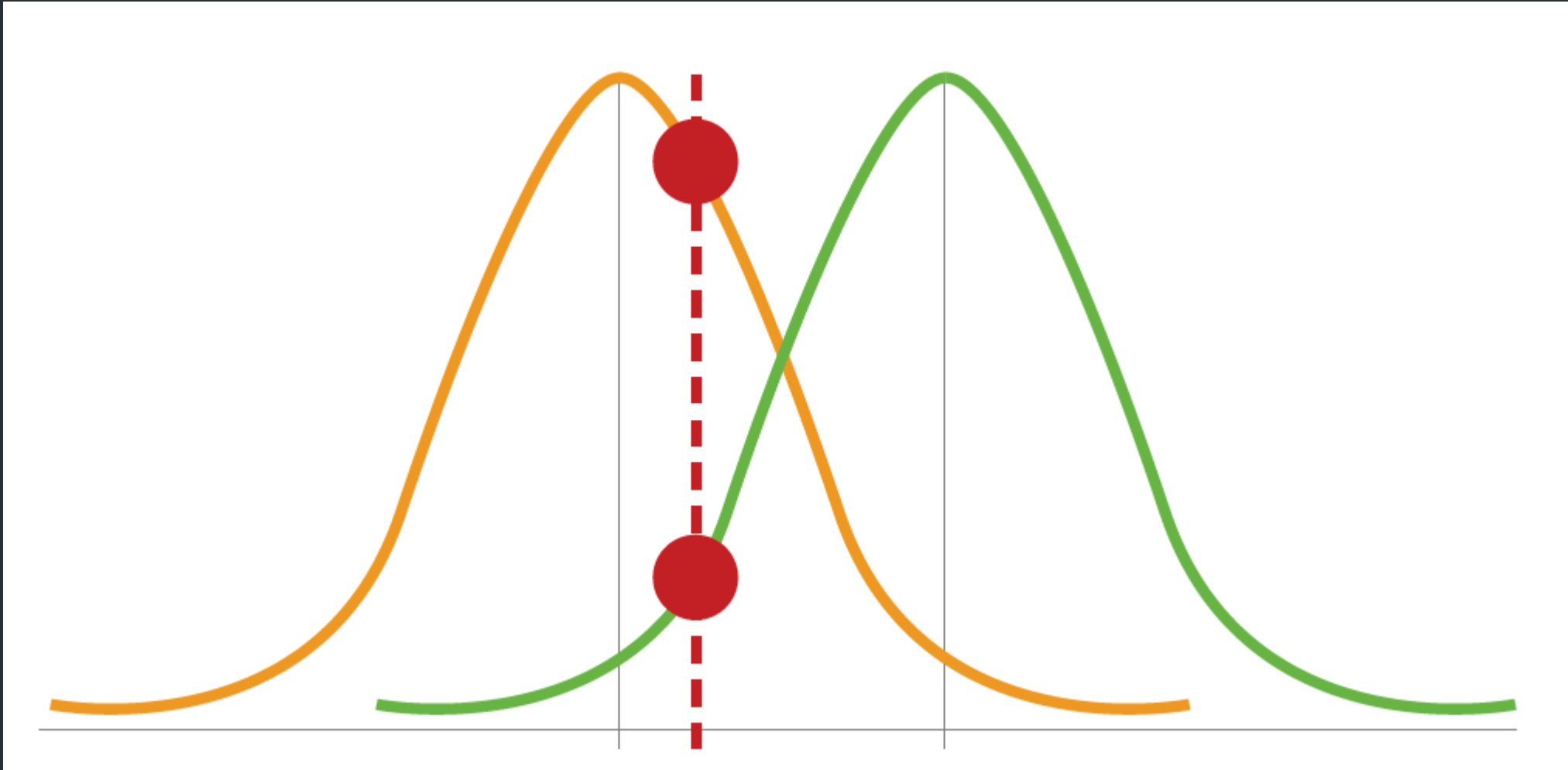
# L'effet *Posthumus* et l'expérience fameuse de Gjorgjevski

« *L'enseignant adapte le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver d'année en année, approximativement la même distribution (gaussienne) de notes* » (De Landsheere, 1980)



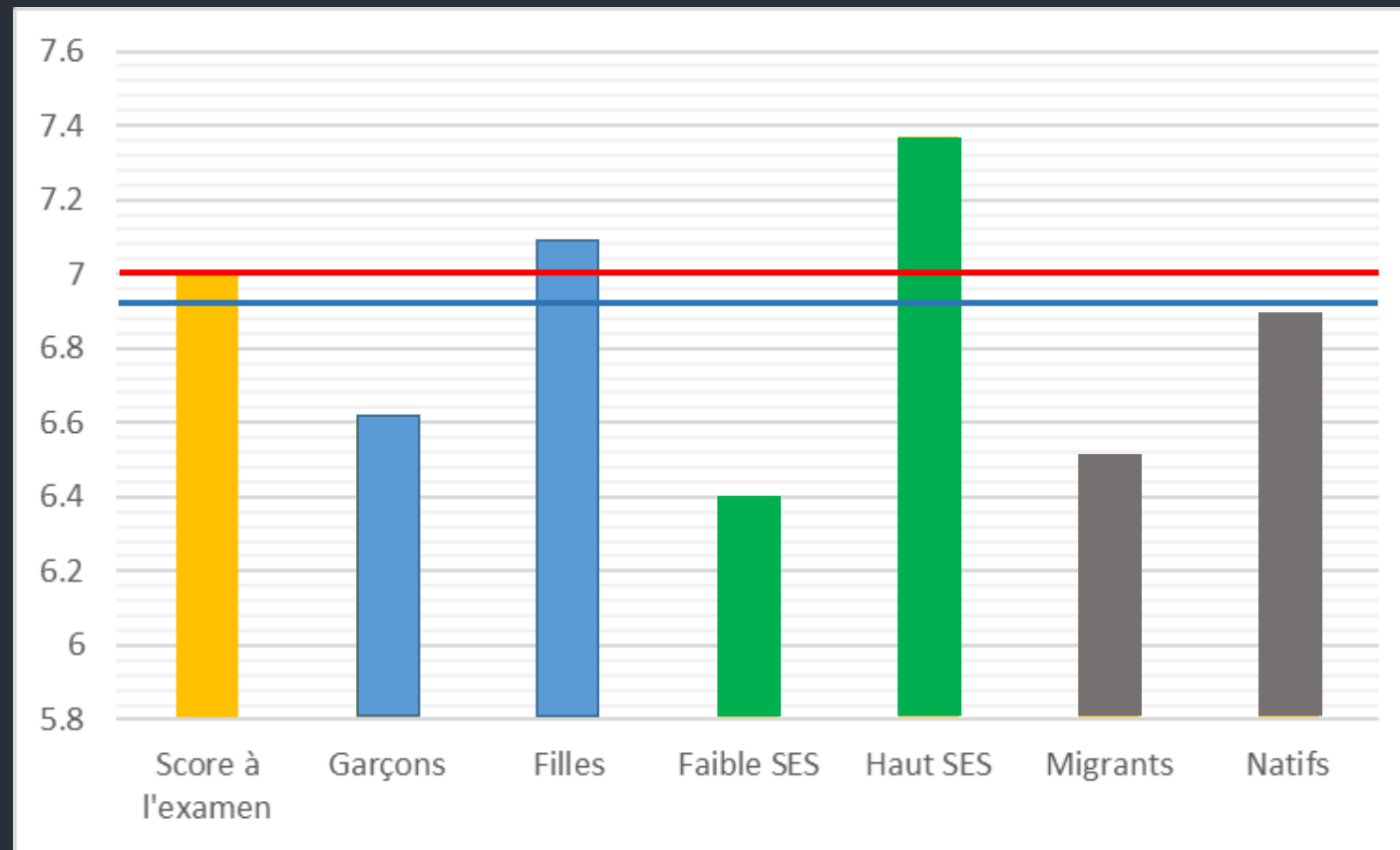
(cité par Rot et Butas, 1959)

# L'effet *Posthumus* et l'expérience fameuse de Gjorgjevski



# Effet de halo lié aux caractéristiques des élèves ?

Comparaison entre : (1) les données issues des examens étatiques externes et (2) les notes des enseignants



*Pour les élèves qui ont obtenu un score identique de 7 au test étatique, les notes des enseignants étaient plus hétérogènes : elles variaient de 6,40 pour les élèves dont les parents ont un faible niveau d'éducation à 7,37 pour les élèves dont les parents ont un haut niveau d'éducation.*

# Les designs de recherche

**Activation du stéréotype**



**11 fautes claires  
6 fautes ambiguës**



**11 fautes claires  
6 fautes ambiguës**

**La recherche expérimentale**

**Non activation du stéréotype**



**11 fautes claires  
6 fautes ambiguës**



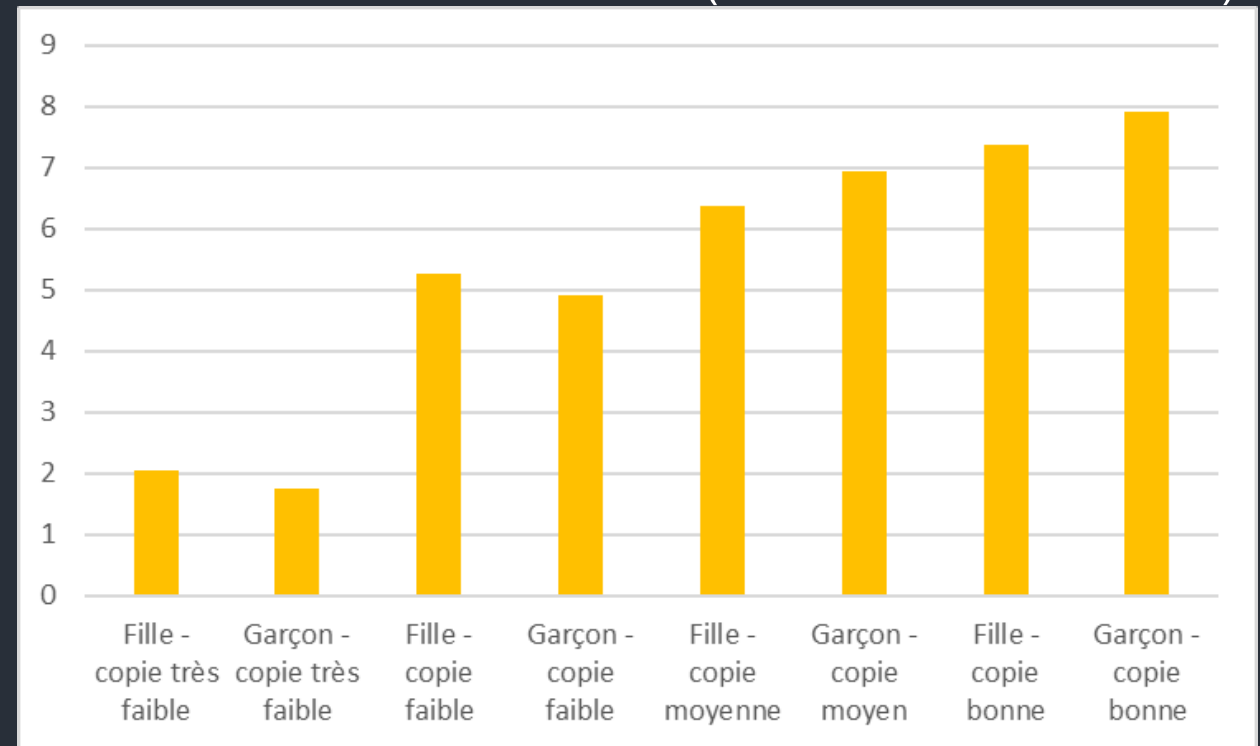
(Lafontaine et Monseur, 2009)

# Effet de halo lié aux caractéristiques des élèves ?

Etude expérimentale :

- (1) Copies fictives construites par un professeur de mathématiques : copie très faible, une copie assez faible, une copie moyenne et une bonne copie
- (2) Seule information disponible → prénom clairement masculin (Bernard et Nicolas) ou féminin (Chloé et Emilie)

48 professeurs de mathématiques

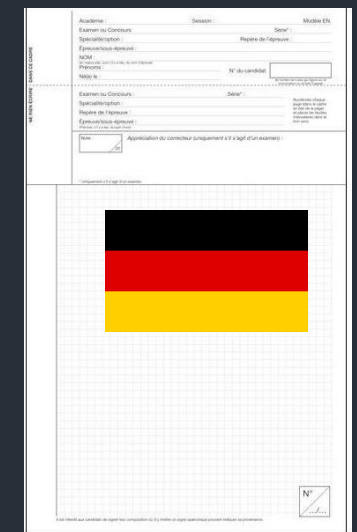
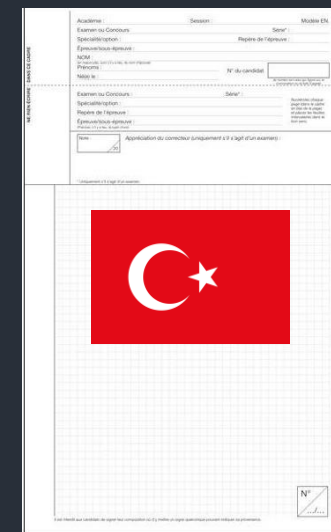


# Effet de halo lié aux caractéristiques des élèves ?

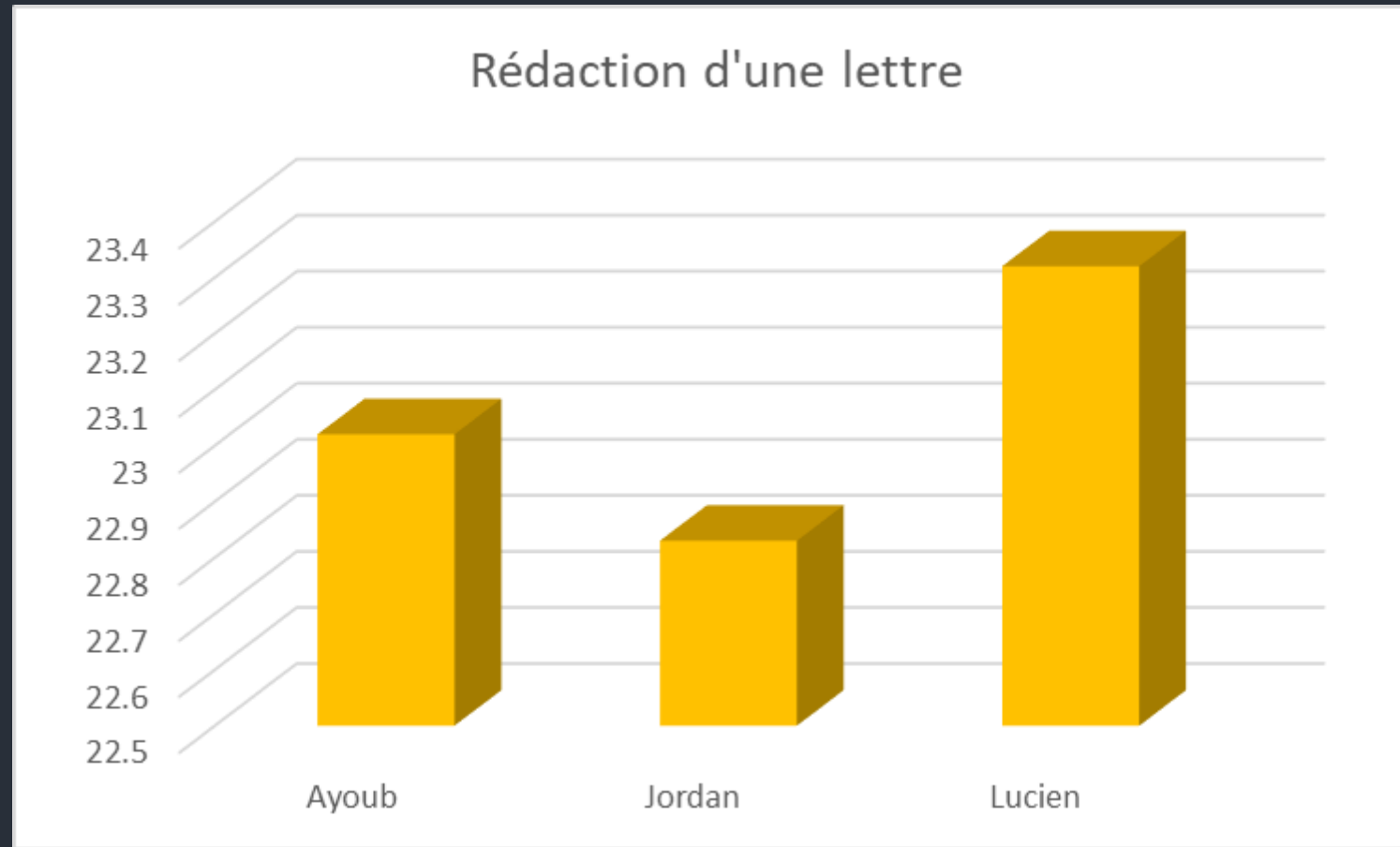
Etude expérimentale :

- (1) Recueil de dissertations rédigées par deux classes anonymes
- (2) Sélection aléatoire de 10 dissertations
- (3) Envoie de la série chaque enseignant avec assignation aléatoire d'un prénom à chaque copie
  - Des prénoms à consonance germanophone (Max, Stefan, ...)
  - Des prénoms à consonance turque (Hakan, Gönül, ...)
  - (une des communautés ethniques les plus représentées)

88 enseignants du primaire (58 écoles différentes)



# Effet de halo lié aux caractéristiques des élèves ?



# Effet des paramètres de l'évaluation



# Effet des paramètres de l'évaluation

	Évaluation normative	Évaluation formative
Définition	utilise des indicateurs tels que des côtes, lettres ou jugements de valeur qui servent parfaitement l'objectif de comparaison avec une norme et entre individus.	fournit un feedback (ici qualitatif) spécifique et détaillé
Visée	comparer les performances de la personne évaluée à celles d'autres personnes	adapter les activités d'enseignement et d'apprentissage et fournir des commentaires pertinents
Contexte	Méritocratique	Correctif/de Maîtrise
Effet	Notation d'un test attribué à un SES - ou SES + Étudiants inscrits dans un programme sélectif $SSE + > SSE -$ (Batruch, Autin & Butera, 2017)	Notation d'un test attribué à un SES - ou SES + Étudiants inscrits dans un programme non-sélectif $SSE + = SSE -$ (Batruch, Autin & Butera, 2017)

# Effet des paramètres de l'évaluation

**Dylan**



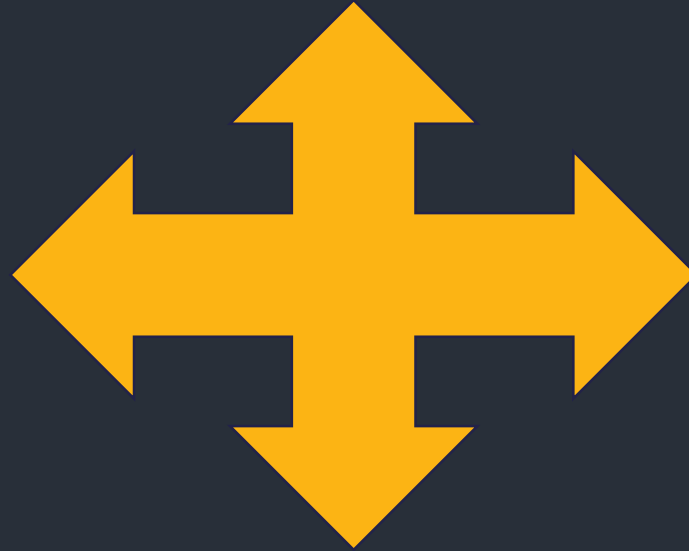
**Dylan**



**Charles**



**Charles**



# Effet des paramètres de l'évaluation

## Expérience 1

### Etude expérimentale:

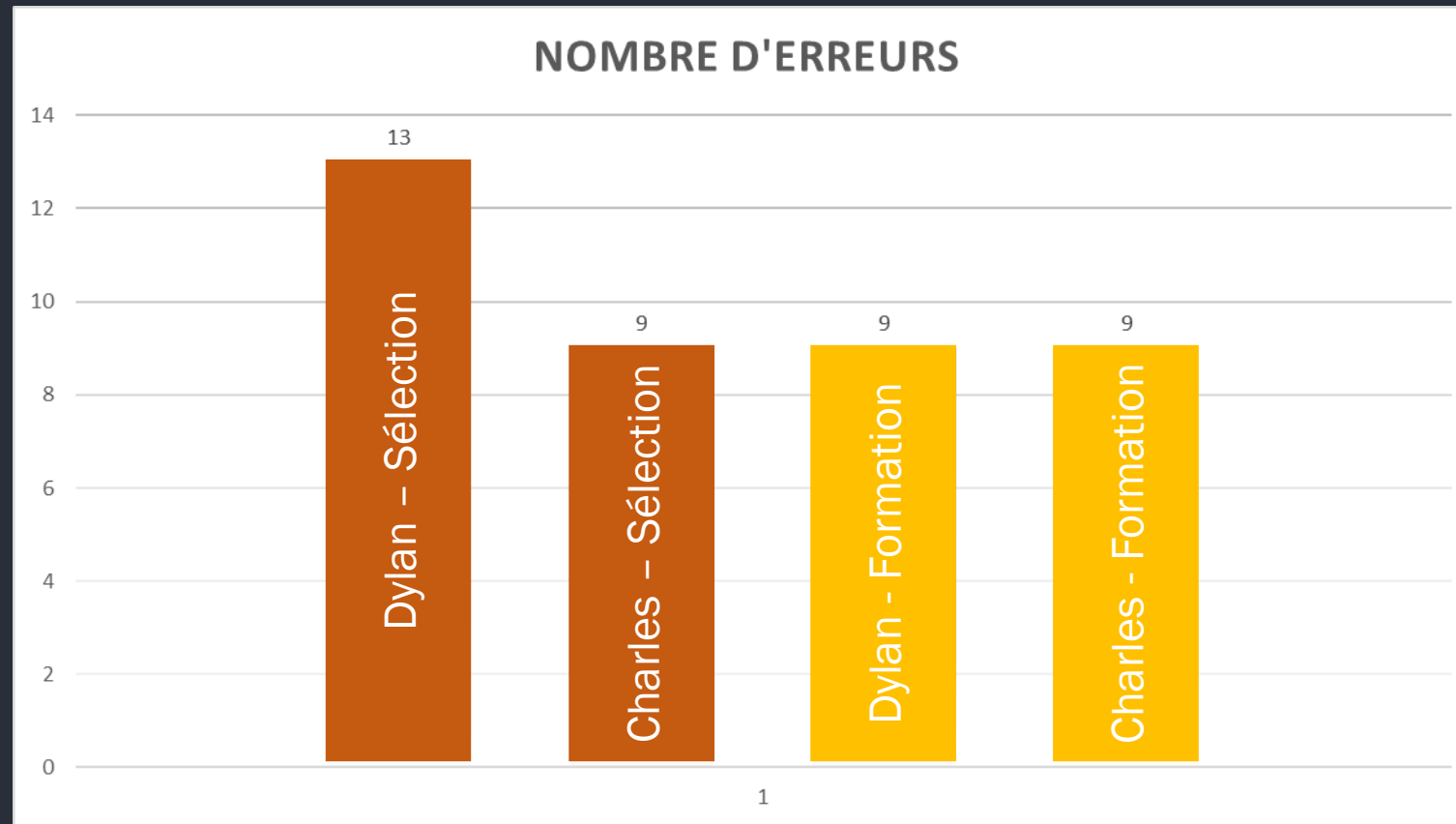
- Double manipulation :
  - Note (normatif) vs. commentaire(s) (formatif)
  - SES faible vs. fort (via des indicateurs tels que le prénom, le statut professionnel des parents, activités extracurriculaires, ...)
- N = 196 étudiants (imaginant qu'ils sont des enseignants)
- Correction d'une dictée (1-6) contenant 11 fautes claires et 6 fautes ambiguës

### Variables étudiées :

- Temps de correction → plus grand dans le groupe « commentaire » que dans le groupe « Note » Nombre de fautes repérées :
  - Note (M = 10.16, SD = 2.53) > Commentaire(s) (M = 8.61, SD = 2.18).
  - SES faible (M = 9.90, SD = 2.54) > SES fort (M = 8.87, SD = 2.16)
- Effet positif du temps sur le nombre de fautes repérées (b = 0.17).

# Effet des paramètres de l'évaluation

## Expérience 1



# Effet des paramètres de l'évaluation

Estimation de l'ampleur de l'effet de la modulation de l'écart de performance dû au SSE de l'élève en fonction de l'orientation de l'évaluation (sélective vs. éducative):

$$D = 0.19$$

$$p = 0.002$$

« Cette méta-analyse interne prouve que **les évaluateurs créent artificiellement un plus grand écart de performance SSE lorsque l'évaluation est utilisée pour sélectionner plutôt que pour favoriser l'apprentissage**. L'ampleur de l'effet est faible, mais nous pensons néanmoins qu'elle doit être interprétée à la lumière de la durée de l'éducation et de la fréquence de l'évaluation. De très petites différences dans les évaluations répétées peuvent avoir des conséquences importantes sur les expériences globales et les résultats scolaires des étudiants lorsqu'elles s'accumulent au fil du temps » (Autun, Batruch & Butera, 2019, p. 729).

# Effet des paramètres de l'évaluation

La manière de poser une question influence les réponses obtenues :

1. A quel point avez-vous réussi dans la vie ?

0	1	2	3	4	5	6	7	8	9	10
-5	-4	-3	-2	-1	0	1	2	3	4	5



2. A quelle fréquence apparaissent les symptômes

<b>Deux fois par mois ou moins</b>	<b>Plus de deux fois par mois</b>	<b>Plus de deux fois par semaine</b>	<b>Plusieurs fois par jour</b>
Jamais	Une fois par an	Une fois par mois	<b>Plus de deux fois par mois</b>

3. Quelle est la chose la plus importante à enseigner aux enfants ?

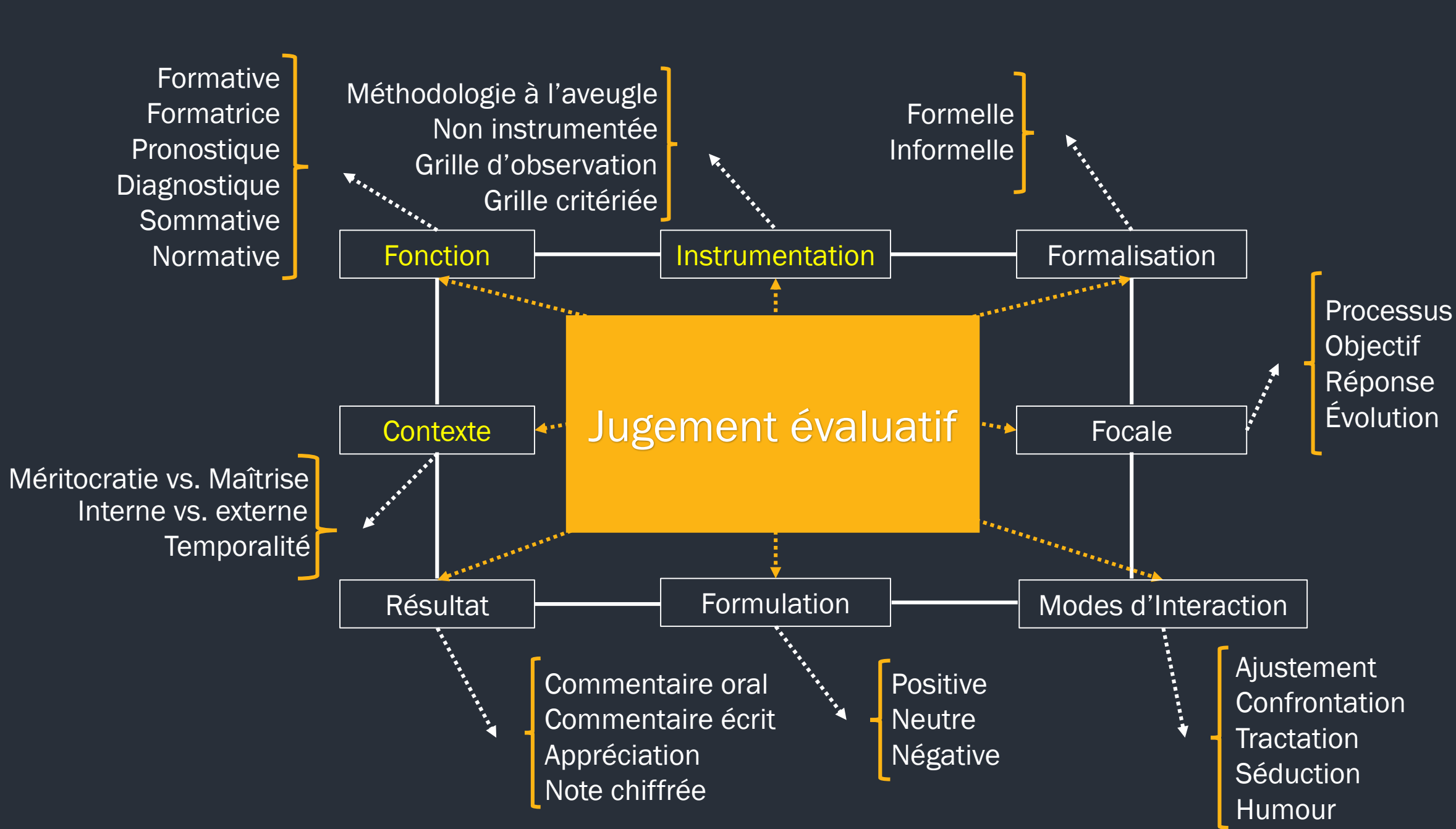
- 61 % cochent la réponse : penser par eux-mêmes
- 5 % le mentionnent spontanément lors d'une question ouverte.

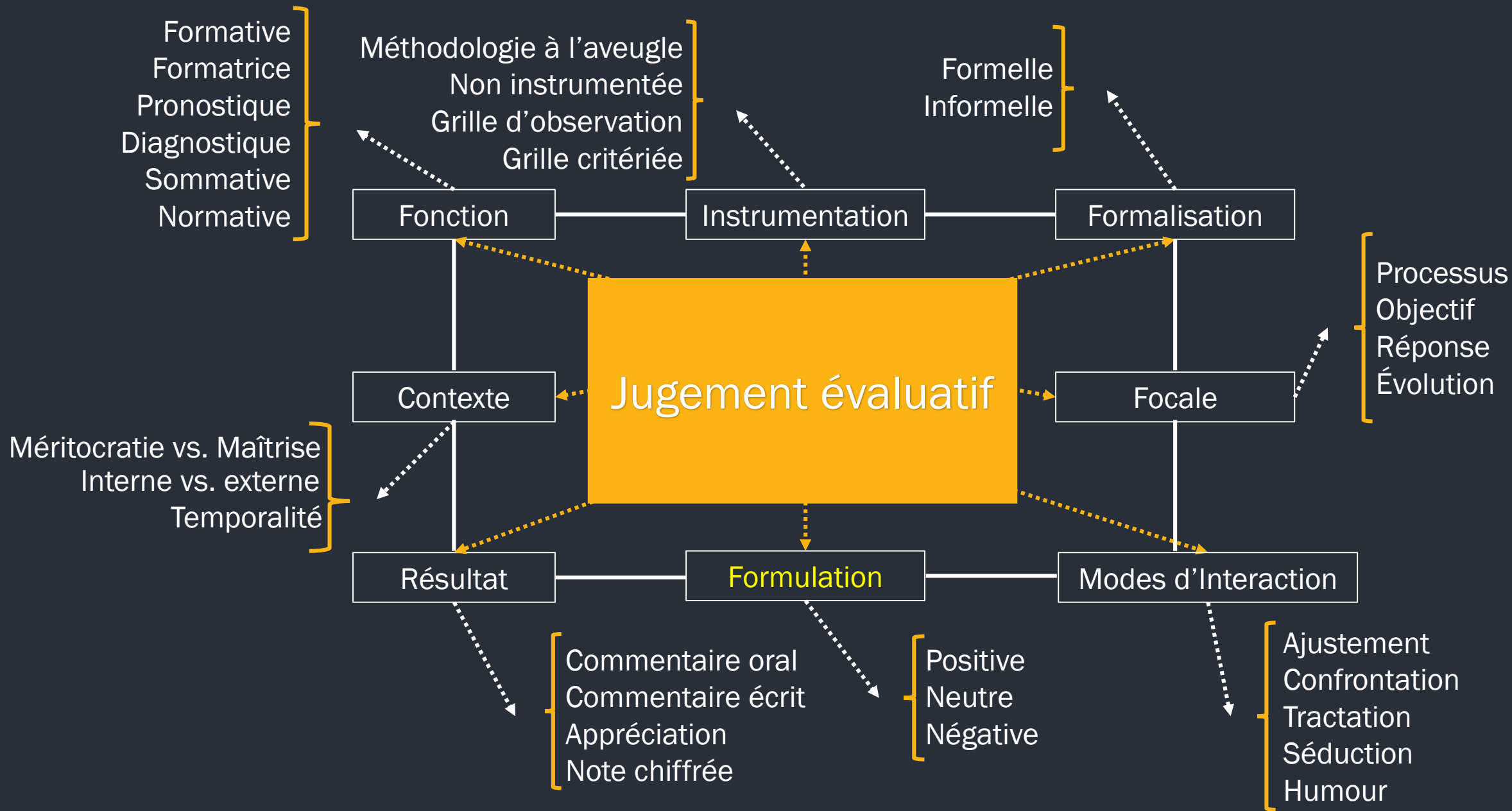
(Crahay, 2019)

L'évaluation des performances des élèves/étudiants par des enseignants peut-elle être biaisée par leur connaissance du milieu social d'un élève ?



- Origine ethnique de l'élève (Dutrévis, 2015 ; Sprietsma, 2013) – et son interaction avec le niveau de la copie (Dachet et al., 2023)
- Niveau d'éducation des parents (Rangvid, 2015)
- Genre – et son interaction avec le niveau de la copie dans les disciplines stéréotypées (Lafontaine & Monseur, 2009)
- Parcours scolaire – redoublement (Rauw, 2022) et génération Covid (Dachet et al., 2022) de l'élève
- Statut socio-économique et son interaction avec le contexte éducatif (méritocratique vs. de maîtrise) (Autin et al., 2019)





# Le renforcement positif, une pratique qui a fait ses preuves ?

Revue systématique (Royer et al., 2018)

QR : *Est-ce que la pratique du renforcement positif spécifique au comportement peut-être considérée comme fondée sur des preuves ?*

- L'éloge = **subordonnée** au comportement, **spécifique**, **sincère**, **variée**, et **crédible**  
= **facteur malléable** plutôt facteur non contrôlable  
= manière dont le **comportement rencontre l'attente** ou **affecte la réussite**

**Méthode** : recherche dans 21 bases de données bibliographiques → 318 références  
recherche à la main → 32 références

- 282 articles exclus pour des raisons de pertinence
- 29 articles exclus après lecture des synthèses
- 33 articles exclus après lecture complète
- 6 articles rencontrent les critères d'inclusion (enseignants ; comportements ; K-12 ; recherche expérimentale)

# Le renforcement positif, une pratique qui a fait ses preuves ?

Revue systématique (Royer *et al.*, 2018)

Étude	Mesure	Effet
Madsen <i>et al.</i> (1968)	Comportements inappropriés	+
Van der Mars (1989)	Comportements hors de la tâche	+
Houghton <i>et al.</i> (1990)		+
Wheatley <i>et al.</i> (2009)	Comportements inappropriés	+
Haydon & Musti-Rao (2011)	Perturbations	+
Hollingshead <i>et al.</i> (2016)	Comportements dans la tâche	/

# Le renforcement positif, une pratique qui a fait ses preuves ?

Revue systématique (Royer *et al.*, 2018)

## Conclusion :

« Les six études examinées montrent que le renforcement positif peut être mis en œuvre avec un **minimum d'effort** et pratiquement **aucune perturbation des enseignements**, alors qu'il **entraîne des changements rapides dans les résultats des élèves**. D'autres études plus rigoureuses sur le plan méthodologiques sont nécessaires pour déterminer si le renforcement positif comme utilisé dans les écoles traditionnelles de la maternelle à la fin de l'enseignement secondaire supérieur peuvent être considérés comme des pratiques *evidence-based*. »

# Le renforcement positif, une pratique qui a fait ses preuves ?

Enquête sur les adolescents (Fefer et al., 2016)

**QR :** *Quelles sont les préférences des élèves adolescents en matière de renforcements positifs et de récompenses relatifs au comportement scolaire ?*

**Méthode :** enquête administrée à 764 élèves de la 5<sup>e</sup> année (P5) à la 12<sup>e</sup> (S6) année d'un district scolaire américain

## Résultats :

- Les adolescents **apprécient** tant les renforcements positifs que les récompenses
- Les adolescents accordent, souvent, **plus d'importance à l'opinion des enseignants** concernant le travail scolaire qu'à l'opinion d'autres (pairs, parents, ...)
- Les **renforcements positifs « silencieux »** (ou privés/discrets) sont plus appréciés que les renforcements positifs publics.
  - Ce dernier constat semble surtout vrai pour les adolescents, moins pour les plus jeunes (Burnett, 2001)

# Une typologie des renforcements positifs

(Morin, 2022)

- Le renforcement positif **personnel**  
Met l'accent sur les **talents naturels**, les **compétences individuelles**, ...  
*« Tu es vraiment doué en algèbre ! »*
- Le renforcement positif fondé sur l'**effort**  
Met l'accent sur le processus d'apprentissage, le temps investi, ... et donc sur des **éléments contrôlables**  
*« Je suis impressionné par le temps que tu as investi pour maîtriser les équations du premier degré. »*
- Le renforcement positif spécifique à un **comportement**  
Met l'accent sur la description du **comportement attendu observé**  
*« Je te félicite d'être resté silencieux pendant que les autres travaillaient ».*

# Qu'est-ce qu'un bon renforcement positif ?

Caractéristiques	Définition	Contre-exemple
Sincère et véritable	Repose sur une forme de vérité // éviter les exagérations (Brummelman et al., 2014)	« Waw, c'est exceptionnel »
Subordonné au comportement	Dépend intimement du comportement observé	« Tu te comportes mieux depuis quelque jours »
Spécifique	Décrit exactement le comportement observé et son intérêt	« J'apprécie que tu restes calme »
Pas directif	Ne décrit pas des tâches futures à accomplir	« C'est bien, tu peux maintenant ... »
Centré sur le processus	Renforce la démarche, le progrès, ... plus que le résultat et/ou la comparaison avec les autres	« Tu as obtenu le meilleur résultat et, pourtant, tu as fini en premier »
Immédiat	Est formulé immédiatement après l'observation du comportement	« La semaine passée, ton comportement dans le rang était plus respectueux »
Cohérent avec les attentes	N'est utilisé que pour valoriser un comportement observé qui correspond au comportement attendu et explicité	
Inattendu	N'est pas prévisible	« C'est bien »
Varié	Porte sur l'ensemble des comportements attendus	

# Conclusion : Quelles évolutions de l'évaluation formative

Conception initiale de Bloom versus conception élargie de l'évaluation formative dans la littérature scientifique francophone (Allal & Mottier Lopez, 2005)

Conception initiale de Bloom	Conception élargie
Insertion de l'EF <u>après</u> la phase d'enseignement	Intégration de l'EF durant <u>tout</u> l'apprentissage
Utilisation de <u>tests</u> formatifs	Utilisation de <u>divers moyens</u> de recueil d'informations
Feed-Back + correction → <u>remédiation</u>	Feed-back + adaptation de l'enseignement → <u>régulation</u>
Gestion de l'EF par l' <u>enseignant</u>	Participation active des <u>élèves</u> à l'EF
Maîtrise des objectifs par <u>tous les élèves</u>	<u>Différenciation</u> de l'enseignement et des objectifs
<u>Remédiation bénéfique</u> aux élèves qui ont été évalués	<u>Régulation à 2 niveaux</u> : pour les élèves évalués et pour les futurs élèves

# Conclusion : changer l'évaluation pour changer l'éducation

Évaluation dans une perspective ...	
D'excellence	De maîtrise
<p>Évaluation <b>normative</b></p> <ul style="list-style-type: none"><li>→ Classement des individus les uns/aux autres</li><li>→ Distribution gaussienne des notes</li><li>→ Questions discriminantes</li></ul>	<p>Évaluation <b>sommative</b> à référence <b>critériée</b></p> <ul style="list-style-type: none"><li>→ Permettre de situer chaque élève par rapport aux compétences assignées comme objectifs</li></ul>
<p>Évaluation <b>pronostique</b></p> <ul style="list-style-type: none"><li>→ Apte à suivre ?</li><li>→ Sélection.</li></ul>	<p>Évaluation <b>formative</b></p> <ul style="list-style-type: none"><li>→ Comprendre les difficultés éprouvées par l'élève<ul style="list-style-type: none"><li>• Diagnostic</li><li>• Analyse des réponses</li><li>• Interactions avec les élèves</li></ul></li><li>→ Réorienter l'apprentissage et/ou l'enseignement</li></ul>

# Bibliographie

Allal, L. et Mottier Lopez, L. (2005). Formative assessment of learning : A review of publications in French. In *Formative Assessment – Improving Learning in Secondary Classrooms* (pp. 241-264). Paris, France : OECD-CERI Publication (What works in innovation in education).

Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assesment leads evaluators to artificially create te social class achievement gap. *Journal of Educational Psychology*, 111, 717-735. <https://doi.org/10.1037/edu0000307>

Burnette, P. C. (2001). Elementary students' preferences for teacher praise. *The Journal of Classroom Interaction*, 36(1), 16-23. Retrieved from: <https://www.jstor.org/stable/23870540>

Crahay, M. (2019). *Peut-on lutter contre l'échec scolaire ?* (4th ed.). de Boeck supérieur.

Durand, T. C. (2019). *Quand est-ce qu'on biaise ?*. Paris, France : HumenSciences (société)

Durand, T. C. (2021). *La science des balivernes*. Paris, France : HumenSciences (société)

Dutrévis, M. (2015). Menace du stéréotype et groupes ethno-raciaux minoritaires. Quel poids sur les performancesdes élèves ? *Revue française de pédagogie*, 191, 61-72. Retrieved from : <http://rfp.revues.org/4751>

Fefer, S., DeMagistris, J., & Shuttleton, C. (2016). Assessing adolescent praise and reward preferences for academic behavior. *Translational Issues in Psychologie Science*, 2(2); 153-162. <https://doi.org/10.1037/tps0000072>

# Bibliographie

Kahneman, D. (2012). *Système 1, système 2 : les deux vitesses de la pensée*. Flammarion

Kahneman, D. Sibony, O., & Sunstein, C. R. (2021). *Noise*. Odile Jacob

Lafontaine, D., & Monseur, C. (2009). Les évaluations des performances en mathématiques sont-elles influencées par le sexe de l'élève ? *Mesure et Evaluation en Education*, 32 (2), 1-106.  
<https://doi.org/10.7202/1024955ar>

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-2109. <https://doi.org/10.1037/0022-3514.37.11.2098>

Morin, A. (2022). *The power of effective praise: a guide for teachers*. Retrieved from:  
<https://www.understood.org/articles/en/the-power-of-effective-praise-a-guide-for-teachers>

Rangvid, B. S. (2015). Systematic differences across evaluation schemes and educational choice. *Economics of Education Review*, 48, 41-55. <http://dx.doi.org/10.1016/j.econedurev.2015.05.003>

Raynal, F., & Rieunier, A. (2014). *Pédagogie, dictionnaire des concepts clés : Apprentissage, formation, psychologie cognitive* (10th ed.). ESF éditeur

Rot, N., Butas, Z. (1959). Les distributions des notes scolaires comparées aux distributions des résultats obtenus aux tests de connaissances. *Le travail humain*, XXII, 1-2.

# Bibliographie

Royer, D. J., Lane, K. L., Dunlap, J. D., & Parks Ennis, R. (2018). A systematic review of teacher-delivered behavior-specific praise on k-12 student performance. *Remedial and Special Education, 40*(2), <http://dx.doi.org/10.1177/0741932517751054>

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics, 45*, 523–538. <http://dx.doi.org/10.1007/s00181-012-0609-x>

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2017). How People Update Beliefs about Climate Change: Good News and Bad News. *Cornell Law Review, 102*(6), 1431-1443. Retrieved from: <https://scholarship.law.cornell.edu/clr/vol102/iss6/1/>

Thalman, Y.-A. (2021). *La tentation du paranormal*. HumenSciences (Débat)

Usabilis. (2018, may 29). *Définition biais cognitive*. Retrieved from: <https://www.usabilis.com/definition-biais-cognitifs/>