

CONFÉRENCE DE CONSENSUS

L'ÉVALUATION EN CLASSE, AU
SERVICE DE L'APPRENTISSAGE
DES ÉLÈVES



RAPPORT DE SYNTHÈSE

#CC_EVALUATION

LES 23 ET 24 NOVEMBRE 2022

En partenariat avec :



**L'ÉVALUATION EN CLASSE AU SERVICE
DE L'APPRENTISSAGE DES ÉLÈVES**
Rapport de synthèse

LAFONTAINE Dominique
Université de Liège (Belgique)

TOCZEK-CAPELLE Marie-Christine
Université Clermont Auvergne

Mars 2023

le cnam
Cnesco

Centre national d'étude des systèmes scolaires

Pour citer ce document, merci d'utiliser la référence suivante :
Lafontaine, D. & Toczec-Capelle, M.-C. (2023). *L'évaluation en classe au service de l'apprentissage des élèves : rapport de synthèse*. Cnesco-Cnam.

Ce texte s'inscrit dans une série de ressources publiées par le Centre national d'étude des systèmes scolaires (Cnesco) sur la thématique : **Évaluation en classe**.

Les opinions et arguments exprimés n'engagent que les auteurs du rapport.

Disponible sur le site du Cnesco : www.cnesco.fr
Publié en mars 2023.
Centre national d'étude des systèmes scolaires
41 rue Gay-Lussac 75005 Paris

Contact : cnesco@lecnam.net - 06 98 51 82 75

Sommaire

Introduction : enjeux et périmètre de la conférence de consensus.....	6
Chapitre 1 : L'évaluation comme soutien d'apprentissage.....	8
A. Bref panorama des publications relatives à l'évaluation dans le primaire et le secondaire	8
1. Revues de littérature et méta-analyses	10
2. Les revues consacrées à l'évaluation	10
B. De l'évaluation formative à l'évaluation soutien d'apprentissage	11
C. Que dit la recherche à propos des effets de l'évaluation soutien d'apprentissage ?.....	16
D. Impact des feedbacks en contexte de classe	17
1. Qu'est-ce qu'un feedback ?	17
2. Effets des différents types de feedbacks	18
3. Caractéristiques des feedbacks les plus efficaces.....	19
Chapitre 2 : Processus cognitifs et processus psychosociaux impliqués dans l'évaluation	21
A. Processus cognitifs	21
B. Processus psychosociaux.....	23
1. Motivation	23
2. Le soi	24
3. Effets des situations évaluatives sur les performances	27
Chapitre 3 : Évaluation, différenciation, équité.....	33
A. Modèles de justice et conceptions de l'égalité.....	33
1. Le modèle de l'égalité des chances.....	33
2. Le modèle de l'égalité de traitement.....	34
3. L'égalité des acquis ou de résultats	34
B. Évaluer et différencier l'enseignement pour amener un maximum d'élèves à la réussite	35
C. Évaluer différemment des élèves différents.....	37
D. Changer les situations d'évaluation et les perceptions sources d'inégalités.....	38
1. L'évaluation sommative amplifie les inégalités, au détriment des élèves d'origine défavorisée.....	39
2. Le pouvoir de la situation et des perceptions	39
Chapitre 4 : Les pratiques d'évaluation en France	42
A. Repères historiques.....	42
B. Évolutions récentes	43
C. État des lieux des pratiques	45
1. Les pratiques évaluatives dans le premier degré.....	45
2. Les pratiques évaluatives dans le second degré	46
3. Les pratiques évaluatives à la lumière de trois études empiriques	48

Chapitre 5 : Enjeux de la formation des enseignants	52
A. L'évaluation soutien d'apprentissage : bien plus qu'un changement de méthode.....	52
B. Le développement d'une littératie de l'évaluation.....	52
C. Programmes de développement professionnel : effets et points d'attention	54
Références	56

Liste des Figures

Figure 1 : Nombre de documents recensés relatifs à l'évaluation en général et par domaine	9
Figure 2 : Nombre de documents recensés relatifs aux pratiques d'évaluation en général et par domaine.....	9
Figure 3 : Nombre de documents traitant de l'évaluation recensés comme revues de la littérature ou méta-analyses	10
Figure 4 : Assessment for learning : sa fonction formative et son interface avec l'évaluation sommative (traduit de Laveault & Allal, 2016, p. 8 – d'après les travaux d'Allal publiés au début des années 2010)	14
Figure 5 : Le modèle MTSS	36
Figure 6 : Le pouvoir de la situation (Goudeau, 2023)	39
Figure 7 : Liens entre typologie des situations évaluatives observées et fonctions de l'évaluation (Genelot, 2023).....	49

Liste des Tableaux

Tableau 1 : Perspective classique <i>versus</i> élargie de l'évaluation formative (inspiré de Allal & Mottier Lopez, 2005, p. 269 ; extrait de Fagnant, 2023).....	13
Tableau 2 : Les caractéristiques d'un feedback efficace selon la revue de la littérature (Calone & Lafontaine, 2023)	19
Tableau 3 : Compétences pour le développement d'une littératie en évaluation (Fontaine, 2023)....	53

Introduction : enjeux et périmètre de la conférence de consensus

Les apprentissages scolaires se construisent au sein de la cellule de base du système éducatif que constitue la classe. C'est en classe que l'évaluation des acquis des élèves prend place – et quelle place ! – tout au long de la scolarité obligatoire. Ainsi, le trio « enseignement-évaluation-apprentissage » forme un continuum qui apparaît indissociable. Les trois concepts sont interdépendants, mais quelle place occupe l'« évaluation » dans ce continuum ? Quels rôles joue-t-elle ?

Pour le grand public, l'évaluation est souvent conçue comme un bilan qui suit un enseignement et des apprentissages. Dans cette optique, elle permet de vérifier si les apprentissages visés sont acquis. Si cette évaluation-bilan a toute sa place à l'école (on parle à ce propos d'*évaluation sommative*), d'autres formes d'évaluation se déroulent en cours d'apprentissage et peuvent nourrir des ajustements ou régulations de l'action à la fois pour les enseignants et les élèves. Ce second type d'évaluation consiste à faire le point en cours d'apprentissage d'une notion ou d'une compétence, à informer l'élève de ce qui est déjà acquis, connu, et de ce qu'il lui reste à acquérir, développer, approfondir. Constatant certaines difficultés chez ses élèves, l'enseignant peut ajuster ou différencier son enseignement pour tenter de surmonter ces difficultés. Dans ce cas, on parle d'*évaluation formative*. La notion d'évaluation formative s'est développée dès la fin des années 1960 avec les travaux fondateurs de Bloom *et al.* (1971) et de Scriven (1967). De nos jours, la distinction entre évaluation formative et sommative tend à s'estomper et on parle plus volontiers d'*évaluation soutien à l'apprentissage*. C'est à cette forme d'évaluation que le Cnesco a décidé de consacrer sa conférence de consensus en 2022.

Les nombreuses études scientifiques qui portent sur l'évaluation relèvent de plusieurs disciplines universitaires : les sciences de l'éducation, la psychologie cognitive, la psychologie sociale, la sociologie, et bien entendu l'ensemble des didactiques des disciplines enseignées à l'école.

L'évaluation comportant de multiples facettes, il a fallu, pour cadrer cette conférence, commencer par en **définir le périmètre**.

Ainsi, l'évaluation concerne différents niveaux (évaluation des élèves, des enseignants, des écoles, du système). Par ailleurs, elle revêt différentes fonctions (sommative, formative, certificative, diagnostique), différentes formes (évaluations externes, nationales ou internationales, internes, à enjeux plus ou moins élevés).

Parmi ces types d'évaluation, le Cnesco a choisi de centrer la conférence de consensus 2022 sur l'évaluation menée par l'enseignant, au service de l'apprentissage (évaluation interne).

Ceci signifie que les travaux menés lors de cette conférence n'ont concerné :

- **ni les évaluations certificatives** (c'est-à-dire les évaluations permettant de délivrer un diplôme, le brevet, le bac) ;
- **ni les évaluations externes** (comme les évaluations standardisées nationales et internationales).

Ils se sont concentrés de manière privilégiée sur l'école du socle et sur la fonction formative de l'évaluation.

Depuis le début des années 2000, les travaux scientifiques consacrés à l'évaluation ont connu un important développement et une diversification tout aussi importante. On en sait aujourd'hui beaucoup plus sur ce qu'il convient de faire pour soutenir les élèves dans leurs apprentissages quotidiens, et encore davantage sur ce qu'il convient d'éviter de faire. On comprend beaucoup mieux les mécanismes qui lient les apprentissages et la dynamique motivationnelle (le concept de soi, les buts, la persévérance, la résignation, etc.) et on peut cerner avec une relative précision le type de pratiques d'évaluation qui engendrent un accroissement des inégalités plutôt qu'une réduction de celles-ci.

C'est sur cet éventail de recherches que la conférence de consensus a décidé d'appuyer ses travaux en sorte que le jury puisse dégager :

- des pistes concrètes d'évolution des pratiques, nuancées et pragmatiques, qui dépassent les clivages stériles (pour ou contre la note, évaluation de savoirs ou de compétences, évaluation formative *versus* sommative, etc.) ;
- des pistes qui contribuent à la réussite de tous les élèves et à la réduction des inégalités scolaires et sociales.

Cette note de synthèse, sans être exhaustive, constitue « la porte d'entrée scientifique » de la conférence de consensus. Elle s'appuie principalement sur :

- a. les rapports scientifiques rédigés en amont de la conférence par Annick Fagnant (Université de Liège) sur *L'évaluation en classe au service des apprentissages des élèves*¹ et par Anthony Calone et Dominique Lafontaine (Université de Liège) sur *L'impact des différents types de feedback en contexte de classe*².
- b. les notes de synthèse rédigées par les différents experts intervenant lors des deux journées de séances publiques de la conférence³.

La présente note tente donc l'exercice difficile de faire la synthèse des synthèses. Dans cette mesure, elle ne comportera pas de partie conclusive et encore moins de recommandations, qui sont du ressort du jury de la conférence de consensus. La note a sa propre cohérence et pourra soit être lue isolément, soit servir de document d'orientation pour lire les documents plus complets auxquels il sera clairement renvoyé.

¹ https://www.unesco.org/fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_FAGNANT.pdf

² https://www.unesco.org/fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_CALONE-LAFONTAINE.pdf

³ https://www.unesco.org/fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Chapitre 1 : L'évaluation comme soutien d'apprentissage

A. Bref panorama des publications relatives à l'évaluation dans le primaire et le secondaire

L'évaluation, les pratiques d'évaluation et les problématiques y afférant – apprentissages, motivation, équité, etc. – sont des thématiques très investiguées en sciences de l'éducation ainsi qu'en psychologie de l'éducation et en psychologie sociale. Pour mieux appréhender l'étendue et la variété des publications sur cette thématique, une recherche a été effectuée par Margault Sacré (Université de Liège) dans une des bases de données les plus importantes.

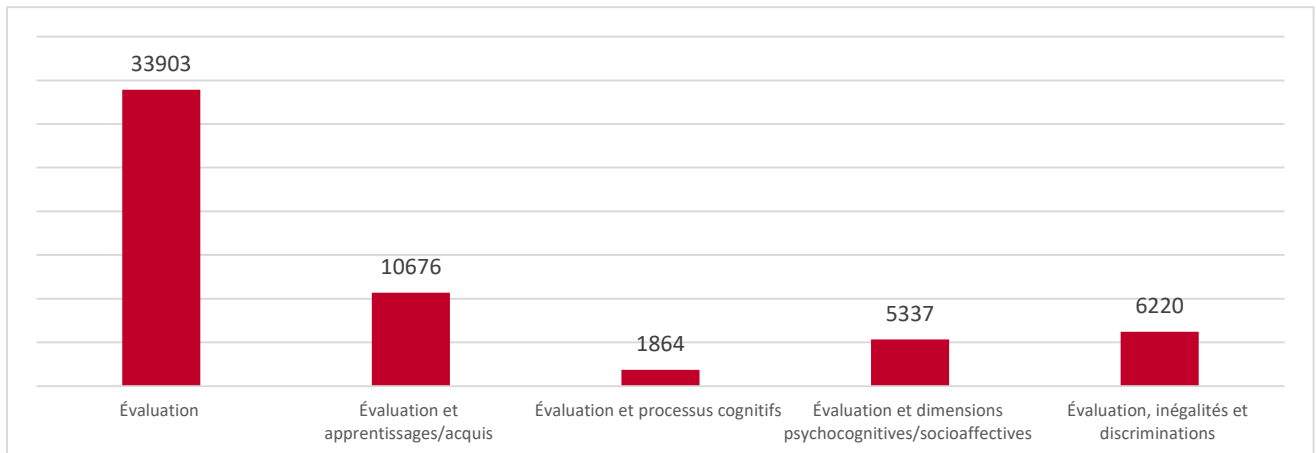
En mars 2023, la base de données *Education Resources Information Center* (ERIC) a été interrogée. Celle-ci a été choisie car plus de 80 % des documents recensés concernent l'éducation. ERIC fonctionne sur la base d'un *thesaurus*, c'est-à-dire un répertoire structuré de mots-clés : les descripteurs. Ainsi, chaque document est indexé sous ces descripteurs et quand on interroge la base de données en entrant un descripteur ou une combinaison de descripteurs, tous les documents correspondant à ces descripteurs sont identifiés.

Dans un premier temps, la recherche a visé à simplement quantifier le nombre de publications sur le thème de l'évaluation depuis l'année 2000. Ensuite, au sein de cet ensemble de publications, on a tenté de caractériser plus finement ce sur quoi portaient ces publications. Combien de publications portent à la fois sur l'évaluation et une série de domaines correspondant aux principaux thèmes de la conférence ?

- a. Évaluation et apprentissages/acquis (focus principal de la conférence de consensus, évaluation soutien d'apprentissage, thème « Enjeux ») ;
- b. Évaluation et processus cognitifs (processus d'apprentissage, stratégies d'apprentissage, métacognition, processus cognitifs ; thème 2 de la conférence : « dans la tête des élèves » ; thème 4 : « évaluation et processus d'enseignement-apprentissage »).
- c. Évaluation et dimensions psychocognitives/socioaffectives (motivation, engagement, attitudes, buts, intérêt, attention, participation, besoins, satisfaction, persistance) (thème 2 de la conférence : « dans la tête des élèves » ; thème 4 : « évaluation et processus d'enseignement-apprentissage » et thème « Enjeux »).
- d. Évaluation, inégalités et discriminations (thème 3 : « Évaluation et équité »).

La Figure 1 présente les résultats de cette recherche.

Figure 1 : Nombre de documents recensés relatifs à l'évaluation en général et par domaine

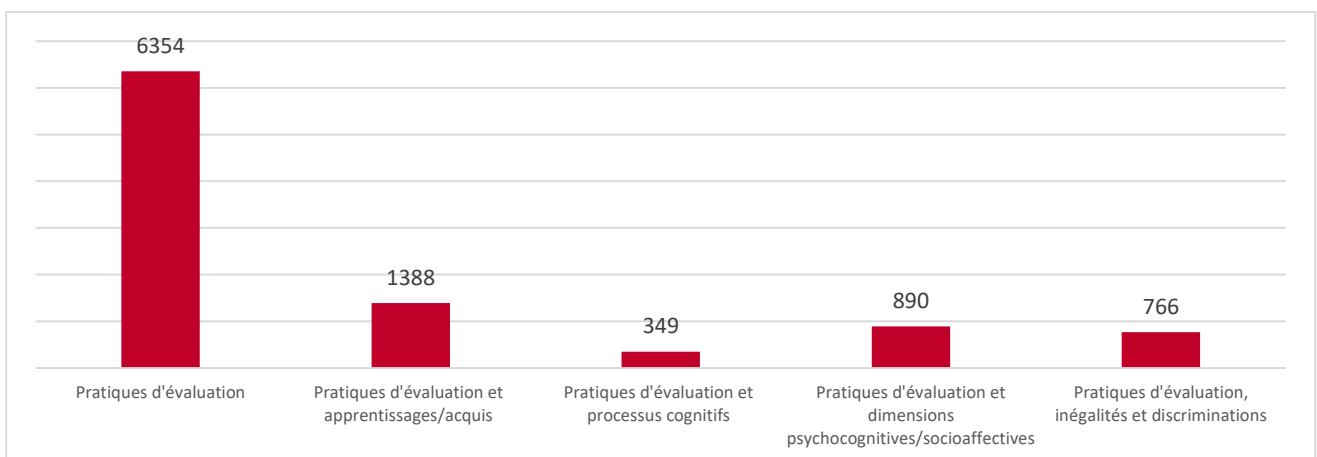


Pas moins de 33 903 documents concernant l'évaluation ont été identifiés. Parmi ces documents, environ un tiers (10 676) concernent les apprentissages ou les résultats des élèves.

En ce qui concerne ce qui se passe « dans la tête des élèves », on soulignera que les travaux consacrés aux processus cognitifs sont nettement moins nombreux (5 % de l'ensemble des documents traitant de l'évaluation) que ceux liés aux processus psychocognitifs ou socioaffectifs (15,7 %). Enfin, les documents relatifs aux inégalités et discriminations représentent une proportion non négligeable (18,3 %) de l'ensemble des documents en lien avec l'évaluation⁴.

Dans un deuxième temps, la même recherche a été lancée mais en incluant le terme « pratiques d'évaluation » dans les descripteurs.

Figure 2 : Nombre de documents recensés relatifs aux pratiques d'évaluation en général et par domaine



6 354 documents traitant des pratiques d'évaluation ont été identifiés. Ceci représente environ un cinquième (19 %) des documents traitant de l'évaluation en général.

⁴ Notons que tous les pourcentages présentés ne sont pas cumulatifs, car rien n'empêche que les documents soient référencés sous plus de deux descripteurs : un document référencé sous « évaluation » et « processus psychocognitifs » peut également être référencé sous « discriminations », en fonction des thématiques traitées dans le document.

La répartition plus fine est relativement similaire à celle observée pour les documents relatifs à l'évaluation en général. Parmi les documents relatifs aux pratiques, 22 % traitent également des résultats d'apprentissage, 5% seulement traitent des processus cognitifs, 14 % des processus psychocognitifs et 12 % des discriminations et des inégalités.

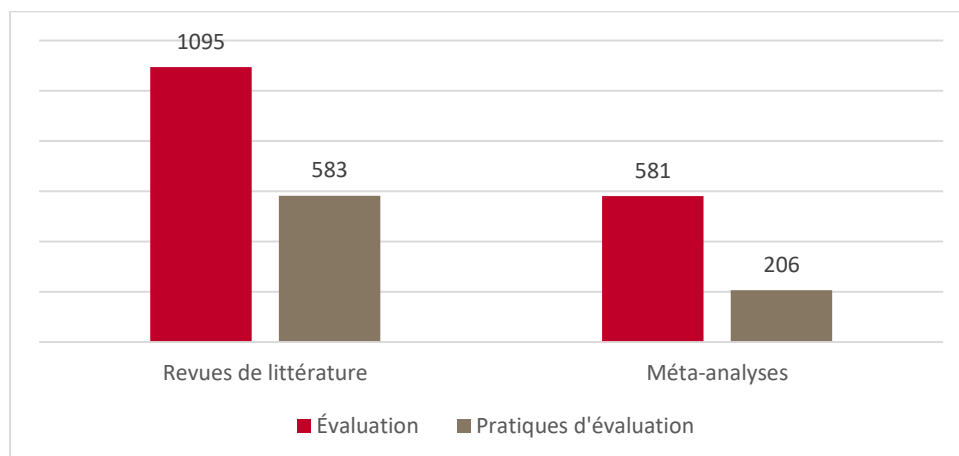
Enfin, dans un troisième temps, nous nous sommes intéressées d'une part au nombre de revues systématiques de la littérature et de méta-analyses consacrées à l'évaluation, d'autre part aux revues spécialisées portant sur le domaine de l'évaluation.

1. Revues de littérature et méta-analyses

L'existence de revues de littérature et de méta-analyses est en soi un indicateur significatif de l'importance et de la dynamique d'un champ de recherche. On ne trouvera en effet pas ou guère de revues systématiques sur des thématiques confidentielles ou marginales.

Pour repérer ce type d'études, les mêmes équations de recherche ont été utilisées en y ajoutant le descripteur *revues de la littérature* d'une part, le descripteur *méta-analyses* d'autre part. La Figure 3 reprend les résultats de chaque équation de recherche.

Figure 3 : Nombre de documents traitant de l'évaluation recensés comme *revues de la littérature* ou *méta-analyses*



Plus de 1 000 documents traitant de l'évaluation et comprenant le descripteur *revue de la littérature* ont été recensés, ce qui représente 3,2 % de l'ensemble des documents sur l'évaluation (33 903), ainsi que 581 méta-analyses. En ce qui concerne les pratiques, 583 documents sont des *revues de la littérature* et 138 documents sont des méta-analyses.

2. Les revues consacrées à l'évaluation

La base de données ERIC recense les documents issus de 1 288 revues avec comité de lecture. Elle propose un groupement par thématique de ces revues et, parmi ces thématiques, figure la thématique *Assessment and Evaluation*. Sous cette thématique sont indexées 84 revues avec comité de lecture (6,5 %) qui publient des articles consacrés à l'évaluation, mais aussi parfois des articles sur d'autres thématiques (revues plus généralistes).

Parmi les 84 revues avec comité de lecture, nous avons pu identifier 18 revues spécialisées qui comportent au moins un des termes *evaluation*, *assessment* ou *testing* dans le titre. 21 % des revues avec comité de lecture recensées sous la thématique *Assessment and Evaluation* sont donc des **revues spécialisées en évaluation**, ce qui montre l'importance du champ de l'évaluation en sciences de l'éducation. Les autres revues sont des revues plus générales qui publient occasionnellement, mais pas exclusivement, des articles sur le thème de l'évaluation.

En bref, de cet inventaire systématique ressortent trois constats :

- **l'évaluation est clairement une problématique centrale des travaux de recherches en éducation**, comme l'attestent le nombre de documents qui y sont consacrés, le nombre de revues systématiques de la littérature, de méta-analyses et le nombre de revues spécialisées ;
- les travaux plus spécifiquement consacrés aux pratiques sont proportionnellement assez peu nombreux (20 % des documents environ) ;
- les **travaux de recherche consacrés aux processus psychocognitifs liés à la dynamique motivationnelle** sont proportionnellement relativement nombreux. C'est sans doute sur ces questions que la recherche au cours des vingt dernières années a fait le plus d'avancées et a débouché sur les résultats les plus consensuels.

B. De l'évaluation formative à l'évaluation soutien d'apprentissage

Dès la fin des années 1960, la distinction entre *évaluation formative* et *évaluation sommative* a été posée par Scriven (1967) et par Bloom *et al.* (1971) dans le cadre de leurs travaux portant sur la pédagogie de maîtrise.

À l'opposé de l'évaluation sommative, qui vise à dresser un bilan des acquis des élèves après une période de formation, **l'évaluation formative se déroule en cours d'apprentissage**. C'est une prise d'informations qui peut déboucher sur des ajustements ou des régulations tant chez les élèves que chez leur enseignant. Scallon (1988) en donne la définition suivante :

« L'évaluation formative est un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage, avec l'intention de modifier la situation d'apprentissage ou le rythme de cette progression, pour apporter (s'il y a lieu) des améliorations ou des correctifs appropriés » (p. 155).

Ainsi, l'évaluation formative, dès l'origine, ne se résume pas à une prise d'informations, elle vise « l'adaptation des activités d'enseignement et d'apprentissage afin de favoriser la progression des apprenants vers les objectifs de formation » (Allal, 2008, p. 312). Comme le souligne Mottier Lopez (2023) dans sa note,

« les résultats qu'elles [les évaluations formatives] produisent fournissent une information à l'enseignant et à l'élève, permettant notamment de différencier les activités didactiques au regard des besoins des élèves (y compris pour les élèves de bon niveau scolaire) et de repérer des problèmes d'apprentissage et/ou d'enseignement. Elles donnent lieu à des rétroactions (ou *feedbacks*) : elles sont exploitées pour réguler (adapter) l'enseignement et soutenir l'autorégulation de l'élève. Elles peuvent être formelles ou informelles, instrumentées (sous forme papier crayon,

avec des grilles d'autoévaluation, des listes de vérification, etc.) ou non (dans l'interaction, par l'observation par exemple ») (Mottier Lopez, 2023, pp. 10-11).

L'évaluation formative est donc conçue d'emblée comme une **évaluation au service de l'apprentissage, au bénéfice de chacun des élèves**. Elle est intrinsèquement liée au postulat d'éducabilité énoncé par Bloom à la fin des années 1960 : tous les élèves sont capables de maîtriser des objectifs d'apprentissage pour autant qu'on leur accorde le temps nécessaire et un enseignement de qualité. Elle s'inscrit en rupture avec le modèle méritocratique de l'égalité des chances pour prôner un modèle d'égalité des acquis, où l'on vise la maîtrise des objectifs fondamentaux par tous les élèves⁵. Dans ce modèle, l'évaluation formative et les feedbacks sont les instruments de pilotage qui permettent de repérer les difficultés, les erreurs, d'ajuster, de différencier l'enseignement, les activités, le rythme, pour conduire un maximum d'élèves vers la réussite.

Au fil des années, plusieurs auteurs, en particulier francophones, ont considéré que « les deux seules catégories " formative " et " sommative " étaient réductrices pour rendre compte de la pluralité des buts et des fonctions de l'évaluation » (Mottier Lopez, 2023, p. 13). Suite à ce constat, certains auteurs (notamment De Ketele, 2010) ont proposé différentes typologies plus spécifiques dont on trouvera des exemples dans la note de Mottier Lopez (2023, pp. 13-14).

Si ces développements ont leur intérêt du point de vue de la recherche, « ces distinctions conceptuelles restent peu opérationnelles dans les pratiques de classe » (Mottier Lopez, 2023, p. 15), notamment parce que les enseignants ne font pas de distinction entre les fonctions de l'évaluation et qu'ils évoluent dans un contexte où la forme dominante de l'évaluation reste la note avec ses enjeux forts de classement, de sélection ou d'orientation. Ainsi, les enseignants sont pris entre des injonctions contradictoires, voire inconciliables, comme celle de noter, juger, classer, et celle d'une évaluation bienveillante selon la loi du 8 juillet 2013 (pour plus de détails, voir Butera, 2023 ; Merle, 2018 ; Mottier Lopez, 2023). Mottier Lopez pose dès lors la question :

« Est-ce alors une bonne idée de différencier de façon aussi pointue les différentes fonctions et démarches évaluatives, ou serait-il préférable de concevoir des approches plus intégratives, caractérisées par la finalité éducative visée plus que par les différentes décisions qui découlent de l'évaluation (régulation, certification, orientation) ? » (Mottier Lopez, 2023, p. 15).

Dans cet esprit, des auteurs, issus tant du monde anglophone que francophone (Laveault & Allal, 2016 ; Black & Wiliam, 2018 ; Crahay, Mottier Lopez & Marcoux, 2019), prônent au contraire une vision intégrative qui, plutôt que de les opposer, rapproche et conjugue évaluation formative et certains aspects de l'évaluation sommative, en insistant davantage sur la visée de l'évaluation que sur sa fonction. Ces travaux nous paraissant plus porteurs pour la pratique de l'évaluation au quotidien dans les classes, nous nous y arrêterons plus en détail⁶.

Tout d'abord, ces travaux s'inscrivent dans la lignée d'une vision élargie de l'évaluation formative, présentée dans le Tableau 1 suivant (extrait de Fagnant, 2023).

⁵ On trouvera plus d'informations à ce propos dans la note de F. Butera (2023). Nous y reviendrons en début de chapitre 3.

⁶ Pour plus d'informations à ce propos, on consultera le rapport de Fagnant (2023) et la note de Mottier Lopez (2023).

Tableau 1 : Perspective classique *versus* élargie de l'évaluation formative (inspiré de Allal & Mottier Lopez, 2005, p. 269 ; extrait de Fagnant, 2023)

Perspective « classique » (Bloom)	Perspective élargie
Insertion de l'évaluation formative après chaque phase d'enseignement/apprentissage	Intégration de l'évaluation formative dans toutes les situations d'apprentissage
Utilisation de tests formatifs	Utilisation de divers moyens de recueil d'information
Feedback + action corrective (remédiateive) ⇔ Remédiation	Feedback + adaptation de l'enseignement (différenciation) ⇔ Régulation
Gestion de l'évaluation formative par l'enseignant	Participation active des élèves dans l'évaluation formative (autoévaluation, évaluation mutuelle, co-évaluation)

Parmi les quatre types d'ouverture que propose la vision élargie, on insistera en particulier sur trois points :

- a) L'utilisation pour l'évaluation formative, non seulement d'évaluations formelles ou instrumentées (tests, contrôles), mais aussi **d'évaluations informelles**, notamment des observations des démarches des élèves ou des interactions entre élèves au sein de groupes.
- b) Le remplacement de la notion de remédiation par celle de **différenciation**. La notion de remédiation est jugée trop réductrice ; elle limite *de facto* les ajustements des activités aux élèves en difficulté alors que le concept de différenciation, plus englobant, concerne l'ensemble des élèves.
- c) Au concept de remédiation vient aussi s'ajouter celui de **régulation**, concept auquel Allal a consacré d'importants travaux⁷. Pour notre propos, il nous paraît important d'insister sur le concept de **corégulation**. Comme le souligne Mottier Lopez (2012), la vision originale de l'évaluation formative donnait à penser d'une part que « toute intervention était régulation », d'autre part que l'enseignant peut « réguler l'élève » de façon directe (Mottier Lopez, 2012, p. 31). En fait, les élèves ne peuvent se réguler eux-mêmes que quand ils se saisissent des interventions ou des feedbacks de l'enseignant et en tirent réellement parti. C'est pourquoi la régulation implique nécessairement l'élève et l'enseignant ; il s'agit donc bien de corégulation enseignant-élève⁸. Nous y reviendrons dans la section consacrée aux feedbacks.

Ensuite, certains auteurs (Bennett, 2011) soulignent que les deux fonctions – formative et sommative – de l'évaluation sont interreliées et qu'il est artificiel de vouloir les distinguer, en particulier quand on cherche à faire évoluer les pratiques des enseignants. Ainsi, les évaluations sommatives peuvent aussi informer les enseignants sur les acquis, les difficultés des élèves et les amener à cerner les besoins des élèves et à différencier leur enseignement. Black et Wiliam (2018) soulignent à ce propos : « de notre travail avec les enseignants, nous avons retiré la conviction que

⁷ Pour plus de détails, voir la note de Mottier Lopez (2023), p. 15.

⁸ À ne pas confondre avec la coévaluation. Voir à ce propos la note de Morales Villabona (2023).

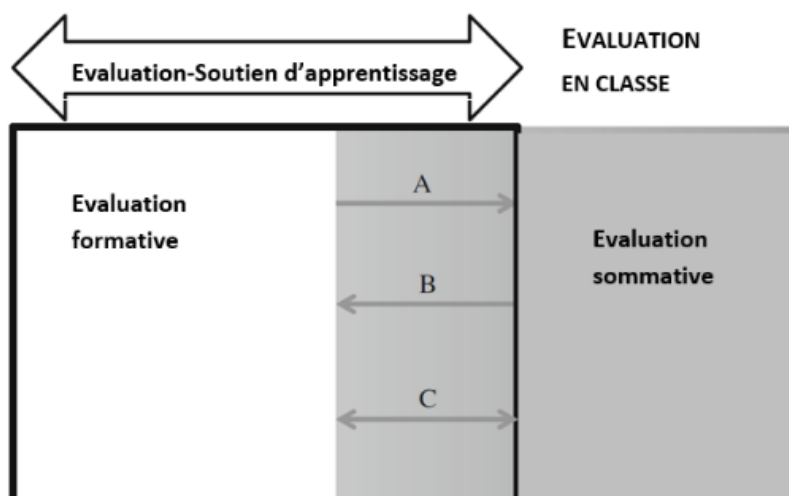
toute tentative d'amélioration des pratiques de classe se concentrant sur l'évaluation doit intégrer tous les aspects de l'évaluation » (Black & Wiliam, 2018, p. 19).

Ce rapprochement entre évaluation formative et sommative trouve son aboutissement dans la notion d'**évaluation soutien d'apprentissage**⁹ qui s'impose à partir des années 2010. Laveault et Allal (2016) en proposent une définition qui englobe les différentes formes d'évaluation en classe :

« L'évaluation-soutien d'apprentissage désigne la collecte et l'interprétation d'informations issues d'évaluations dont l'utilisation intentionnelle permet aux enseignants et aux élèves, qu'ils agissent individuellement ou de manière interactive, de prendre des décisions qui ont un impact positif sur l'enseignement et l'apprentissage » (Laveault & Allal, 2016, p. 7).

Voici comment Laveault et Allal schématisent les relations et le recouvrement des pratiques d'évaluations formative et sommative au sein de la classe.

Figure 4 : Assessment for learning : sa fonction formative et son interface avec l'évaluation sommative (traduit de Laveault & Allal, 2016, p. 8 – d'après les travaux d'Allal publiés au début des années 2010)



La zone en grisé au centre correspond au recouvrement entre les deux types d'évaluation.

La *flèche A* correspond aux situations où des informations issues des évaluations formatives alimentent l'évaluation sommative, par exemple lorsque l'enseignant tient compte, dans sa note, d'observations plus informelles menées dans le quotidien de la classe.

À l'inverse, la *flèche B* correspond aux situations où l'enseignant utilise des informations issues d'évaluations sommatives ou de bilans/contrôles à des fins formatives pour adapter son enseignement (revenir sur des points de matière non maîtrisés, proposer des activités différenciées, analyser les erreurs, etc.)

Enfin, la *flèche C* correspond à la zone d'interaction directe entre les fonctions formative et sommative. Laveault et Allal (2016, cités par Fagnant, 2023) donnent à ce propos l'exemple du

⁹ L'expression « évaluation soutien d'apprentissage » a été proposée par Allal et Laveault (2009) pour traduire le concept « *Assessment for learning* ».

portfolio qui peut donner lieu à une note finale, mais qui, en cours d'élaboration, peut faire l'objet de feedbacks destinés à améliorer la production de l'élève.

Pour Laveault et Allal, les éléments se situant dans la zone de recouvrement (flèches A à C) témoignent « d'une volonté de l'enseignant de soutenir les apprentissages des élèves tout en répondant aux attentes institutionnelles en matière d'évaluation sommative, conduisant à mettre des points ou des appréciations dans des bulletins ou autres documents officiels (Fagnant, 2023, p. 21).

C'est dans cet esprit que s'inscrivent les travaux de recherche et de formation de Pasquini qui visent le développement de pratiques de notation constructives au service des apprentissages des élèves (Pasquini, 2023). Pasquini part du principe que « l'évaluation notée est parfois exploitée à des fins non pédagogiques [...] : maintenir les élèves dans le travail, les motiver à apprendre, gérer la discipline, asseoir le pouvoir de l'enseignant, ou encore récompenser (un effort par exemple) ». Considérant que dans l'état actuel des pratiques, les enseignants continuent à noter les élèves et y consacrent beaucoup de temps, Pasquini entend intégrer l'évaluation sommative notée « dans une réflexion systémique centrée sur les processus d'enseignement-apprentissage » en sorte que celle-ci « participe à faire progresser chaque élève » (Pasquini, 2023, p. 119). « L'enjeu est de pouvoir offrir des possibilités de rétroactions aux élèves afin de soutenir leurs apprentissages à l'aide de notes à haute valeur informative. En cela, nous parlons de note constructive » (Pasquini, 2023, p. 115).

Pour amener les enseignants à développer des notes constructives, Pasquini s'appuie sur trois axes de travail : l'alignement curriculaire (dont Pasquini propose un modèle dit « élargi »), la planification à rebours¹⁰ et l'approche critériée.

L'alignement curriculaire désigne la triple concordance entre les objectifs, les méthodes et l'évaluation.

« Le modèle élargi – qui désigne le fort lien de cohérence systémique existant entre les compétences évaluées (référenciation restreinte), les tâches évaluatives (design), les points et/ou les critères attribués (pondération) et l'échelle (notation) (Pasquini, 2018, 2019a) - permet d'appréhender les conditions concrètes nécessaires à l'élaboration d'une note constructive » (Pasquini, 2023, p. 116).

Les étapes de la planification à rebours sont les suivantes :

- « déterminer les compétences du programme à atteindre ;
- identifier les critères clés caractérisant l'atteinte de ces compétences ;
- construire des situations et des tâches d'apprentissage alignées avec les compétences et les critères ;
- collecter des informations sur les apprentissages des élèves tout au long de la séquence en lien avec les attentes formulées en pensant aux attentes relatives à la certification » (Pasquini, 2023, pp. 116-117).

¹⁰ Sur la question de la planification, voir aussi la note de Yerly (2023).

L'approche critériée se révèle plus constructive qu'une approche normative car elle permet d'identifier, tant pour l'élève que pour l'enseignant, quels savoirs ou compétences sont acquis et lesquels il convient de retravailler. De surcroît, elle permet de clarifier les attendus et fait passer la comparaison entre élèves à l'arrière-plan. Amené à réfléchir à des critères,

« l'enseignant, pour établir des seuils de notes, change son questionnement, qui passe de "à quel pourcentage de points met-on le seuil de suffisance ?" à "quelles preuves d'apprentissage l'élève doit-il fournir dans cette situation pour obtenir telle note, synonyme d'apprentissages en construction ou réalisés ?". Dans une telle perspective, les points peuvent toujours être exploités, mais ils s'appartiennent à des critères qui renvoient aux apprentissages évalués » (Pasquini, 2023, pp. 117-118).

C. Que dit la recherche à propos des effets de l'évaluation soutien d'apprentissage ?

Cette section s'appuie essentiellement sur le chapitre II du rapport scientifique de Fagnant (2023) consacré à ces effets. Nous y renvoyons le lecteur pour plus de détails.

Pour répondre à cette question, Fagnant s'est essentiellement appuyée sur des articles de synthèse qui font eux-mêmes le bilan des recherches et des connaissances sur le sujet à partir d'un ensemble nourri de publications (notamment celui de Black & Wiliam, 1998).

Ces articles de synthèse concluent que l'évaluation formative améliore l'apprentissage : même si ces articles n'ont pas quantifié les effets, les progrès observés sont notables et se situent parmi les plus importants observés pour des interventions éducatives. La récente revue systématique de Lee *et al.* (2020) identifie quant à elle un effet moyen de 0,29, qui peut être considéré comme un effet faible à modéré¹¹.

Black et Wiliam (1998) identifient quelques points d'attention :

- a) les effets observés semblent particulièrement bénéfiques pour les élèves en difficulté d'apprentissage ;
- b) plusieurs travaux mettent en évidence des tensions entre les fonctions formative et sommative de l'évaluation, tant pour les enseignants que pour les élèves ;
- c) pour être efficace, l'évaluation formative nécessite bien plus qu'un changement marginal ; elle doit s'intégrer pleinement dans les pratiques de classe, d'où l'importance de prendre en compte les perceptions, les croyances et le climat de classe (voir 0 sur ces questions).

Au terme de leur article, Black et Wiliam (1998) soulignent une importante difficulté qu'ils ont rencontrée dans leur entreprise de synthèse : le concept d'évaluation formative est souvent flou et

¹¹ Une taille d'effet doit dépasser 0,20 pour être considérée comme statistiquement significative. Une taille d'effet de 0,20 est considérée comme faible, à partir de 0,50 comme modérée et à partir de 0,80 comme forte. Précisons toutefois qu'en matière de pédagogie, bien peu de « traitements » (pratiques, méthodes, dispositifs, politiques éducatives, etc.) atteignent un effet fort et même modéré. Ceci ne signifie en rien que les effets considérés comme faibles soient négligeables et ne méritent pas d'être pris en considération, ne serait-ce qu'en raison du fait que plusieurs traitements peuvent être combinés.

mal défini dans de nombreuses études ; il recouvre ainsi une série de pratiques aussi diverses que la pédagogie de maîtrise, les évaluations en classe, ou encore les feedbacks.

Bennett (2011), dans sa revue critique, souligne aussi la faiblesse conceptuelle des définitions de l'évaluation formative ou de l'*Assessment for learning* dans les études recensées. Celle-ci rend complexe toute entreprise de synthèse systématique ou de méta-analyse « car si nous ne pouvons pas clairement définir une innovation, il n'est pas possible de documenter son efficacité de façon sensée » (Bennett, 2011, p. 8).

En 2005, Allal et Mottier Lopez, de leur côté, avaient pointé le manque de travaux empiriques solides dans le monde francophone pour étayer les effets et les conditions d'efficacité des pratiques d'évaluation formative.

Au terme de cette section, il est difficile de contenir un sentiment de frustration face à cette relative absence de résultats empiriques concluants (articles de synthèses, méta-analyses), alors qu'il est pourtant communément admis que l'évaluation peut soutenir l'apprentissage de manière efficace. Il n'y a cependant pas de quoi s'en alarmer outre mesure. La difficulté est bien celle sur laquelle Black et Wiliam (1998) et Bennett (2011) ont mis le doigt : celle d'un ensemble de pratiques ou de composantes de l'évaluation formative ou de l'évaluation soutien d'apprentissage mal défini et sans doute trop large. Cette difficulté de réaliser des revues systématiques ou des méta-analyses sur des ensembles trop hétérogènes ou trop larges a été observée à maintes reprises dans le domaine de l'éducation. Un autre exemple connu est celui des pédagogies dites actives. Celles-ci sont trop diverses et trop mal définies pour qu'un quelconque effet bénéfique puisse être dégagé. Si la difficulté est connue des chercheurs avertis, la solution à ce type d'obstacle l'est aussi. Il convient alors de restreindre l'étendue de la pratique, de la démarche ou de la méthode à laquelle on s'intéresse et de porter son attention sur certaines de ses composantes mieux définies.

À cet égard, les feedbacks sont considérés comme une composante essentielle et incontournable de l'évaluation soutien d'apprentissage. C'est pourquoi nous allons maintenant nous intéresser aux recherches qui ont porté sur les effets des feedbacks et qui ne posent pas ce problème de flou conceptuel.

D. Impact des feedbacks en contexte de classe

Cette section emprunte ses principaux résultats et conclusions au rapport scientifique rédigé en amont de la conférence de consensus par Calone et Lafontaine (2023).

1. Qu'est-ce qu'un feedback ?

« Selon une conception relativement simple, le feedback est une information fournie par une personne (enseignant, pair, parent...) voire par un ordinateur concernant la performance d'un individu » (Hattie & Timperley, 2007). Souvent, le feedback est perçu comme une simple transmission d'information qui met en rapport l'état actuel de performance avec une performance attendue (Nicol & Macfarlane-Dick, 2006). Selon cette conception, les enseignants transmettent aux élèves une information sur ce qui est correct ou pas. Dans le meilleur des cas, ils leur font un commentaire sur leurs forces ou leurs faiblesses, et c'est aux élèves d'utiliser ces informations pour s'améliorer. Cependant, depuis près de vingt-cinq ans, de nombreux chercheurs ont essayé de

nuancer et d'enrichir cette représentation (Boud, 2000 ; Hattie & Timperley, 2007 ; Sadler, 1998 ; Yorke, 2003).

En effet, considérer le feedback comme un processus cognitif impliquant uniquement un transfert d'informations suppose que celui-ci ne peut influencer ni la motivation, ni l'affect et les croyances des élèves. Or, Dweck (2002) a montré que les feedbacks externes influencent la perception que les élèves ont d'eux-mêmes et leurs croyances motivationnelles. L'impact du feedback dépasse donc le champ de l'apprentissage au sens strict (Calone & Lafontaine, 2018). Allal (2007) parle de « régulation sur les composantes cognitives et métacognitives d'une part, de régulation sur les composantes affectives et motivationnelles d'autre part » (Calone & Lafontaine, 2023, p. 8).

2. Effets des différents types de feedbacks

Plusieurs méta-analyses (Hattie & Timperley, 2007 ; Hattie, 2009) ont été consacrées à l'efficacité du feedback élaboré, comparativement au feedback sous forme de simple note sans commentaire. Nous ne reprendrons ici que les résultats de la dernière méta-analyse en date.

En 2020, Wiesniewski, Ziéer et Hattie ont réalisé une méta-analyse ; celle-ci synthétise les résultats de 435 études menées sur les effets du feedback sur les apprentissages, le comportement et la motivation.

La taille d'effet moyenne est de 0,55. Cet effet est important, même s'il est un peu moins élevé que dans les trois méta-analyses précédentes ($d = 0,79$ dans Hattie & Timperley, 2007).

Plus important, les résultats confirment une importante variation des tailles d'effet selon les études. Ceci signifie que le feedback ne peut pas en soi être considéré comme une pratique dont les effets seraient garantis quels qu'en soient la forme ou le contenu.

Deux résultats sont particulièrement importants à épingle :

- *le feedback a davantage d'effet sur les gains d'apprentissage ($d = 0,51$) que sur la motivation ($d = 0,33$). On considère généralement que le feedback peut influencer la motivation intrinsèque, le sentiment d'efficacité perçue et le sentiment de contrôlabilité des tâches. Les effets sont toutefois relativement faibles.*

- *le feedback élaboré a nettement plus d'effet ($d = 0,99$) que de simples feedbacks comme le renforcement ou la sanction ($d = 0,24$). Un feedback dit élaboré contient de l'information sur la tâche, le processus ou parfois l'autorégulation. Selon les auteurs de la méta-analyse, l'effet substantiel observé "suggère que les élèves bénéficient du feedback quand il les aide non seulement à comprendre quelles erreurs ils ont faites, mais aussi pourquoi ils les ont commises et comment ils peuvent les éviter à l'avenir" (Wiesniewski, Zierer & Hattie, 2020, p. 12). (Calone & Lafontaine, 2023, p. 30)*

Certains types de feedback sont donc un moyen puissant de soutenir l'apprentissage.

Mais quelles sont les caractéristiques des feedbacks les plus efficaces mises en évidence par les travaux de recherche ?

3. Caractéristiques des feedbacks les plus efficaces

Tableau 2 : Les caractéristiques d'un feedback efficace selon la revue de la littérature (Calone & Lafontaine, 2023)

Recommandations	Description détaillée et références
Le feedback doit mettre en évidence ce que l'enseignant attendait de l'élève, ce qui est produit (forces/faiblesses) et les moyens de remédier aux difficultés.	Ne pas axer seulement le feedback sur les forces et les faiblesses, fournir à l'élève des suggestions pour l'amélioration (Hattie & Timperley, 2007 ; Orsmond <i>et al.</i> , 2002 ; Rakoczy <i>et al.</i> , 2019 ; Sadler, 1989 ; Wollenschläger <i>et al.</i> , 2016).
Centrer le feedback sur la tâche, le processus et non sur la personne de l'élève.	Éviter les louanges ou réprimandes (Hattie & Timperley, 2007 ; Kluger & DeNisi, 1996) et proposer un feedback centré sur le processus mis en œuvre (Rakoczy, 2013) ⁹ , en s'attardant davantage sur la bonne réponse et la démarche à entreprendre que sur la mauvaise réponse (Kluger & DeNisi, 1996).
Le feedback doit être axé sur des critères.	Communiquer les critères d'évaluation à l'élève et fournir un feedback en rapport à ces critères (Nicol & Macfarlane-Dick, 2006). Les commentaires doivent être précis et personnalisés (Galand, 2006), limités en nombre et s'en tenir aux éléments essentiels.
Le feedback doit être fourni et réutilisé rapidement.	Procéder à deux moments d'évaluation relativement rapprochés, entrecoupés d'un feedback élaboré (Crahay & Lafontaine, 2012 ; Gibbs & Simpson, 2004).

Au terme de leur revue des travaux consacrés aux effets des feedbacks, Calone et Lafontaine (2023, pp. 35-37) tirent un ensemble de conclusions que nous reprenons de manière synthétique ici :

- Même si certaines caractéristiques des feedbacks efficaces ont pu être identifiées, il n'existe pas de véritable « recette » pour un bon feedback. Tout dépend de l'objectif de cette pratique (est-il destiné à aider l'élève ou à aider l'enseignant à mieux connaître son élève ?), de la matière enseignée et des caractéristiques des élèves.
- En tout état de cause, le feedback le plus porteur sera celui qui fournira des informations claires en fonction de l'objectif d'apprentissage poursuivi. À cet égard, le feedback normatif sous forme de note isolée est sans conteste le moins bénéfique, et peut même s'avérer délétère sur le plan socioaffectif et motivationnel (Butera, 2011 ; Deci, Koestner & Ryan, 1999¹²). Il est donc à éviter autant que possible. Des avatars à première vue plus sympathiques, comme des smileys, des échelles de couleurs ou encore des appréciations sous forme de lettres, s'ils ont l'avantage de mettre un peu à l'arrière-plan le côté normatif des notes chiffrées, n'en demeurent pas moins pauvres et ne permettront pas réellement à l'élève de s'améliorer.
- Pour être efficace, un feedback doit présenter certaines caractéristiques : il doit être fondé sur **des critères de réussite et mettre en évidence les processus nécessaires à la résolution de la**

¹² Voir chapitre 2 pour plus de développements.

tâche. Les feedbacks sont également plus porteurs lorsqu'ils ne sont **pas trop chargés en informations** et lorsqu'ils sont **fournis assez rapidement**¹³.

- Le feedback n'est pas uniquement un outil à destination des élèves. Le pouvoir du feedback semble bien plus fort lorsque l'enseignant recherche des informations sur les élèves : que savent-ils vraiment ? Qu'ont-ils bien ou mal compris ? Où commettent-ils des erreurs ? Quelles stratégies utilisent-ils ? Interpréter les réponses des élèves permet à l'enseignant de fournir un feedback plus adapté. Ainsi, le feedback s'inscrit pleinement dans une dynamique d'apprentissage et d'enseignement. C'est pour cela qu'il n'est pas qu'un simple transfert d'informations. **Pour donner un feedback à un élève, l'enseignant réalise un travail en amont – interpréter la réponse de l'élève et ses démarches d'apprentissage – et un travail en aval – s'assurer que le feedback soit compris et que l'élève sache l'utiliser pour s'améliorer.**
- **Pour que le feedback soit profitable, il faut que les élèves en perçoivent l'utilité** et sachent effectuer le lien entre le feedback reçu et la manière dont il faut s'y prendre pour progresser. Pour cela un apprentissage de la manière de décoder un feedback – comment en dégager les informations utiles en vue de progresser – est indispensable pour maximiser l'impact de ce dernier.
- La surcharge de travail que ceci peut occasionner pour les enseignants est un élément non négligeable à prendre en compte. Un enseignant peut se trouver enfermé dans la logique des « trimestres », des « bilans sommatifs » et des « notes chiffrées dans les bulletins scolaires », et il peut lui rester très peu de marge de manœuvre pour pratiquer une évaluation formative.
- En définitive, si les feedbacks fournis aux élèves sont potentiellement porteurs d'efficacité, ce n'est pas tant la forme du feedback qui prime (pour autant qu'il soit constructif sur le fond), que le contexte plus large dans lequel il est formulé qui peut conduire à un enseignement plus efficace, motivant et équitable. Une pratique occasionnelle du feedback constructif, même reflétant toutes les conditions propices identifiées par les recherches, ne peut donner que des effets limités. **Les feedbacks élaborés ne peuvent donner toute leur puissance que dans des systèmes éducatifs où la culture de l'évaluation privilégie résolument une évaluation formative, bienveillante, au service d'une dynamique de l'apprentissage où les erreurs ont toute leur place.** À cet égard, la France et la Belgique francophone notamment ont encore du chemin à parcourir pour, au-delà des déclarations d'intention, ancrer dans les pratiques de classe une autre culture de l'évaluation au service de l'apprentissage de tous les élèves.

¹³ Sur les possibilités qu'offre la technologie pour fournir des feedbacks plus rapides voire automatiques, on consultera la note de Silvestre et Parmentier (2023).

Chapitre 2 : Processus cognitifs et processus psychosociaux impliqués dans l'évaluation

A. Processus cognitifs

Plusieurs travaux de recherche se sont intéressés aux processus cognitifs en jeu lors des évaluations en classe au service de l'apprentissage des élèves. Des recherches ayant l'objectif de modéliser l'évaluation soutien d'apprentissage (Fagnant, 2023) et des travaux cherchant à articuler des processus de mémorisation et des processus d'évaluation (Gaonac'h, 2023).

Concernant la modélisation des processus soutenant les apprentissages en classe, même si nous les avons développés dans le chapitre précédent, rappelons simplement qu'Allal (2007 ; 2020) et Wiliam (2011) ont proposé une modélisation dans laquelle les **processus d'autorégulation sont centraux**.

« Au cœur du modèle, le processus d'autorégulation correspond à la régulation interne qui se produit au niveau de l'apprenant [...]. Toutefois, dans ce modèle situé, Allal (2020) considère que tout apprentissage en classe est coréglé et que l'autorégulation n'existe pas en tant qu'entité pleinement indépendante. Les propos suivants clarifient cette idée : "Même lorsque les élèves travaillent en autonomie sur une tâche, contrôlent activement leur propre avancement, utilisent des stratégies pour orienter et ajuster leur progression dans la tâche, ils le font dans un contexte et avec des outils qui sont des constructions sociales et culturelles. Dans cette perspective, la corégulation n'est pas un processus transitoire amenant à l'autorégulation autonome. À un certain moment, les apprenants – en interaction avec les enseignants et/ou avec leurs pairs – intériorisent des stratégies de régulation spécifiques qui deviennent partie intégrante de leur répertoire d'autorégulation" (Allal, 2020, p. 341) » (Fagnant, 2023, p. 36).

Soulignons ici que **ces processus d'autorégulation mettent en jeu des activités cognitives et métacognitives de celui qui apprend**.

À ce sujet, plusieurs recherches en psychologie (par exemple, Gaonac'h, 2022 ; Brown, Roediger & McDaniel, 2016) s'intéressent aux processus cognitifs impliqués lors de certaines situations évaluatives. De tels travaux sont particulièrement intéressants dans la mesure où ils mettent en évidence le fait que les modalités évaluatives sont susceptibles de réguler la qualité du traitement de l'information réalisé par celui qui apprend et de ce fait, détermine la qualité de son apprentissage. Plus précisément, les recherches de Roediger (Karpicke & Roediger, 2008) ont montré que **tester régulièrement les connaissances est un excellent moyen de les réactiver en mémoire et ainsi, de lutter contre l'effacement en mémoire à long terme**.

Qu'entendent les psychologues cognitifs par « tester des connaissances » ? Il s'agit simplement d'un exercice de rappel des connaissances à maîtriser sous forme de quiz ou de questions. Les effets bénéfiques de ce type d'exercice sont connus des chercheurs depuis des décennies.

En d'autres termes, l'effet-test apparaît bénéfique lorsque celui qui apprend révise en se testant par rapport à des élèves qui réviseraient avec d'autres méthodes de travail, comme relire le cours par exemple.

En référence à cet effet positif, Gaonac'h (2023) explique de manière très précise les conditions de réalisation de ces tests. Il souligne notamment que l'effet est plus marqué si on introduit un délai

entre le moment où l'on apprend et le test (Butler, Karpicke & Roediger, 2007). Et même si ce délai est dépendant de la nature de l'apprentissage, il poursuit en précisant qu'un « test immédiat n'a que très peu d'effet, et la nécessité d'un délai suffisamment long apparaît surtout lorsque la mémorisation finale est évaluée à long terme (rappel final plusieurs semaines ou plusieurs mois après les exercices) » (Gaonac'h, 2023, p. 123).

L'explication sous-jacente à ce processus est intéressante puisqu'elle éclaire les processus cognitifs impliqués à la fois dans les processus de mémorisation et les processus d'évaluation. En effet,

« la récupération d'une information en mémoire est en elle-même une activité cognitive, qui demande plus ou moins d'effort. On peut admettre que lorsque la phase de test n'est pas immédiate, le rappel demande davantage d'effort qu'un rappel immédiat : cet effort est centré sur la récupération de traces en mémoire. L'hypothèse qu'on peut alors avancer est que c'est cet effort lui-même qui joue un rôle favorable pour la mémorisation, à travers une activité cognitive spécifique qui a des conséquences notamment sur l'élaboration de traces incluant des indices de récupération. On a d'ailleurs montré à ce propos que l'effet-test est plus important si la phase de test impose un rappel, et non la simple reconnaissance (à travers un QCM par exemple) des informations apprises (Rowland, 2014) » (Gaonac'h, 2023, p. 124).

Pour aller plus loin, Gaonac'h (2023) développe les processus d'intégration des informations en mémoire en présentant les processus de consolidation et la réorganisation des circuits neuronaux concernés par l'information acquise.

Des travaux ont également étudié les bénéfices d'une évaluation formative considérée comme favorisant une activité métacognitive de celui qui apprend. Néanmoins,

« si de nombreuses données permettent d'affirmer que les compétences métacognitives d'un individu peuvent avoir des effets bénéfiques sur la réussite d'un apprentissage (par exemple de Boer et al., 2018, pour une synthèse), il faut cependant d'abord admettre que ce n'est pas toujours le cas, et la divergence entre croyances affichées et stratégies mises en œuvre constitue un des thèmes de prédilection des recherches sur la métacognition » (Gaonac'h, 2023, p. 125).

Ainsi, Gaonac'h (2023) s'intéresse au cas des évaluations formatives en questionnant les effets potentiels d'une mauvaise représentation des processus d'apprentissage réellement en jeu par celui qui apprend. Son argumentation scientifique s'appuie notamment sur les processus explicites et processus implicites convoqués lors des activités métacognitives et ce, en lien avec les processus mnésiques. À ce sujet, le chercheur propose de retenir que :

« La fréquence, et surtout la variété, des situations d'évaluation, constituent en soi des conditions favorables au développement de compétences de régulation qui ne relèvent pas nécessairement d'une régulation explicite, liée à la prise de conscience de l'état d'un apprentissage. Les situations d'évaluation doivent aussi permettre à l'apprenant d'être confronté, à travers différents contenus d'apprentissage, à des contraintes d'apprentissage suffisamment variées pour que les mécanismes implicites, dont on connaît maintenant la puissance, puissent devenir opérationnels » (Gaonac'h, 2023, p. 127).

En parallèle, d'un point de vue didactique, les travaux de Roussel (2013) par exemple, développent une manière d'utiliser l'évaluation pour favoriser l'apprentissage des langues vivantes étrangères en se centrant sur les processus cognitifs des élèves. Plus précisément, elle étudie deux types de démarches pouvant être mises en œuvre afin d'entraîner les élèves à une meilleure compréhension orale : l'enseignant peut entraîner les processus de bas niveau¹⁴ et/ou encourager la mise en œuvre des processus de haut niveau¹⁵.

L'ensemble de ces travaux centrés sur les processus cognitifs permettent de mieux comprendre l'articulation entre les trois concepts interdépendants que sont l'évaluation, l'apprentissage et l'enseignement. Toutefois, au cœur des processus d'autorégulation des apprentissages sont également convoqués d'autres processus complémentaires intégrant à la fois la dimension sociale de la situation et/ou la dimension sociale des acteurs concernés. Il s'agit des processus psychosociaux.

B. Processus psychosociaux

De tels processus sont étudiés par la psychologie sociale qui

« s'intéresse, quels que soient les stimuli ou les objets, à ces événements psychologiques fondamentaux que sont les comportements, les jugements, les affects et les performances des êtres humains en tant que ces êtres humains sont membres de collectifs sociaux ou occupent des positions sociales (en tant donc que leurs comportements, jugements, affects et performances sont en partie tributaires de ces appartenances et positions) » (Beauvois, Dubois & Doise, 1999, p. 311).

1. Motivation

Un premier processus apparaît comme fondamental dans tout type d'apprentissage puisqu'il influence à la fois le choix des activités des apprenants, mais également leur persévérance et leurs performances dans la réalisation de ces activités (Cosnefroy & Fenouillet, 2009 ; Viau, 1994). Il s'agit de la motivation. La motivation correspond aux « forces internes et/ou externes produisant le déclenchement, la direction, l'intensité et la persistance du comportement » (Vallerand & Thill, 1993).

Certains travaux ont étudié les liens entre motivation, évaluation et apprentissage (Calone & Lafontaine, 2023). Ils synthétisent **les effets délétères de l'évaluation normative** et ainsi présentent **les effets néfastes de l'utilisation de la note isolée** lorsqu'elle est utilisée en tant qu'évaluation normative.

« De nombreux chercheurs (Butler & Nisan, 1986 ; Pulfrey et al., 2011) ont montré l'effet délétère de l'évaluation normative sur l'intérêt des élèves à apprendre, que ce soit pour les élèves les plus faibles, mais aussi parfois, de manière moins attendue, pour les meilleurs. Ceci s'expliquerait par un effet de sur-justification. L'octroi d'une récompense (la note) pour un apprentissage réussi ou sur la base d'un engagement de l'élève aurait un impact fortement négatif sur la motivation intrinsèque

¹⁴ Certaines opérations sont dites « de bas niveau » parce qu'elles concernent la manipulation directe du langage (traitements phonologiques, décodage, segmentation, traitements lexicaux et syntaxiques, etc.).

¹⁵ Les opérations dites de « haut niveau » renvoient par exemple à la régulation d'une tâche, à l'anticipation, au maintien de la cohérence du discours, à l'activation des connaissances préalables en mémoire à long terme.

de l'élève, qui se transformerait en motivation extrinsèque. En d'autres termes, la note tuerait le plaisir d'apprendre et favoriserait les stratégies d'apprentissage superficielles permettant de performer aux évaluations pour tous les types d'élèves (Archambault & Chouinard, 2006 ; Deci, Koestner & Ryan, 1999). D'ailleurs, il est intéressant de s'interroger sur ce qui pousse l'élève à travailler davantage dans un système compétitif : montrer sa performance ou éviter de montrer son incompetence ? Elliot et Church (1997) apportent des données empiriques intéressantes en rapport avec les buts d'accomplissement. La poursuite de buts de performance, évitement ou approche, serait corrélée négativement avec la motivation intrinsèque. Une étude de Pulfrey et al. (2011) va encore plus loin. Annoncer aux élèves, avant une évaluation, que celle-ci va être notée renforcerait les buts de performance-évitement et pas nécessairement les buts de performance-approche, qui sont plus favorables aux apprentissages. Si les notes "motivent" l'élève, ce n'est donc certainement pas dans le sens attendu, elles renforceraient des stratégies d'études superficielles et seraient par ailleurs source d'anxiété » (Calone & Lafontaine, 2023, p. 14).

« La théorie des buts d'accomplissement est née dans les années 1970-80 avec les travaux de Nicholls (1978, 1979, 1984) et Dweck (1975). [...] Dans le milieu des années 1990, plusieurs chercheurs, dont le principal est Elliot (1997, 1999 ; Elliot & Harackiewicz, 1994, 1996) ont proposé d'intégrer, dans la définition des buts, une distinction entre les buts d'approche et les buts d'évitement. C'est d'abord aux buts de performance que cette distinction a été appliquée : les buts de performance-approche (chercher à se montrer supérieur à autrui) sont alors différenciés des buts de performance-évitement (éviter de se montrer inférieur à autrui). Plus tard, la question de la valence a également été intégrée dans les buts de maîtrise (Elliot & McGregor, 2001 ; Pintrich, 2000), permettant de distinguer les buts de maîtrise-approche (chercher à maîtriser la tâche et progresser) des buts de maîtrise-évitement (éviter de ne pas maîtriser la tâche ou de régresser) » (Darnon, 2014, p. 11).

2. Le soi

La plupart des auteurs s'accordent pour reconnaître un caractère multidimensionnel au concept de soi, c'est-à-dire que tout individu possède un concept de soi global constitué de différentes conceptions de soi (Markus & Wurf, 1987).

Le concept de soi d'une personne (ensemble des éléments qui la définissent) est enraciné dans l'image que lui renvoient les autres individus. Dès l'enfance, entre 5 et 7 ans, les enfants développent des capacités cognitives permettant d'internaliser les évaluations d'autrui. Le processus de soi-miroir se met en place. Au fil du temps, l'entourage de l'enfant, puis l'entourage de l'adolescent, transmettent diverses connaissances de soi à la fois actuelles mais aussi futures (autrement dit, ce qu'il pourrait devenir) c'est-à-dire, des sois possibles (Martinot, 2004 ; 2008). On comprend alors que le soi d'un élève est fortement influencé par des facteurs sociaux, que les réactions des autres s'intériorisent progressivement et que **les évaluations transmises par des enseignants représentent l'une des sources de cette construction.**

Pour aller plus loin, plusieurs recherches ont montré qu'à l'école, **les représentations de soi sont susceptibles d'influencer les productions cognitives des élèves** : certains de ces travaux sont centrés sur les situations évaluatives, d'autres sur les outils d'évaluation comme la note, et d'autres sur les élèves.

a. Les situations évaluatives et le soi

Les connaissances des élèves à propos d'eux-mêmes, construites à partir de ce qui est vécu en classe, appartiennent au concept de soi scolaire ainsi qu'à un sentiment particulier appelé sentiment d'efficacité personnelle qui désigne la confiance que les individus ont dans leur capacité à réussir une tâche donnée. Or, de telles connaissances de soi peuvent entrer en résonance avec la situation, notamment la situation évaluative. Autrement dit, certains éléments de la situation peuvent agir comme des indices permettant de récupérer en mémoire les expériences plus ou moins couronnées de succès – et donc émotionnellement positives ou négatives – que l'élève a déjà vécues dans cette même situation. C'est de cette interaction entre contextes interne et externe que naîtraient des conséquences cognitives pesant plus ou moins sur les performances (Toczek & Souchal, 2017).

Par exemple, Monteil & Huguet (2002) ont révélé le poids du passé scolaire de collégiens en géométrie lors d'une situation évaluative qui était présentée selon deux conditions expérimentales. La première condition était présentée comme un exercice de géométrie et entrait en écho avec les niveaux scolaires des élèves et donc avec leur soi scolaire – connaissances de soi liées à une réussite en géométrie *versus* connaissances de soi liées à des difficultés en géométrie. La seconde condition présentait le même exercice dans un contexte indépendant du passé scolaire en géométrie des élèves et donc du soi scolaire de ces élèves (il s'agissait d'un simple dessin). Les résultats montrent que les représentations construites au fil du temps par les élèves à propos d'eux-mêmes sont déterminantes. En effet, on observe **une concordance entre la nature des connaissances de soi scolaire des élèves (en réussite *versus* en difficulté) et leurs performances réalisées en situation**. Autrement dit, lorsque les caractéristiques de la situation évaluative activent des connaissances sur soi liées à une réussite en géométrie, les performances sont élevées. Et inversement, lorsque la situation évaluative entre en écho avec des connaissances sur soi liées à des difficultés en géométrie, les performances sont significativement plus faibles. Alors que, si cette même activité cognitive est présentée dans une situation évaluative indépendante du passé et du soi scolaire des élèves - ici, le dessin -, quel que soit leur niveau scolaire en géométrie, les élèves réussissent de la même façon.

D'une manière générale, les chercheurs expliquent qu'une **situation évaluative potentiellement menaçante amènerait un individu à focaliser son attention sur le soi** et que ce mécanisme aurait pour effet de diminuer les ressources attentionnelles disponibles pour accomplir la tâche (Brunot, Huguet & Monteil, 2000).

b. Les outils d'évaluation et le soi

Certains outils, comme la note, peuvent-ils menacer le soi des élèves ? Certaines recherches sont en mesure de répondre par l'affirmative à cette question. Effectivement, comme nous l'avons évoqué précédemment au sujet de la motivation (voir Chapitre 2. B. 1.), **la note chiffrée sans commentaires, quand elle est utilisée dans une perspective normative, représente un feedback préjudiciable pour l'apprentissage, la motivation mais aussi le soi de l'élève**.

« Seule, la note contribue à attribuer l'échec à des causes stables et incontrôlables ("je suis nul en maths", "le prof ne m'aime pas") et ne soutient pas le besoin de compétence pour les élèves les plus faibles (Dresel & Hugwitz, 2008 ; Harks, Rakoczy, Hattie, Besser & Klieme, 2014). La note incite les élèves à formuler des attributions causales dysfonctionnelles qui engendrent l'abandon plus rapide d'une activité et peut conduire sur le long terme à la résignation apprise. Pour Viau (1994), cette

résignation n'est ni plus ni moins que la forme la plus extrême de perception d'incontrôlabilité de la tâche : l'élève pense que, quoi qu'il fasse, il n'arrivera à rien. Le développement de la résignation s'explique par le fait que l'élève est pris dans un momentum psychologique négatif (ou élan négatif), c'est-à-dire que les variables psychologiques telles que l'engagement dans la tâche ou la persistance sont négativement affectées par le fait que l'élève a vécu un ou des événements qui l'ont éloigné du but qu'il désirait atteindre (Briki & Gernigon, 2009). Ainsi, l'élève développe le sentiment qu'il a perdu le contrôle de sa réussite. C'est pourquoi, celui qui développe de l'incapacité acquise a souvent vécu une histoire scolaire parsemée d'échecs malgré les efforts fournis (Crahay, 2007). L'incapacité acquise est très résistante au changement, ceci pourrait s'expliquer par le fait que l'élève en situation de résignation développe très souvent un style attributionnel pessimiste (Paquet, 2009). Ce type d'élève aurait tendance à attribuer ses échecs à des dimensions internes, stables dans le temps et globales ou générales. En revanche, les succès sont vécus comme une "anomalie" dans leur parcours scolaire ("cette fois-ci, j'ai eu de la chance") et, à leurs yeux, ne peuvent s'expliquer que par des facteurs externes : influence d'une autre personne, chance... Par leur pauvreté, les notes peuvent participer au développement de la résignation chez les élèves en situation d'échec scolaire » (Calone & Lafontaine, 2023, p. 11).

c. Le soi des élèves « embarqué » dans leurs apprentissages

Pansu (2023) va plus loin en développant le rôle du soi dans les apprentissages. Il rappelle que

« s'il est une conclusion qui fait consensus aujourd'hui, c'est bien celle selon laquelle **se percevoir comme une personne ayant peu d'habiletés pour réaliser une tâche peut nuire à l'exécution des actions nécessaires à sa réalisation** et, in fine, affecter sa réussite, alors que se percevoir comme compétent pour réaliser une tâche peut soutenir la mise en œuvre des comportements nécessaires à l'exécution de celle-ci et, in fine, mener à sa réussite (Bandura, 1977, 1986 ; Harter, 1999 ; Marsh & Craven, 2006) » (Pansu, 2023, p. 73).

Or, l'évaluation qu'un élève fait de sa compétence n'est pas nécessairement liée à ses capacités, à ses performances réelles, mais bien au jugement ou à la perception que l'élève mobilise. Les psychologues sociaux étudient ce **biais d'auto-évaluation** de compétence scolaire en lien avec les performances.

« Par exemple, Vaillancourt et Bouffard (2009) ont montré que la présence d'un biais négatif d'auto-évaluation était associée à un traitement négatif des situations positives, à une exagération de l'importance des événements négatifs mineurs et à un sentiment de responsabilité accru envers ces événements. Les résultats indiquent aussi que se combinent à ce schéma des croyances relevant de forces incontrôlables. Par exemple, les élèves peuvent croire que leurs résultats sont dus à la chance ou au hasard et qu'ils ne peuvent rien faire pour en changer l'issue. Ils peuvent alors présenter des manifestations proches de ce que Seligman et collaborateurs ont appelé l'impuissance acquise, connue pour générer des déficits aux plans cognitif, affectif et motivationnel (cf. Miller, & Seligman, 1975). Au plan cognitif, ce déficit se traduit par un sentiment de non-contrôle (échec perçu comme inévitable) et interfère avec les processus de résolution de problèmes, de réflexion critique, et l'attention que l'élève peut porter à la tâche. Sur le plan émotionnel, il s'exprime dans un état dépressif exprimé par de l'apathie, des sentiments envers soi négatifs, de la passivité, de la colère et/ou de la rébellion. Au plan motivationnel, ce sentiment d'impuissance paralyse l'initiation des efforts et la persévérance à la tâche dès que survient un obstacle au profit de ruminations sur son état. Au fil des années, enfermés dans une telle spirale, les élèves

consolident leur système de réponses inappropriées et développent des attitudes peu favorables envers l'école, des comportements les poussant à s'écarter des apprentissages et pouvant induire chez eux un sentiment d'aliénation scolaire propice à un décrochage prématuré » (Pansu, 2023, p. 73).

Ce **biais négatif d'auto-évaluation de compétences scolaires** (illusion d'incompétence) **apparaît largement préjudiciable, aussi bien du point de vue des apprentissages que du point de vue du bien-être psychologique.** Il est souvent associé à l'adoption de **stratégies de protection de soi aux effets délétères pour la motivation de l'élève** (Pansu, 2023).

En revanche, selon des travaux récents, présenter une auto-évaluation biaisée positivement de sa compétence semble accroître la motivation de l'élève tout en facilitant l'autorégulation de ses apprentissages.

Les travaux de Martinot (2001) s'inscrivent dans cette perspective en montrant que **favoriser le développement de conceptions de soi de réussite se révèle bénéfique pour l'élève.** En effet, de telles conceptions favorisent la construction d'un sentiment d'auto-efficacité élevé qui, par voie de conséquence, est susceptible d'accentuer les efforts et la persistance de son comportement en cas de besoin (Bandura, 1982 ; 1997).

En complément, Pansu (2023) insiste sur l'importance de considérer les deux facettes du soi emboîtées hiérarchiquement :

« le biais général d'auto-évaluation de "sa" compétence scolaire, en englobant le rapport à l'école dans son intégralité, est d'un ordre supérieur, alors que le biais spécifique qui se focalise sur un apprentissage particulier (français, mathématiques, géographie, etc.) est d'un ordre inférieur » (Pansu, 2023, p. 75).

Les recherches sur ce thème conduisent Pansu (2023) à proposer des repères pour l'action en classe, des « gestes professionnels » susceptibles de prévenir à la fois une auto-évaluation biaisée négativement de ses compétences et de restaurer la confiance des élèves en cas de besoin (Pansu et al., 2021).

« Finalement, les résultats des études présentés invitent à réfléchir à différents leviers pour favoriser l'apprentissage des élèves : l'outillage dont les enseignants ont besoin (1) pour être en capacité de repérer les élèves présentant des attributions dysfonctionnelles pour leurs apprentissages ; (2) pour centrer leurs feedbacks sur la tâche et non sur les capacités intrinsèques de l'élève (ce qui reviendrait à renforcer chez lui une vision stable et immuable de l'intelligence) ; (3) pour éviter un climat de trop grande compétitivité et référer l'élève à ses progrès plutôt qu'aux performances des autres (en faisant en sorte de ne pas lui imposer, même involontairement, de comparaison sociale forcée) ; (4) pour procéder par petites étapes, donner des buts proches en ayant des attentes explicites et claires, de manière à permettre aux élèves de mettre en place des stratégies d'autorégulation adaptées pour atteindre leurs objectifs immédiats et futurs » (Pansu, 2023, p. 77).

3. Effets des situations évaluatives sur les performances

Tout un pan de la littérature scientifique montre comment **la situation évaluative dans laquelle un élève met en œuvre ses compétences est déterminante**. Ces travaux révèlent que les contextes évaluatifs sont en mesure de réguler les performances de certains élèves, c'est-à-dire qu'ils sont susceptibles de les faire réussir dans certains cas et de les faire échouer dans d'autres. Plusieurs recherches montrent que le contexte de traitement de la tâche est déterminant pour l'activité cognitive de l'élève. Ces recherches ont identifié des mécanismes psychosociaux responsables des différences de performances. Elles soulignent l'intérêt de chercher à saisir les dynamiques psychosociales dans leur dimension temporelle et historique. Une telle perspective montre l'importance d'étudier l'impact cognitif des catégorisations et des conceptions des différents acteurs inscrits dans des situations d'enseignement et d'apprentissage : catégorisations sociales véhiculées par les contextes, catégorisations liées à l'histoire des sujets (échec ou réussite scolaire), catégorisations liées au genre, à l'appartenance socioéconomique, aux conceptions de l'intelligence des différents acteurs de la situation ou bien aux buts d'apprentissage des élèves.

En d'autres termes, ces travaux révèlent que la mobilisation de normes et valeurs intériorisées par les élèves ou par les enseignants au fil de leur histoire interfère avec le traitement cognitif de la situation d'évaluation dans laquelle ils sont insérés. Afin d'illustrer cette perspective théorique, nous allons exposer les effets de plusieurs contextes évaluatifs.

a. L'effet de menace du stéréotype

Le phénomène de menace du stéréotype a été mis en lumière et nommé ainsi pour la première fois en 1995 par Steele et Aronson. Ces auteurs s'intéressaient à la population noire américaine et s'étonnaient de constater les moindres performances systématiquement enregistrées par les étudiants appartenant à cette population comparativement à celles obtenues par les étudiants blancs. Plus précisément, ces auteurs se sont intéressés aux étudiants noirs d'universités prestigieuses, ayant donc atteint un niveau d'excellence académique indéniable et ne pouvant donc pas être considérés comme désidentifiés de ce domaine académique ou comme possédant des connaissances de soi négatives dans le contexte scolaire.

Dans l'étude réalisée, Steele et Aronson proposent à des étudiants – noirs et blancs – de l'université de Stanford de réaliser un test verbal dont seul l'habillage diffère entre les deux groupes expérimentaux. Tandis qu'une première moitié des participants croit réaliser un test diagnostique des capacités intellectuelles, l'autre moitié est avertie que cette tâche est un simple exercice de laboratoire. L'hypothèse des chercheurs est que le contexte externe – la tâche présentée comme diagnostique des capacités intellectuelles – active la connaissance du stéréotype d'infériorité intellectuelle des participants noirs et génère alors des conséquences cognitives conduisant à la baisse de leurs performances uniquement dans cette situation. Les résultats obtenus confirment cette hypothèse : les étudiants noirs de la condition « test d'intelligence » enregistrent de moins bonnes performances que les étudiants blancs de cette même condition, mais également que tous les participants de la condition « test de laboratoire ». L'explication donnée par les chercheurs est que les individus appartenant à des groupes sociaux stigmatisés par un stéréotype d'infériorité ressentent, lorsqu'ils sont placés dans une situation où ils sont susceptibles de confirmer à leurs yeux ou aux yeux des autres le stéréotype, un inconfort psychologique qui se traduit par une perturbation de leurs fonctions cognitives et donc, entre autres, par une diminution de leurs performances (Steele, 1997 ; Steele & Aronson, 1995). En fait, ce qui semble menacer l'individu stigmatisé est le

fait que l'obtention d'une mauvaise performance dans cette situation stéréotypée pourrait être expliquée uniquement en faisant appel au stéréotype négatif. L'individu ne se sent donc plus pris en compte comme une personne avec des caractéristiques propres, mais comme un exemplaire de son groupe d'appartenance. Désert, Croizet & Leyens (2002) parlent à ce titre de *dé-individuation*.

Cet effet, initialement obtenu auprès d'étudiants noirs stigmatisés sur le domaine de l'intelligence, a également été depuis étudié auprès de nombreux groupes sociaux. En fait, il semble que la menace du stéréotype puisse atteindre tout membre d'un groupe stigmatisé par un stéréotype négatif. Ainsi, le groupe des femmes et le stéréotype d'infériorité intellectuelle dans le domaine scientifique qui le stigmatise ont-été tout particulièrement étudiés. Les résultats montrent généralement que les femmes placées dans une situation pouvant être interprétée à la lumière de ce stéréotype obtiennent de moins bonnes performances que les hommes et que les femmes et les hommes placés dans une condition plus neutre.

Les individus issus de groupes sociaux de faible statut socioéconomique sont également sensibles à cet effet de menace du stéréotype. Croizet et Claire (1998) montrent ainsi que les performances de ces étudiants à un test verbal sont moindres que celles des autres participants de haut statut lorsqu'ils considèrent ce test comme diagnostic des capacités verbales et que leur appartenance groupale a été rendue saillante par des questions concernant la profession et le niveau d'études de leurs parents (voir aussi Spencer & Castano, 2007 ; Harrison, Stevens, Monty & Coakley, 2006 ; Désert, Préaux & Jund, 2009).

Les groupes constitués sur la base des filières d'études suivies sont également la cible de stéréotypes plus ou moins négatifs et donc c'est tout logiquement que leurs membres peuvent également être touchés par cet effet de menace. Ainsi, Croizet, Dutrévis & Désert (2002) montrent que les étudiants titulaires d'un baccalauréat technologique sont menacés face à un test verbal. Ils y obtiennent en effet de moins bonnes performances lorsqu'on le leur présente comme diagnostique des capacités intellectuelles que lorsqu'ils pensent que ce test concerne le fonctionnement mnésique. Comme dans les autres études, les performances des participants non stigmatisés – c'est-à-dire les titulaires d'un baccalauréat général – ne sont pas impactées par le contexte de passation. Ces résultats ont également été répliqués chez des jeunes enfants de cours préparatoire (Désert *et al.*, 2009).

En somme, ce phénomène de **menace du stéréotype peut être défini comme la crainte de confirmer ou d'être vu comme confirmant un stéréotype négatif propre à son groupe d'appartenance** – et donc d'être réduit à ce trait négatif. Il entraîne une pression évaluative sur l'individu et cette pression peut être suffisamment forte pour perturber le fonctionnement cognitif « normal » de la personne.

Ainsi, comme le souligne Goudeau (2023),

« cette "menace du stéréotype" engendre du stress, mais également des émotions et des pensées négatives, ce qui peut perturber la performance lors de la réalisation d'une tâche (Schmader et al., 2008). En effet, elle impose à l'élève de faire deux choses : la tâche à réaliser, mais aussi gérer le stress qu'entraîne la peur d'être perçu comme incompetent. Cela s'avère délétère pour les performances intellectuelles car une partie de l'attention n'est pas allouée à la tâche » (Goudeau, 2023, p. 78).

b. La comparaison sociale

La classe est un lieu de multiples évaluations, positives ou négatives (Huguet, 2006). Évaluations à partir desquelles les élèves se comparent les uns avec les autres. Or, ces comparaisons ont des effets aussi bien sur les états émotionnels et affectifs que sur les fonctionnements cognitifs, et donc sur les performances des apprenants. La classe réunit les conditions d'émergence de représentations de soi et de comportements qui, pour une part au moins, sont le produit résumé de ce qu'une culture choisit d'honorer, de récompenser et de cultiver. Dans notre culture, et plus généralement dans les cultures nord-américaines et d'Europe de l'Ouest, l'emprise de la compétition est telle qu'elle contraint les individus à se comparer entre eux en permanence (Huguet, 2006).

Il y a plusieurs décennies, des expériences publiées par Monteil & Huguet (2002), mettaient déjà en évidence les effets délétères de la comparaison sociale en classe. Ces études opérationnalisait la comparaison sociale par le biais d'une pratique pédagogique singulière : le fait de clamer à voix haute les notes lors de la remise d'un devoir, et ce, avant de commencer un nouveau cours. Sans surprise, les résultats mettent en évidence des effets de la comparaison sociale sur les performances des élèves à l'issue du cours.

D'une manière générale, les recherches en psychologie sociale ont mis en évidence qu'en classe, constater qu'un camarade est en train de mieux réussir que soi peut représenter une menace pour l'image de soi. Plus précisément, **la réussite des autres peut atteindre le besoin de maintenir une bonne image de soi, ce qui génère du stress et un cortège d'émotions et de pensées négatives très coûteuses en ressources attentionnelles** (Goudeau, 2023). Goudeau poursuit son argumentation en expliquant que les plus ou moins grandes réussites des élèves sont souvent perçues comme le reflet de différences individuelles relativement stables d'intelligence ou de motivation (Croizet & Claire, 1998). Or, croire que l'on est moins intelligent que les autres, et que l'intelligence est une donnée stable représente une réelle menace pour soi (Normand & Croizet, 2013). Cette interprétation est renforcée par le fait que la classe est supposée être un environnement juste.

Goudeau (2023), dans sa note, aborde de nombreuses situations évaluatives concrètes aux effets non neutres. Il évoque différents moments d'échanges collectifs en classe qui peuvent avoir lieu quotidiennement, comme par exemple l'enseignant qui pose des questions à ses élèves, les feedbacks qu'il leur adresse, la réaction des élèves, le fait qu'ils lèvent la main ou pas. Ce chercheur explique qu'en dépit de leur caractère informel, de telles situations invitent les élèves à formuler une bonne ou mauvaise réponse, à être interrogés ou non, et par voie de conséquence sont susceptibles de donner lieu à des inférences sur de supposées capacités de tel ou tel élève.

« Dans une série d'études, Goudeau et Croizet (2017) ont mis en évidence le fait que les situations de comparaison sociale imposées par le contexte de classe (par ex., un élève qui lève la main pour dire qu'il a terminé un exercice), jouaient un rôle dans l'amplification des inégalités scolaires. Ces comparaisons délétères résultent du fait que les situations scolaires mettent en scène les écarts de réussite entre élèves d'une manière qui ignore l'inégale familiarité vis-à-vis de l'arbitraire culturel scolaire, ce qui oblige les élèves à interpréter ces (dés)avantages comme le reflet de différences de capacités intellectuelles, interprétation qui peut interférer avec le fonctionnement cognitif des élèves désavantagés » (Goudeau, 2023, p. 80).

En revanche, une autre étude montre qu'il est possible de neutraliser de tels effets délétères en conduisant les élèves à interpréter les différences de réussite comme la conséquence de différences d'entraînement plutôt que de différences de réelles capacités intrinsèques aux élèves (Goudeau & Croizet, 2017). Ces résultats sont cohérents avec les travaux plus anciens de Lockwood et Kunda (1997) qui montrent que **voir réussir un camarade mieux que soi peut être motivant à la condition de penser qu'il est possible d'atteindre ce niveau de réussite.**

Dans la même veine,

« **plusieurs recherches ont été menées pour tenter de réduire ces effets de "menace" de l'image de soi.** Une façon de réduire ces phénomènes est d'agir sur la métacognition, c'est-à-dire la manière dont les individus pensent et se représentent le fonctionnement cognitif. Ainsi, amener des étudiants à penser que le stress est bon pour la performance ou encore expliquer à des collégiens que l'intelligence est un attribut malléable qui se développe plutôt qu'une capacité fixe, permet de contrecarrer les effets de menace du stéréotype (Johns, Inzlicht & Schmader, 2008 ; Good, Aronson & Inzlicht, 2003). » (Goudeau, 2023, pp. 82-83).

c. La compétition et la poursuite de buts de performance

« De nombreuses études montrent que pratiquer la compétition ou donner des feedbacks de comparaison sociale (via les notes notamment) pourrait se révéler néfaste à l'apprentissage et **tendrait à diminuer les performances** (Ames, 1992 ; Butler, 1987 ; Deci & Ryan, 2000 ; Galand, 2006). De plus, elles pourraient conduire à des comportements antisociaux – la tricherie par exemple – (Dupeyrat, Escribe & Mariné, 2006 ; Perrenoud, 1999 ; Pulfrey, 2011), et pourraient renforcer la peur de l'échec et les différences entre élèves performants et élèves faibles (Butler & Nissan, 1986 ; Kim, Lee, Chung & Bong, 2010). [...]

Un **système compétitif** favoriserait chez l'élève la poursuite de **buts de performance** au lieu de buts de maîtrise (Butler, 2006 ; Darnon et al., 2011). Or les types de buts qu'un élève va poursuivre influencent sa conception de l'apprentissage, sa motivation, le type de stratégies qu'il va mettre en place, l'autorégulation et in fine l'apprentissage. C'est ce qui expliquerait ce phénomène de "fin de trimestre" où les élèves se bousculent pour connaître leur "moyenne", mais ne se pressent pas pour comprendre leurs erreurs et comment y remédier.

En somme, le type d'évaluation annoncé aux individus est susceptible de rendre plus ou moins saillants certains types de buts – buts de performance ou buts de maîtrise. Il n'est donc pas étonnant qu'en fonction des types de comparaisons suggérés par les résultats des évaluations, les élèves puissent modifier leurs buts de réussite. C'est par ce biais que certains chercheurs (Souchal et al., 2014) ont testé des contextes de classe qui, tout en restant évaluatifs avec l'utilisation de la note, pouvaient présenter des caractéristiques non menaçantes pour les individus habituellement en échec.

Plus précisément, leur étude s'est déroulée à l'occasion d'un cours de physique-chimie face à des lycéens fréquentant des classes de seconde. Elle incluait trois conditions. Tandis que les membres du premier groupe ne pensaient pas être évalués, les élèves du deuxième et du troisième groupes étaient prévenus qu'ils passeraient une évaluation. Entre ces deux groupes, c'est l'orientation de l'évaluation qui était manipulée : pour les uns, l'évaluation était orientée vers la performance et la comparaison entre élèves ; pour les autres elle était orientée vers la maîtrise car on expliquait que l'évaluation aiderait les élèves dans leur apprentissage. Dans les deux cas, les élèves pensaient recevoir une note qui compterait dans leur moyenne trimestrielle de physique-chimie.

Conformément à leurs hypothèses, les chercheurs montrent que la situation évaluative orientée vers la maîtrise est la seule des trois conditions testées qui permet à l'ensemble des participants de préserver leurs performances alors même que les élèves pensaient recevoir une note comptant pour leur moyenne. **Le pouvoir des contextes évaluatifs orientés vers la maîtrise semble donc être plus important que la simple utilisation de l'outil évaluatif "note" »** (Calone & Lafontaine, 2023, p. 12).

Chapitre 3 : Évaluation, différenciation, équité

A. Modèles de justice et conceptions de l'égalité

La question des inégalités face à l'évaluation ou engendrées par les situations évaluatives est une question centrale. Elle peut se révéler encore plus cruciale quand il s'agit d'évaluations à enjeux élevés pour le parcours de l'élève (évaluations sommatives ou certificatives sur lesquelles sont fondées des décisions de passage ou d'orientation). Néanmoins, la question du type d'égalité ou de l'équité que l'on vise en pratiquant une évaluation soutien d'apprentissage est tout aussi essentielle.

Il se trouverait peu d'acteurs de l'école pour contester que les élèves doivent être égaux face à l'évaluation ou que l'évaluation doit être juste ou équitable. Mais de quelle égalité parle-t-on exactement ? Qu'est-ce qui est juste en matière d'éducation ? Existe-t-il des inégalités « justes » ou « acceptables » ? Plusieurs modèles ou conceptions développés par des chercheurs en sciences de l'éducation ou des sociologues coexistent ; cela vaut la peine de les clarifier à l'entame de ce chapitre.

Quelles qu'elles soient, les différentes conceptions de l'égalité se rattachent à un modèle de la justice (Rawls, 1987 ; Walzer, 2013). Les auteurs s'accordent (Crahay & Lafontaine, 2012) pour distinguer deux grands modèles de la justice :

- Une justice *méritocratique* : selon cette conception, il est considéré comme juste que chacun reçoive, soit rémunéré ou noté selon ses mérites (les élèves qui font mieux ou plus vite sont mieux notés, valorisés, ceux qui ont des difficultés sont sanctionnés, etc.).
- Une justice *corrective* ou *compensatoire* : selon cette conception, il est considéré comme juste de tenter de corriger les inégalités de départ, d'éviter de les accroître ou tenter de les réduire, de donner plus aux moins nantis ou plus fragiles, etc.¹⁶

À ces deux conceptions de la justice se rattachent trois modèles d'égalité¹⁷ : le modèle d'égalité des chances, le modèle d'égalité de traitement et le modèle d'égalité des acquis. Nombre de pratiques et de politiques éducatives peuvent être lues ou analysées au prisme de ces modèles d'égalité.

1. Le modèle de l'égalité des chances

Ce modèle est le plus répandu et est communément admis, sauf dans les sociétés non démocratiques ou économiquement peu développées¹⁸. Il a joué un rôle central dans la généralisation de la scolarité obligatoire. Il se rattache au modèle de justice méritocratique et est fortement axé sur l'égalité des chances d'accès (à différents parcours). Selon ce modèle, à *aptitudes*

¹⁶ Ces deux modèles de la justice sont aussi à rapprocher des deux fonctions principales de l'école, telles que les présente Butera dans sa note (2023). La justice méritocratique est étroitement liée à la fonction de sélection, tandis que la justice corrective est davantage liée à la fonction de formation.

¹⁷ D'autres auteurs (Dubet, 2004 ; Felouzis, 2014) ont proposé des modèles légèrement différents, mais qui s'articulent aussi étroitement à des principes de justice. À ce propos, Dubet écrit : « je suis convaincu que l'école ne peut être définie que par une articulation prudente entre plusieurs principes de justice, par une combinaison relativement complexe dans laquelle chaque principe vise à corriger les effets destructeurs des autres » (Dubet, 2004, p. 10).

¹⁸ Ainsi, en Afghanistan et dans plusieurs pays d'Afrique subsaharienne, les filles ont un accès plus limité à l'enseignement primaire que les garçons (voir <https://uis.unesco.org/fr/news/consultez-les-dernieres-donnees-sur-leducation-des-filles-et-des-femmes>).

égales, chaque élève a le droit d'accéder aux formations les plus longues, les plus prestigieuses¹⁹, etc. L'origine sociale ou ethnique, le genre, les revenus, etc. ne peuvent constituer des obstacles dans l'accès à l'éducation. Toutefois, pour pouvoir accéder aux cursus les plus avancés, il faut être « méritant » (être bon élève, « en vouloir », faire des efforts, etc.). **La méritocratie est une composante essentielle du modèle de l'égalité des chances.** Un exemple de ce modèle d'égalité des chances en France est l'existence d'un quota réservé aux jeunes d'origine défavorisée particulièrement « doués » dans les Grandes écoles.

Une autre composante de ce modèle est l'idéologie du don ou une conception fixiste de l'intelligence : chacun doit pouvoir se développer sans entrave en fonction de ses potentialités « naturelles », pouvoir accéder à ce pour quoi il est fait ou prédestiné, aller au bout de ses possibilités. Toutefois, selon cette conception, les capacités sont différentes ou inégales dès le départ, certains élèves sont « doués » pour les études longues, alors que d'autres ont « l'intelligence de la main ». La prégnance de ce modèle est particulièrement visible dans les débats récurrents qui opposent les partisans du collège unique à ceux qui prônent une orientation précoce vers des filières techniques et professionnelles.

2. Le modèle de l'égalité de traitement

En plus de l'égalité des chances, certains prônent une égalité de traitement. Les deux conceptions ne sont pas contradictoires : on peut défendre l'égalité de traitement jusqu'à un certain âge (fin de l'enseignement primaire ou fin du collège), puis l'égalité des chances après.

Les défenseurs de ce modèle se sont surtout battus pour imposer le modèle de l'école unique dans le primaire, puis dans le secondaire (collège unique en France, tronc commun, école unique dans les pays nordiques). La généralisation d'évaluations externes nationales obéit aussi à ce type de conception : le principe est que chaque élève soit mis sur un pied d'égalité pour les évaluations les plus décisives comme le brevet ou le bac.

Réaliser cet idéal d'égalité de traitement suppose une homogénéisation, une standardisation, une politique assez centralisée de gestion, de contrôle et de régulation. Ce modèle est assez influent en France.

3. L'égalité des acquis ou de résultats

Ce troisième modèle diffère à maints égards des deux précédents et est incompatible avec l'égalité de traitement. Il est plus récent et, à la différence des deux précédents, trouve ses fondements dans des recherches en sciences de l'éducation datant des années 1960 (Bloom *et al.*, 1971 ; Carroll, 1963). Ces deux pères fondateurs de l'évaluation formative et de la pédagogie de maîtrise ont posé le postulat d'éducabilité : tous les élèves sont capables de maîtriser des objectifs d'apprentissage fixés pour autant qu'on leur accorde le temps nécessaire à cet effet et un enseignement de qualité. En termes d'apprentissage, il s'agit d'un bouleversement conceptuel majeur. Dans le modèle de

¹⁹ Aujourd'hui dans les pays économiquement avancés, l'enjeu est un accès aux parcours dans l'enseignement supérieur, mais dans d'autres contextes et à une époque pas si lointaine, l'enjeu était l'accès à l'instruction tout simplement, ensuite l'accès aux études générales dans le secondaire ou à certaines filières.

Carroll, l'aptitude de l'élève ne prédit pas le niveau qu'il peut atteindre (comme dans le modèle d'égalité des chances), mais le temps qui sera nécessaire pour parvenir au but. Selon ce postulat, il n'existe pas de don ou d'aptitude naturelle qui ne soit modifiable, pas d'intelligence qui ne puisse se développer. Le modèle d'égalité des acquis s'appuie sur une conception incrémentale de l'intelligence, alors que le modèle d'égalité des chances est imprégné d'une conception fixiste de l'intelligence²⁰.

Modèle de Carroll (1963)

$$\text{Degré d'apprentissage} = \frac{\text{Temps consacré}}{\text{Temps nécessaire}}$$

Dans un enseignement traditionnel, le temps consacré est à peu près le même pour tous les élèves d'une classe ; comme le temps nécessaire pour acquérir un savoir ou une compétence est variable selon les individus, les acquis des élèves varient en fin d'apprentissage – ce qui est considéré comme « normal » dans un modèle d'égalité des chances. Chacun a eu sa chance, mais certains sont mieux équipés pour s'en saisir. L'inégalité de résultats n'apparaît donc pas comme injuste.

Dans la pédagogie de maîtrise, on fixe un objectif/degré d'apprentissage constant que tous doivent atteindre, et on fait varier le temps pour y parvenir. C'est l'égalité de résultats qui est visée.

Le modèle de Carroll est à l'origine des pratiques pédagogiques de différenciation. Les pratiques d'évaluation formative, ainsi que les feedbacks, sont quant à eux les principaux instruments de cette différenciation : ils aident à poser un diagnostic dans l'optique de remédier, comme on disait auparavant, de différencier comme on le dit aujourd'hui.

Le modèle d'égalité des acquis se rattache à une conception correctrice de la justice et est aussi au fondement de toutes les politiques compensatoires. Celles-ci, par définition, consistent à rompre avec une stricte égalité de traitement et à donner davantage, assurer de meilleures conditions ou occasions d'apprendre à ceux qui ont moins au départ, tout cela dans le but de les amener au même résultat, la maîtrise des savoirs essentiels ou des socles communs.

B. Évaluer et différencier l'enseignement pour amener un maximum d'élèves à la réussite

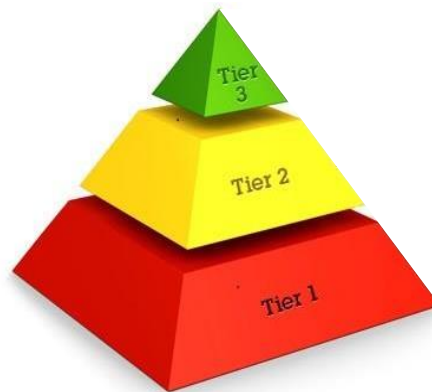
Jusqu'ici, dans cette note, la question de l'évaluation comme soutien à l'apprentissage a été abordée sans l'articuler explicitement à un modèle d'égalité ou une visée particulière en termes de résultats. Certes, la notion de pédagogie de maîtrise a été évoquée à quelques reprises, mais sans la développer. Des modèles plus récents et plus systémiques, encore peu connus en France, méritent que l'on s'y attarde. L'un de ces modèles est développé dans la note de Coertjens (2023) à laquelle l'essentiel de ce qui suit est emprunté.

Au début des années 2000, un modèle a été développé qui combine évaluation des élèves (suivi des progrès) et soutien différencié pour amener un maximum d'élèves à la réussite. Ce modèle, d'abord

²⁰ Pour plus d'informations sur ces deux conceptions de l'intelligence, voir la note de Fenouillet (2023).

dénommé « modèle de la réponse à l'intervention », est aujourd'hui appelé « système de soutien à différents niveaux²¹ ».

Figure 5 : Le modèle MTSS



Tier signifie « niveau ». Repris de <https://www.fargo.k12.nd.us/Page/2059> (Coertjens, 2023, p. 85)

Le « système de soutien à différents niveaux » comporte trois niveaux d'intervention. Le premier niveau correspond à l'enseignement de base dispensé à toute la classe ; pour beaucoup d'élèves, voire pour tous, celui-ci est suffisant pour faire acquérir un nouveau contenu. Pour un certain nombre d'élèves toutefois, cela ne suffira pas et un enseignement plus ciblé sera nécessaire (réexpliquer, aborder la matière autrement, faire plus d'exercices). C'est le niveau 2 du modèle : cette intervention se fait généralement en petits groupes. Enfin, pour une minorité d'élèves, une intervention encore plus soutenue sera nécessaire : cette dernière correspond au niveau 3 et se fait généralement de manière individualisée²². Chaque classe étant différente, la proportion d'élèves nécessitant des interventions de niveaux 2 et 3 est variable, il n'y a pas de norme en la matière. Comme ce type d'intervention nécessite un suivi en petits groupes (niveau 2) ou individualisé (niveau 3), il faudra faire appel à d'autres intervenants que le titulaire de la classe pour l'aider à organiser les interventions spécifiques. Comme le précise Coertjens dans sa note (2023), le modèle MTSS nécessite un travail d'équipe : l'aide d'autres enseignants, voire le recours à des spécialistes comme des orthophonistes pour le niveau 3.

Comment s'organise l'évaluation dans le modèle MTSS et quel est son rôle ?

Tout d'abord, pour déterminer quels élèves ont besoin d'un soutien de niveau 2, le niveau de compétence de tous les élèves est mesuré plusieurs fois par an en utilisant un « *curriculum-based test* ». Ce type de test n'est pas élaboré par les enseignants, mais par des experts externes. Ce sont des tests standardisés validés qui permettent d'évaluer de manière fine le niveau des élèves. Comme le souligne Coertjens,

« ces tests ont une visée purement formative : certes, ils permettent de faire un état des acquis de chaque élève (ce qui peut faire penser à une évaluation sommative), mais les résultats ne sont

²¹ « *Multi-Tiered System of Support* » (MTSS) dans la littérature anglo-saxonne.

²² Pour plus de détails sur l'organisation concrète de ces différents niveaux d'intervention, nous renvoyons le lecteur intéressé à la note de Coertjens (2023).

utilisés que pour déterminer le meilleur accompagnement possible pour chaque élève » (Coertjens, 2023, p. 87)

Quand des élèves ont été identifiés comme nécessitant un soutien de niveau 2, les évaluations se font plus régulières et plus spécifiques, portent sur de plus petites unités d'apprentissage et s'articulent plus étroitement à ce qui a été enseigné ou travaillé avec les élèves. Il s'agit d'évaluations formatives plus classiques qui viennent soutenir l'apprentissage.

Que dit la littérature de recherche sur les effets du modèle MTSS ?

Cinq méta-analyses ont été publiées sur le sujet entre 2017 et 2021. Parmi celles-ci, trois portent spécifiquement sur la mise en place des interventions à trois niveaux. Les méta-analyses portant sur l'efficacité de la mise en place d'un soutien de niveau 2 montrent que celle-ci améliore les résultats en lecture dans l'enseignement primaire avec des tailles d'effet moyennes faibles à modérées (respectivement 0,24, 0,36 et 0,38). Une seule méta-analyse a examiné l'efficacité du soutien individualisé de niveau 3. L'effet moyen est 0,41. Ce soutien est efficace « pour aider à apprendre à lire aux élèves en primaire pour lesquels les niveaux 1 et 2 n'ont pas suffi » (Coertjens, 2023, p. 95).

Même si des recherches supplémentaires sont nécessaires, les premières méta-analyses consacrées à l'efficacité du modèle MTSS donnent des résultats prometteurs. L'originalité de ce modèle est **qu'il articule étroitement des évaluations diagnostiques visant à repérer les élèves « à risque »** pour lesquels l'enseignement de base ne suffit pas, **avec la mise en place d'interventions ciblées** dites de niveau 2 (menées en petits groupes) et de niveau 3 (menées individuellement). La visée explicite du MTSS est de conduire un maximum d'élèves à la maîtrise des savoirs et connaissances fondamentaux, dans une optique de pédagogie de la maîtrise conforme au modèle de Carroll. C'est bien entendu une approche emblématique du modèle de l'égalité des acquis : les élèves en difficulté ou à risque reçoivent un soutien supplémentaire et différencié en fonction des besoins diagnostiqués par des évaluations régulières.

C. Évaluer différemment des élèves différents²³

La question d'une évaluation différente pour les élèves différents abordée par Jury (2023) dans sa note renvoie aux enjeux et tensions qui traversent le champ de l'évaluation et aux différentes conceptions de l'égalité et de la justice présentées au début de ce chapitre, en d'autres termes à des débats ou conflits de valeurs.

Ainsi, une question qui peut de prime abord apparaître comme technique – quels aménagements apporter à l'évaluation pour permettre aux élèves porteurs d'un handicap avéré ou en difficulté d'apprentissage de pouvoir montrer ce dont ils sont capables ? – se mue rapidement en débat autour de ce qui est *juste* ou *équitable* en matière d'évaluation.

Les rares²⁴ travaux de recherche sur les effets des aménagements (donner plus de temps, lire les consignes à voix haute) montrent que ceux-ci peuvent avoir des effets positifs, soit particulièrement

²³ Cette section emprunte son titre à la note de Jury (2023).

²⁴ Ces travaux sont particulièrement rares pour les niveaux primaire et secondaire. Ils sont un peu plus nombreux dans l'enseignement supérieur.

pour les élèves en situation de handicap, soit pour tous les élèves. Les études sont non seulement peu nombreuses, mais elles ne débouchent pas sur des résultats consensuels.

Toutefois, la vraie difficulté à la mise en place d'adaptations n'est pas là. Comme le souligne Jury (2023),

« le fonctionnement du système, à travers sa mission de sélection et la promotion de la méritocratie, peut venir implicitement créer des barrières dans la mise en place d'adaptations pédagogiques pour l'évaluation. En effet, cela pourrait laisser à penser que l'on combat une injustice (la situation de handicap) en en installant une autre (vis-à-vis des pairs ordinaires). Aussi s'agit-il, pour les enseignants, de fournir des adaptations aux élèves en situation de handicap afin de leur permettre d'exprimer pleinement leurs compétences tout en ne créant pas de rupture au principe d'égalité avec les autres élèves » (Jury, 2023, p. 105).

Pour dépasser ce dilemme entre deux conceptions antagonistes de la justice et de l'égalité, Jury invite, en conclusion de sa note, à une approche plus universelle pour soutenir une évaluation vraiment inclusive :

« l'incompatibilité entre sélection et évaluations adaptées nous semble reposer principalement sur une conception normative des évaluations (c'est-à-dire, utile pour comparer les niveaux des élèves). Une conception beaucoup plus formative de celles-ci, centrée notamment sur les progrès des élèves, rendrait la tension évoquée ici beaucoup plus faible » (Jury, 2023, p. 105).

Ainsi, plutôt que de procéder à des adaptations individuelles, « à la carte », qui peuvent aider, mais aussi se révéler stigmatisantes pour les élèves qui en bénéficient, c'est tout l'édifice de l'évaluation qu'il serait souhaitable de repenser, dans la perspective qui est précisément au cœur de cette conférence de consensus du Cnesco :

« Une réflexion plus en amont sur les barrières potentielles lors d'une évaluation et une réponse plus collective que singulière (par ex., si je lis les consignes à haute voix pour un élève, pourquoi ne pas les lire pour toutes et tous ?) s'accorderait probablement davantage avec l'idée d'une école dans laquelle chacun, quelle que soit la situation de handicap qu'il rencontre, ses difficultés passagères ou persistantes ou son environnement de vie, a, de droit, sa place » (Jury, 2023, p. 106).

D. Changer les situations d'évaluation et les perceptions sources d'inégalités

Dans ce chapitre, nous avons jusqu'ici mis en avant des exemples de pratiques ou de politiques de *discrimination positive*, consistant à traiter différemment les élèves les moins avantagés de la société, en respectant des principes d'équité²⁵ et en visant la réussite de tous les élèves.

Il existe cependant d'autres points de vigilance, déjà évoqués dans le chapitre 2, mais sur lesquels nous souhaitons revenir de manière synthétique ici. Nombre de situations d'évaluation courantes dans le quotidien de la classe peuvent ainsi engendrer des inégalités selon certaines caractéristiques

²⁵ L'équité est intrinsèquement liée à la notion de justice. L'équité n'est pas l'égalité, c'est une égalité ou une inégalité considérée comme « juste » au vu d'un certain nombre de principes. Prenons un exemple simple, celui d'une famille de deux enfants âgés de 6 et de 16 ans. Quelle somme d'argent de poche leur donner par semaine ? Si le plus jeune reçoit 10 euros et l'aîné 25 euros, ils ne sont pas traités de manière égale, mais de manière équitable. La somme reçue tient compte de la situation des deux enfants ; cette inégalité de traitement peut être considérée comme juste.

des élèves (l'origine sociale ou ethnique, le genre, le fait d'être en retard scolaire, le fait d'être en surpoids au cours d'EPS, etc.²⁶). Les nombreuses études menées en psychologie sociale ont attiré l'attention sur ce type de phénomènes et ont aussi dégagé des pistes pour y sensibiliser les enseignants et développer des pratiques qui favorisent une réduction de ces inégalités plutôt que leur accroissement.

1. L'évaluation sommative amplifie les inégalités, au détriment des élèves d'origine défavorisée

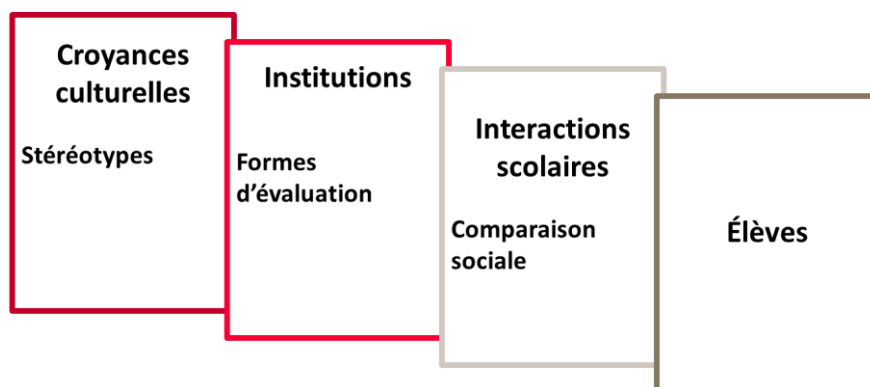
Une étude citée par Butera dans sa note (2023) montre que les évaluateurs d'une dictée mis en situation d'évaluation normative et sélective trouvent davantage d'erreurs dans la copie d'un élève de classe sociale défavorisée que dans celle d'un élève de classe sociale favorisée, **alors que le nombre d'erreurs est strictement le même**. Quand les enseignants sont mis en situation d'évaluation formative, la différence en fonction de l'origine sociale disparaît (Autin *et al.*, 2019).

La même différence est observée si on demande aux enseignants soit de noter simplement la dictée soit de rédiger des feedbacks/commentaires. Le nombre d'erreurs repérées est significativement plus élevé pour les élèves de milieu défavorisé en situation de notes isolées, alors qu'en situation de feedbacks sous forme de commentaires, la différence disparaît. Le simple fait d'induire une situation d'évaluation normative conduit donc les enseignants à repérer davantage d'erreurs dans une dictée, ce qui pourrait encore se comprendre. Le souci est que le repérage d'un plus grand nombre d'erreurs ne se produit que pour les élèves de classe sociale défavorisée, alors que – rappelons-le – le nombre d'erreurs est strictement identique.

« Ces études montrent donc que la fonction de sélection de l'école, et son association avec l'évaluation normative, amènent à des évaluations qui perpétuent les inégalités qui existent par ailleurs dans le système scolaire » (Butera, 2023, p. 8).

2. Le pouvoir de la situation et des perceptions

Figure 6 : Le pouvoir de la situation (Goudeau, 2023)



²⁶ Sur cette question, voir l'intervention de Gottsman (2023) : <https://www.youtube.com/watch?v=s3Lv-avj0J0&list=PLWZT1gcRUA4rES9UxFMAB8MZa3yfPOuRT&index=23>

Comme l'illustre déjà l'étude présentée au point précédent et la sous-partie B. 3. du chapitre 2 (voir p. 27), les situations d'évaluation et la perception²⁷ qu'en ont d'un côté les enseignants, de l'autre les élèves, jouent un rôle déterminant dans la génération ou l'amplification des inégalités.

Alors que l'influence des formes d'évaluation (sommative ou formative) relève de l'institution scolaire elle-même, le rôle des stéréotypes renvoie à des croyances culturelles plus larges qui débordent du cadre scolaire. Les études de psychologie sociale portant sur la menace du stéréotype ont clairement montré que des stéréotypes liés à certaines caractéristiques des élèves (genre, origine sociale, filière fréquentée, etc.) peuvent influencer leurs performances dans une tâche d'évaluation.

Une autre série d'études (Goudeau & Croizet, 2017) a aussi montré que les situations de comparaison sociale imposées par le contexte de classe contribuaient à accentuer les inégalités en fonction de l'origine sociale. De manière intéressante, ces mêmes auteurs ont montré qu'il était possible de limiter ces effets délétères en conduisant les élèves à interpréter les différences de résultats comme résultant de causes contrôlables (par exemple différence d'entraînement) plutôt que de causes internes non contrôlables (différence de capacité ou d'intelligence)²⁸.

En résumé, cet ensemble d'études de psychologie sociale illustre que **ce qui se passe dans la classe au quotidien peut générer des inégalités scolaires** en fonction de l'origine sociale et du genre notamment, ainsi qu'amplifier les inégalités sociales existantes. À cet égard, on a vu que **les situations d'évaluation et les perceptions qu'en ont les différents acteurs jouent un rôle crucial**. Il suffit parfois de pas grand-chose pour changer la perception que les élèves se font d'une tâche d'évaluation et creuser ainsi les écarts ou au contraire les réduire. Ainsi, le message positif qui se dégage de ces travaux est que des leviers peuvent être activés pour renverser la tendance et que certaines situations d'évaluation ont montré leur efficacité pour réduire les inégalités ou éviter de les accentuer.

Sans surprise, les situations d'évaluation qui ont montré un potentiel pour tendre vers plus d'équité sont :

- les situations d'évaluation formative (plutôt que normatives) ;
- les situations qui n'encouragent pas ou minimisent la comparaison sociale ;
- l'interprétation des différences de résultats par des facteurs externes et contrôlables plutôt qu'internes et non changeables ;
- le développement d'une conception incrémentale de l'intelligence plutôt que fixiste.

Comme le souligne Goudeau en conclusion de sa note (2023),

²⁷ Cette question du rôle des perceptions peut aussi concerner les parents. Bénit (2023) dans sa note intitulée « Quelle perception par les parents et par les élèves des pratiques évaluatives en classe ? » met en évidence d'importantes différences en fonction de l'origine sociale tant du côté des parents que des élèves.

²⁸ Le rôle-clé que joue la conception que l'élève se fait de l'intelligence est aussi abordé par Fenouillet dans sa note (2023). Selon que l'élève a une conception fixiste de l'intelligence (voyant celle-ci comme une entité constituée de traits stables, fixes, innés, non contrôlables, que l'on peut difficilement changer) ou une conception incrémentale de l'intelligence (un ensemble de qualités malléables, contrôlables, qui peuvent s'améliorer avec le temps), il va interpréter différemment ses réussites et ses échecs, ce qui impactera sa motivation et les buts qu'il se fixe. Selon Fenouillet, « ces résultats ont conduit à la mise en place de programmes éducatifs visant à orienter les conceptions de l'intelligence vers la conception incrémentale, programmes qui ont montré un impact significatif sur la réussite scolaire des élèves les plus en difficulté (Yeager *et al.*, 2019) » (Fenouillet, 2023, p. 67).

« bien sûr, il existe d'autres causes économiques et structurelles à ces inégalités, mais ce qui se passe pendant la classe n'est pas à négliger. S'il paraît difficile de modifier directement les contextes culturels et institutionnels, les enseignants ont le pouvoir en revanche, dans une certaine mesure, de modifier certaines croyances et pratiques à l'intérieur de la salle de classe. Conduire les élèves à changer la façon dont ils interprètent leurs réussites, difficultés et échecs, modifier le sens associé aux différences de performances dans la classe, limiter la compétition, ou encore prendre en compte le fait que les élèves arrivent en classe inégalement préparés aux apprentissages et que par conséquent ce qui a été acquis à travers l'expérience familiale pour certains doit faire l'objet d'un travail explicite en classe pour les autres, sont autant de pistes possibles pour réduire la construction des inégalités sociales » (Goudeau, 2023, p. 83).

Chapitre 4 : Les pratiques d'évaluation en France

A. Repères historiques

Depuis l'origine de l'école, les enseignants estiment la valeur du travail de leurs élèves. Ils portent des appréciations, des jugements sur les compétences, les savoirs, les comportements ou les performances produites par leurs élèves.

L'histoire de l'école et plus particulièrement l'histoire des pratiques évaluatives en France, nous permet de mieux comprendre la culture de l'évaluation présente dans la société française et pourrait même éclairer certains malentendus entre les différents acteurs concernés par l'évaluation des élèves. Présentons quelques-uns de ces repères.

Au Moyen-Âge, l'école fonctionne selon des modalités très différentes de celles que l'on connaît actuellement. Le préceptorat et les maîtres privés exercent leur mission sans avoir recours à l'évaluation et aux examens.

Ce fonctionnement est remis en cause par les Jésuites (Merle, 2015). En effet, au milieu du **XVI^e siècle**, Ignace de Loyola et ses disciples définissent une nouvelle organisation de l'enseignement. Cet ordre entend sélectionner une nouvelle élite et développer une **pédagogie à finalité éminemment sélective, fondée sur le classement et la compétition**. L'ensemble des classements réalisés au fil de l'année permet de catégoriser les élèves selon trois niveaux : *optimi* (les meilleurs élèves ayant l'autorisation de passer dans la classe supérieure), *dubli* (les élèves admis dans la classe supérieure à l'essai), *inepti* (les élèves considérés comme inaptes et donc invités à quitter l'établissement). Les pratiques évaluatives de ces collèges s'inscrivent dans un contexte de rivalité permanente où l'enseignant ne note pas, mais classe constamment en référence à un niveau de connaissances global attendu. En d'autres termes, pour cet ordre, la valeur d'un élève dépend strictement de la valeur des autres élèves.

À la fin du XVII^e siècle, l'Institut des Frères des écoles chrétiennes met en œuvre une pédagogie dont les principes sont antagonistes à ceux des collèges jésuites. En effet, l'objectif de cet institut est d'éduquer et d'instruire les enfants des catégories populaires et non de repérer une élite. Ces écoles organisent les apprentissages selon une progression personnelle. « Le passage à une leçon supérieure repose sur une forme d'évaluation constituée, non par un classement, mais par une appréciation globale des compétences acquises réalisées par le maître et confirmée par le Frère Directeur ou l'inspecteur » (Merle, 2015, p. 80). Dès lors, sans classement, les élèves évoluent et sont répartis selon trois catégories : les commençants, les médiocres et les avancés/parfaits.

L'analyse de ces deux pédagogies conduit certains chercheurs (comme Merle, 2015) à mettre en exergue les différences importantes caractérisant ces deux institutions : discipline, compétition, sélection, gloire, hiérarchie des honneurs pour les collèges jésuites et absence de classement, parcours d'apprentissage défini par des niveaux scolaires et progression dans les apprentissages pour les écoles chrétiennes. Au fil de l'histoire, ces pratiques évaluatives semblent avoir influencé les choix pédagogiques de telle ou telle école ou tel ou tel niveau d'enseignement.

Ainsi, dans la première partie du XIX^e siècle, même si une grande variété de pratiques peut être inventoriée, l'enseignement secondaire semble s'inspirer du modèle jésuite en finalisant la réussite

des élèves avec un classement final mettant à l'honneur les meilleurs éléments (Chervel, 1992)²⁹. Alors que l'enseignement primaire, ne mettant en place ni note, ni classement, utilise un système de récompenses et punitions, davantage inspiré des écoles chrétiennes.

Dans cette même première moitié du XIX^e siècle, l'évaluation des candidats à l'examen du baccalauréat se fait à l'aide d'un boulier où les boules de couleur symbolisent l'appréciation des examinateurs : rouge (favorable à l'obtention), blanche (abstention) et noire (défavorable). Merle (2015) souligne le fait qu'une telle pratique évaluative fonctionne sur le principe de la sentence : obtention versus non-obtention. À ce sujet, il faut attendre le Second Empire pour observer un changement de pratique concrétisé par l'utilisation d'une échelle de 0 à 5 pour les candidats bacheliers. Et l'utilisation de la note sur 20 (pour les épreuves écrites) apparaît en 1890 de manière concomitante avec le baccalauréat « moderne ».

1890 représente une date importante dans l'histoire des pratiques évaluatives puisque l'article 21 de l'arrêté du 5 juillet 1890 annonce la notation sur 20 des compositions en mentionnant précisément que pour « les compositions, chaque copie aura sa note chiffrée de 0 à 20 ». Avec la note, l'enseignant doit rendre compte de la qualité des productions de ses élèves (Merle, 2015). À partir de cette date, la note s'installe durablement dans le système éducatif français. Avec ce nouvel outil pédagogique, le système de « sentence » à l'examen est abandonné au profit d'une technique de « mesure » qui apparaît comme plus objective. Ainsi, par rapport au classement, la notation sur 20 est considérée comme un progrès pédagogique (Lelièvre, 2014).

B. Évolutions récentes

Les critiques de la note sont nombreuses au cours de l'histoire de l'éducation, de Claparède à Freinet.

Ainsi, en mars 1968, lors du colloque d'Amiens présidé par Alain Peyrefitte, les prix, les classements et les notes sur 20 sont critiqués. Ces différentes critiques conduisent le ministre de l'Éducation nationale – Edgar Faure – en **1969**, à préconiser une évaluation par lettres : A, B, C, D, E en abandonnant la note sur 20 ainsi que les classements.

Si l'abandon des classements n'a jamais été l'objet de réformes ultérieures, le sujet de la notation sur 20 a rapidement été retravaillé par le ministre qui a succédé à Edgar Faure - Olivier Guichard - puisque dès 1971, les notes sont à nouveau prônées surtout dans les classes préparant un examen (Deubel, Huart, Montoussé, & Vin-Datiche, 2007).

Après ces différents textes institutionnels publiés juste après 1968, la notation sur 20 poursuit son installation dans le système éducatif français. Pendant de nombreuses années, alors que le thème de l'évaluation est l'objet d'un nombre important de publications scientifiques, peu de textes ministériels se prononcent sur les pratiques à valoriser ou à éviter.

²⁹ Cela serait lié à l'émergence de la note sur 20 pour départager les candidats au concours d'entrée de l'École Polytechnique. Cette pratique se serait répercutée dans les classes préparatoires et se serait diffusée jusque dans les établissements d'enseignement secondaire. Pour plus de détails, voir Chesné, Piedfer-Quênay, & Jeanneau (2023), pp. 6-7. (https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_CHESNE-PIEDFER-JEANNEAU.pdf)

Il faudra attendre 2006 pour que le ministère de l'Éducation nationale choisisse de mettre en place de nouvelles modalités évaluatives à la suite de la création du socle commun de connaissances et de compétences (loi du 23 avril 2005), des modalités organisées sous la forme d'un Livret personnel de compétences (LPC) qui propose, pour la première fois de l'histoire, une évaluation par compétences, sans aucune directive au sujet de la notation sur 20.

Puis, en 2012, dans une lettre adressée à tous les personnels de l'Éducation nationale, Vincent Peillon – alors ministre de l'Éducation nationale – annonce à la fois la conception d'un nouveau socle commun de connaissances, de compétences et de culture et un nouveau livret d'évaluation moins complexe que celui de 2006. Le ministre insiste sur le fait que ce livret doit être un outil de dialogue de qualité avec les parents et qu'il est également nécessaire de penser des outils de suivi pour les élèves, clairs et compréhensibles. Et il ajoute que cette révision sera réalisée dans un cadre concerté et transparent. Notons aussi que c'est à l'occasion de la loi de refondation de l'école de 2013 que la notion de « bienveillance » (qui se décline en « école bienveillante », mais aussi en « évaluation bienveillante ») est apparue dans les textes officiels.

Suite à cette annonce, Benoît Hamon – qui succède à Vincent Peillon – organise une conférence nationale sur l'évaluation des élèves en 2014 dans la perspective de construire une véritable politique de l'évaluation.

À l'issue des travaux de cette conférence nationale, un certain nombre de changements sont proposés. En effet, c'est à compter de la rentrée 2016 (Bulletin officiel n°27 du 2 juillet 2015) que les livrets scolaires de l'école et du collège évoluent vers un nouveau livret scolaire commun pour l'ensemble de la scolarité du CP à la 3^e, qui devient le livret scolaire unique numérique (LSU ou LSUN) en 2017.

Ce livret scolaire est établi pour chaque élève en âge d'obligation scolaire. Il comporte, pour chaque cycle, les bilans périodiques des acquis scolaires de l'élève, les bilans de fin de cycle, les attestations obligatoires. Il remplace le livret personnel de compétences à compter de la rentrée scolaire 2016. Ainsi, ce nouveau livret rend compte aux parents des acquis et des progrès de leurs enfants en restituant une évaluation complète.

La ministre de l'Éducation nationale, Najat Vallaud-Belkacem, à l'origine de ce nouveau texte, ne donne aucune directive sur l'utilisation de la notation sur 20, les équipes enseignantes sont libres de leurs choix pédagogiques. Néanmoins, un cadre est défini pour concevoir les nouvelles pratiques évaluatives à mettre en place. Il est attendu que l'évaluation des acquis scolaires des élèves soit réalisée dans un but précis : l'amélioration de l'efficacité des apprentissages, et ce, en permettant à chaque élève d'identifier ses acquis et ses difficultés afin de pouvoir progresser. À partir de la rentrée scolaire 2016, les modalités d'évaluation doivent privilégier une évaluation positive, simple et lisible, qui valorise les progrès, soutient la motivation et encourage les initiatives des élèves, une évaluation bienveillante est à mettre en place. Suite à ce texte, l'évolution des pratiques d'évaluation des élèves est en mouvement tout au long de la scolarité obligatoire.

La loi de 2013 a ajouté l'alinéa suivant à l'article L. 311-1 du code de l'éducation : « L'évaluation sert à mesurer et à valoriser la progression de l'acquisition des compétences et des connaissances de chaque élève ». **Cet alinéa n'a pas été modifié ensuite** (Loi n° 2019-791 du 26 juillet 2019 pour une école de la confiance).

C. État des lieux des pratiques

Les pratiques évaluatives restent assez différentes entre l'école primaire et le collège, examinons-les successivement.

1. Les pratiques évaluatives dans le premier degré

L'enquête Epode³⁰ 2018 montre que les professeurs des écoles jugent les pratiques visant l'implication active des élèves dans les apprentissages moins faisables et moins prioritaires que des pratiques relevant d'un enseignement explicite et donc qu'ils les adoptent moins. Moins de trois enseignants sur dix déclarent s'engager fréquemment dans des démarches d'expérimentation ou d'innovation, attribuer fréquemment des rôles aux élèves lors des débats, ou proposer fréquemment des outils d'auto-évaluation (Note d'information de la Depp n°20.32, 2020).

Néanmoins, cette même enquête relève que

« l'évaluation formative et la pédagogie active sont progressivement mises en place du CP au CM2, niveau où leur score de fréquence se rapproche de celui du collège. Au CP, 13 % des professeurs déclarent attribuer à certains élèves des rôles lors des débats (contre 33 % au CM2). Une proportion similaire déclare donner une grille d'évaluation aux élèves pour qu'ils comprennent leurs réussites et leurs erreurs (contre 30 % au CM2) » (Note d'information de la Depp n°20.23, 2020).

De manière complémentaire, les observations et enquêtes de l'Inspection générale de l'Éducation, du Sport et de la Recherche (IGÉSR) montrent que les projets d'école intègrent souvent les grandes lignes d'une évaluation positive. Ce constat témoigne du fait que les professeurs des écoles se concertent quant au rôle des différents outils de suivi des acquis des élèves. Par exemple, le LSU est considéré comme un outil de communication aux familles, et il peut être complété, par exemple, par des graphiques représentant la progression de l'élève (une grande variété d'outils sont mis en place).

« Dans certaines circonscriptions, une politique volontariste a permis de mettre en place, voire de généraliser, des modalités d'évaluation nouvelles visant par exemple à associer l'élève à son évaluation et à lui laisser le choix des moments auquel il se sent prêt pour être évalué, dans une dynamique progressive et positive » (Hazard, 2023, p. 27).

Une enquête de l'IGÉSR (2022) met en évidence des pratiques différentes selon les disciplines pour le cours moyen (CM1-CM2) :

« Pour les enseignements artistiques, l'histoire des arts et l'EMC, les élèves ne sont pas évalués. Pour les langues vivantes, quelques évaluations ont été rencontrées dans environ un quart des classes observées. Il s'agit généralement d'une évaluation par période, en fin de séquence, portant sur des thématiques pouvant être rencontrées au cycle 2 comme au cycle 3 [...]. Ces évaluations portent sur du lexique avec, la plupart du temps, des réponses à donner sous forme de mots isolés.

³⁰ Les enquêtes Épode (Enquête PériODique sur l'Enseignement), menées par la Depp (Direction de l'évaluation, de la prospective et de la performance), ont pour objectif de décrire les pratiques professionnelles des enseignants des premier et second degrés.

Pour l'histoire, la géographie, les sciences et la technologie, des évaluations sont mises en place, souvent au rythme d'une évaluation par période et en fin de séquence. Elles évaluent principalement des connaissances qui doivent être restituées sous forme de groupes nominaux ou de phrases courtes. [...]

En mathématiques et en français, les évaluations se limitent souvent à des tests bilans de fin de séquence, les interrogations écrites très courtes sont excessivement rares.

Pour le français, la situation est assez hétérogène selon les classes. Dans de très nombreuses classes, l'orthographe est évaluée à un rythme soutenu sous forme de dictées (une dictée chaque semaine avec généralement un travail quotidien pour la préparer). Le second type d'évaluation le plus fréquent en français, rencontré dans plus de la moitié des classes, concerne les conjugaisons. [...].

Pour les mathématiques, les évaluations sont plus fréquentes et plusieurs évaluations par période sont mises en place dans presque toutes les classes. [...].

L'évaluation est donc le plus souvent un moment distinct de la leçon dramatisant parfois excessivement l'évaluation » (Hazard, 2023, p. 29).

2. Les pratiques évaluatives dans le second degré

L'enquête internationale TALIS 2018 (OCDE, 2019) nous offre un certain nombre d'éléments pour caractériser les pratiques évaluatives des enseignants de collège en France. On y apprend que 96,2 % des enseignants de collège français participant à l'enquête TALIS 2018 déclarent gérer eux-mêmes leur processus d'évaluation fréquemment ou toujours (contre 79,6 % en moyenne dans les pays de l'Union européenne) ; que 77,4 % d'entre eux déclarent ajouter fréquemment ou toujours un commentaire écrit à la note (contre 62,6 % dans l'UE) ; que 20,5 % déclarent laisser les élèves s'évaluer eux-mêmes fréquemment ou toujours (contre 36,2 % dans l'UE) ; et que 78,3 % (contre 80,3 % dans l'UE) déclarent observer les élèves pendant qu'ils effectuent en classe une tâche particulière et leur apporter immédiatement des commentaires sur leur travail (Kiwani-Zacka, 2023).

En parallèle, la Depp en collaboration avec une équipe de chercheurs et d'enseignants a réalisé un panorama des pratiques évaluatives déclarées d'enseignement de collège en mathématiques par l'intermédiaire d'une enquête nationale à grande échelle, Praesco (2019). Horoks & Pilet (2013) en présentent les principaux résultats. Selon ces chercheuses,

« L'évaluation pour faire apprendre les mathématiques aux élèves ne semble pas suffisamment présente dans la classe pour favoriser la prise d'initiatives chez les élèves et leur donner des moyens de contrôle de leur action. Cet état des lieux sur les pratiques déclarées des enseignants sur l'évaluation conduit à interroger les leviers possibles pour les faire évoluer » (Horoks & Pilet, 2013, p. 156).

De manière complémentaire, l'enquête (Epede) indique que les pratiques visant l'implication active des élèves dans les apprentissages sont moins fréquentes et considérées comme moins faisables et moins prioritaires. Ainsi, seulement un professeur interrogé sur deux (54 %) déclare proposer fréquemment des démarches d'investigation (ou des démarches inductives) aux élèves et un sur trois (33 %) déclare proposer fréquemment aux élèves des outils d'auto-évaluation (grilles d'indicateurs sur les capacités, connaissances et attitudes) pour qu'ils évaluent leurs progrès, leurs réussites et leurs difficultés (Note d'information de la Depp n°20.23, 2020).

Les observations et enquêtes de l'IGÉSR suggèrent que même si des évolutions de pratiques évaluatives sont repérées, globalement,

« l'évaluation reste focalisée sur la note et la moyenne, notamment dans le second degré, obérant la pertinence de l'évaluation. L'évaluation formative reste peu développée, l'évaluation des apprentissages l'emportant sur l'évaluation pour les apprentissages. Les évaluations formative et sommative sont fréquemment confondues au sein d'une seule et même note » (Hazard, 2023, p. 37).

En outre, l'évaluation par compétences est souvent considérée comme opposée à la notation (Genelot, 2023)³¹. Ce phénomène est amplifié par la coexistence d'un système de suivi du niveau de maîtrise des composantes du socle commun sans note d'une part, et des systèmes de suivi des acquis des élèves laissés au libre choix des équipes pédagogiques dans chaque discipline d'autre part. De plus, au moment du brevet des collèges, le degré de maîtrise des composantes du socle est converti en points pour la prise en compte du socle commun dans l'obtention du diplôme. Ainsi, au fil de l'année, les enseignants sont amenés à situer les élèves sur les niveaux de maîtrise des compétences du socle et, en parallèle, ils notent des devoirs. En outre, les travaux de l'IGÉSR pointent la coexistence de compétences disciplinaires d'une part et des compétences du socle d'autre part comme une potentielle source de confusion.

« En fin de cycle 4, la note du socle peut résulter d'une moyenne pondérée des notes issues des disciplines mais également du calcul automatisé réalisé par ce type d'outil, le résultat étant parfois à la base d'un conseil de socle visant à valider ces calculs automatiques. Il n'est alors pas étonnant que les professeurs déplorent bien souvent que les élèves s'y retrouvent difficilement entre le socle et les disciplines. La prise en charge du socle par des logiciels interroge quand on paramètre le logiciel et que de façon automatique, celui-ci renseigne les éléments d'un profil de l'élève. Les pistes actuelles sont dans des collèges qui font évoluer leurs conseils de classe vers de véritables conseils de socle où chaque professeur met son expertise d'évaluateur disciplinaire au service du positionnement des élèves sur les compétences transversales posées par le socle commun » (Hazard, 2023, p. 32).

Quant aux parents, les travaux de l'Inspection générale tendent à montrer que la question des compétences (associée à une évaluation qui fonctionne autrement qu'avec des notes) rencontre des résistances. Les élèves, comme leurs parents, semblent très attachés à l'utilisation de la note qui constitue, à leurs yeux, un repère de réussite très simple à comprendre.

« Elle participe à leur positionnement social et scolaire. Elle est une forme de reconnaissance entre pairs et au-delà. Elle est un passeport pour leur avenir. Cette appréhension identitaire est d'autant plus graduée entre le lycée professionnel (moins forte, mais très présente également), le lycée technologique et le lycée général, que l'élève appartient aux catégories socioprofessionnelles dites supérieures. Et pourtant, 46 % des lycéens interrogés pensent qu'il y a trop d'évaluations notées et que les notes sont des éléments de stress (cités à la hauteur de 63,8 %) » (Hazard, 2023, p. 34).

Ces observations convergent avec les résultats de l'enquête conduite par le Crédoc à la demande du Cnesco, puisque 47 % des parents d'élèves interrogés considèrent que la notation chiffrée est le

³¹ https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_GENELOT.pdf

système de notation qui rendrait le mieux compte des acquis de leur enfant. Cette proportion augmente au fil de la scolarité (30 % à l'école élémentaire contre 60 % au lycée). Les parents détenteurs d'un diplôme supérieur ou égal au baccalauréat sont plus nombreux que les autres à estimer que la notation chiffrée constitue le meilleur système de notation (50 % versus 41 % pour les non-bacheliers). Il en va de même pour les parents qui considèrent leur enfant comme un bon élève par rapport à ceux qui indiquent que leur enfant fait partie des plus faibles de sa classe (49 % contre 37 %).

3. Les pratiques évaluatives à la lumière de trois études empiriques

Deux recherches complémentaires (Genelot, 2023 et Bénit, 2023) visent à rendre compte des pratiques évaluatives mises en œuvre au quotidien dans les classes de l'école primaire et du collège en se centrant sur les différents acteurs concernés : les enseignants, pour la première étude, et les élèves et les parents pour la seconde étude. Par ailleurs, une enquête menée par le Crédoc rend compte du point de vue de 1 200 parents au sujet de l'évaluation de leur enfant.

a. L'activité évaluative des enseignants

Suite à une commande du Cnesco, Sophie Genelot a conduit une recherche pour mieux documenter l'activité évaluative des enseignants de l'école élémentaire et du collège. Elle a ainsi mené une enquête qualitative en filmant, sur un temps long, des enseignants volontaires de 13 classes allant du CP à la troisième. Son objectif était **d'identifier toutes les situations évaluatives dans le déroulement ordinaire et quotidien des classes.**

351 situations évaluatives ont été identifiées sur les vidéos puis décrites, et ce, afin de nommer et de catégoriser les différentes situations évaluatives. Pour ce faire, l'auteure définit ce qu'elle entend par situation évaluative (Genelot, 2019). Il s'agit de

« toute situation dans laquelle l'enseignant et/ou les élèves prélèvent des informations à propos des apprentissages en cours, les interprètent en fonction de critères (plus ou moins explicites) pour formuler une appréciation qui est communiquée et fonde une prise de décision (plus ou moins explicite, différée ou immédiate) » (Genelot, 2023, p. 38).

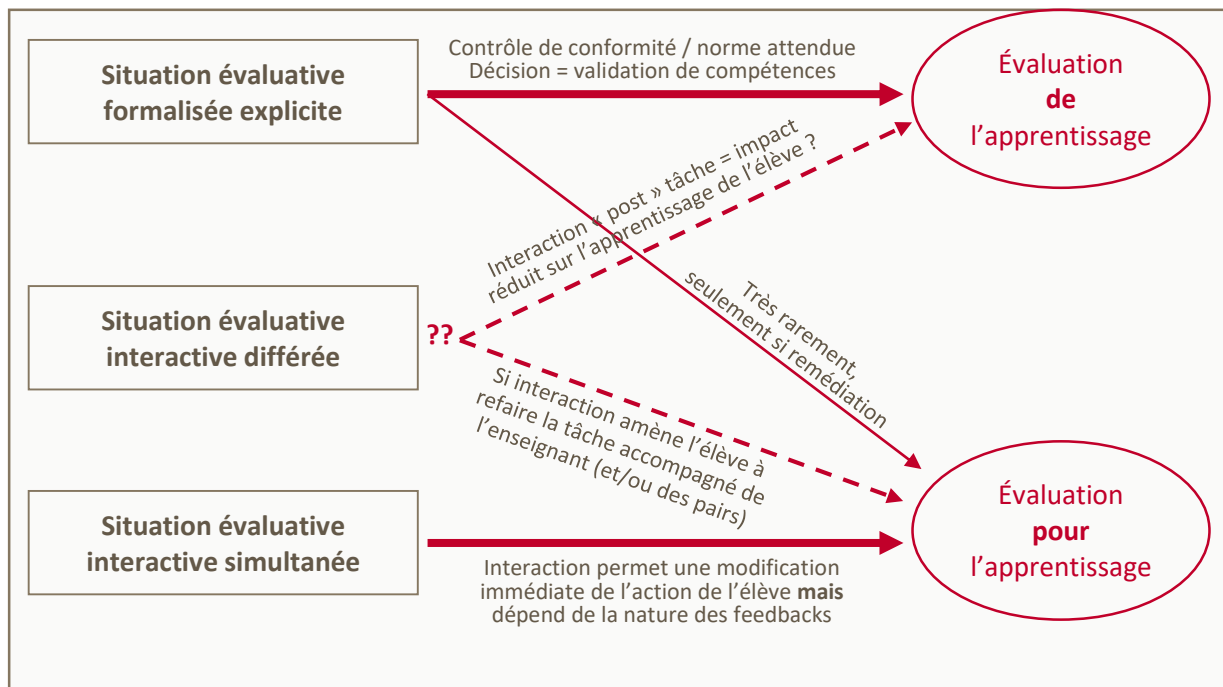
Ainsi, trois grandes catégories de situations évaluatives caractérisent les pratiques ordinaires des enseignants :

- **les situations évaluatives formalisées explicites**, comme par exemple, les dictées en français, les devoirs surveillés, les contrôles flash, les exposés ;
- **les situations évaluatives interactives différées**, sous forme de dialogues évaluatifs adressés à un élève en particulier ou à un groupe d'élèves lors de corrections collectives par exemple ;
- **les situations évaluatives interactives simultanées** qui sont observées lorsque les élèves sont en activité. Ces dialogues évaluatifs peuvent être adressés à un élève ou à plusieurs.

Ces différentes catégories de situations évaluatives sont inégalement représentées selon les niveaux scolaires et les disciplines concernés. Genelot (2023) illustre ce point en présentant une répartition des situations évaluatives selon les niveaux scolaires.

En réalisant une analyse croisée de ces pratiques évaluatives avec les représentations des enseignants – décrites par les enseignants eux-mêmes lors d’entretiens – Genelot (2023) propose l’établissement de relations entre chacun des trois types de situations évaluatives observées et les deux principales fonctions de l’évaluation décrites par la littérature scientifique, « *assessment of learning* » versus « *assessment for learning* ».

Figure 7 : Liens entre typologie des situations évaluatives observées et fonctions de l’évaluation (Genelot, 2023)



Ainsi, selon cette chercheuse,

« les situations évaluatives interactives simultanées constituent, potentiellement, une forme d'évaluation favorable à la régulation de l'activité des élèves, mais c'est la nature des feedbacks qui leur sont adressés lors de ces situations qui est déterminante pour qu'elles constituent un réel soutien à l'apprentissage des élèves » (Genelot, 2023, p. 51).

En poursuivant son analyse, elle va encore plus loin en affirmant que « plus que la situation en elle-même, ce sont les différents choix que vont opérer les enseignants, dans leurs interactions avec les élèves, qui vont rendre celle-ci plus ou moins favorable à l'apprentissage des élèves » (Genelot, 2023, p. 52). Cette étude empirique est riche d'enseignements concernant les pratiques évaluatives effectives des enseignants, elle corrobore notamment des observations formalisées par Joigneaux (2023) et Bénil (2023). En effet, **de nombreuses pratiques évaluatives ne sont pas conscientisées par les enseignants**. À ce sujet, Genelot (2023) emploie les expressions « **l'évaluation qui dit son nom** » versus « **l'évaluation qui ne dit pas son nom** ». Un tel résultat

« contribue, nous semble-t-il, à contredire la représentation la plus courante de l'évaluation chez les enseignants, du moins ceux de notre échantillon, et à conclure qu'un certain nombre d'entre eux font sans doute souvent "de l'évaluation sans le savoir". C'est une particularité sur laquelle nous attirons l'attention des acteurs qui s'emploient à vouloir faire évoluer les

pratiques évaluatives des enseignants : car comment faire évoluer des pratiques qui ne sont qu'en partie conscientisées comme évaluatives par les enseignants eux-mêmes ? »

b. Les perceptions des parents et des élèves à propos des pratiques évaluatives

Pour rendre compte des perceptions de l'évaluation par les élèves – grande section, CP, CM2, sixième et cinquième et leurs parents, Bénit (2023) a mis en place une étude qualitative. Dans la continuité des travaux de Genelot (2019), il montre que l'évaluation ne revêt pas une forme unique en classe et surtout qu'elle fait l'objet de **nombreux échanges au sein de la famille et de plusieurs stratégies**, surtout dans les familles qui ont la volonté de préparer leurs enfants à réussir à l'école.

Les catégories socioprofessionnelles des parents et le statut scolaire des élèves apparaissent comme les deux variables les plus déterminantes au niveau de l'intériorisation de l'impératif de réussite scolaire (Merle, 2005) et du statut accordé à l'erreur.

« L'expérience subjective des élèves face à l'évaluation dépend de différentes instances socialisatrices, l'école et les parents jouant les rôles les plus importants au début de la scolarité » (Bénit, 2023, p. 63). En conséquence, certains élèves construisent des compétences leur permettant, très tôt, un autocontrôle de leur comportement en classe, ils analysent finement leurs erreurs, ils repèrent les attentes des professeurs et celles de leurs parents. Inévitablement, de telles compétences acquises par certains concourent à creuser les écarts entre élèves, notamment avec ceux dont l'environnement familial n'a pas permis la construction de telles compétences.

En parallèle, de nombreux élèves semblent être conscients des différentes situations d'évaluation non formelles – les situations évaluatives interactives simultanées – puisqu'ils disent **avoir l'impression d'être évalués constamment par leurs enseignants** et semblent être tout à fait **lucides quant à l'enjeu** de telles évaluations qui préfigurent, selon eux, des évaluations plus formelles.

Pour ce qui est des parents, plusieurs travaux ont montré (par exemple, Garcia, 2018 ; Merle, 2005) que certains d'entre eux préparent leurs enfants à identifier, comprendre et anticiper les attentes de leurs enseignants en matière d'évaluation, et ce, pour toutes sortes d'évaluation, notes/sans notes, couleurs, tableau de compétences, etc.

Bénit (2023) ajoute que « ce travail prolonge la lente intériorisation de l'impératif de réussite scolaire (Merle, 2005), puisque les enfants comprennent rapidement que ces méthodes leur permettent fréquemment d'obtenir de bons résultats » (Bénit, 2023, p. 63).

« Au final, le travail pédagogique parental quotidien soulève la question de la création de potentielles inégalités : "le fait qu'une partie des élèves réussissent bien scolairement grâce à un travail invisible réalisé au sein de la famille conduit à valider une norme scolaire inaccessible sans travail parental" » (Garcia, 2018, p. 15. Cité par Bénit, 2023, pp. 63- 64).

c. Le point de vue des parents – Enquête Crédoc

Le Cnesco a réalisé en partenariat avec le Crédoc (2023) une enquête auprès de parents d'élèves scolarisés du CP à la terminale. Globalement, cette enquête sur un échantillon de 1 200 parents montre que l'évaluation préoccupe largement les parents, qui sont par exemple nombreux à

demander aux enfants les notes obtenues par leurs camarades de classe, même s'ils considèrent que l'apprentissage est plus important que la compétition.

Quels sont les principaux résultats de cette enquête ?

- La grande majorité des parents se dit bien informée à la fois des progrès, mais aussi des difficultés rencontrées par leur enfant. Pour près de 8 parents sur 10, c'est à travers l'évaluation qu'ils savent ce que fait leur enfant à l'école, même s'il semble qu'ils ont de moins en moins accès à la façon dont se déroule l'évaluation de leur enfant à mesure de sa scolarité.
- La réussite scolaire représente un enjeu important pour eux : c'est un sujet de discussion avec leur enfant mais aussi avec leur entourage par exemple. S'ils accèdent aux résultats scolaires principalement *via* des logiciels ou applications de vie scolaire, c'est aussi par le biais de leur enfant qu'ils se tiennent au courant, soit que celui-ci les informe, soit que les parents lui posent des questions. Leur degré d'inquiétude est plus important lorsque les difficultés touchent les mathématiques et le français.
- La majorité des parents interrogés considèrent que ce que l'on apprend à l'école, au collège ou au lycée est plus important que d'avoir de bonnes notes, mais une proportion similaire indique que les résultats scolaires sont très importants. Dans le même temps, près de 8 parents sur 10 indiquent que ce n'est pas important si leur enfant n'est pas le meilleur de sa classe, mais plus d'un sur deux se renseigne sur les résultats de ses camarades à la suite d'une évaluation.
- Les parents d'élèves indiquent que l'évaluation est un facteur de stress important pour leurs enfants. D'après leurs réponses, près d'un enfant sur deux est stressé à l'approche d'un contrôle (57 %), perd ses moyens pendant la passation d'une évaluation (45 %) et appréhende l'annonce des résultats (51 %).

Chapitre 5 : Enjeux de la formation des enseignants

A. L'évaluation soutien d'apprentissage : bien plus qu'un changement de méthode

En préambule de ce chapitre consacré aux enjeux de la formation des enseignants, il importe de rappeler que le passage d'une évaluation des apprentissages à une évaluation soutien de l'apprentissage représente un véritable changement de paradigme ; celui-ci prend place dans le contexte d'un système éducatif qui présente des caractéristiques susceptibles d'entrer en tension avec cette nouvelle conception de l'évaluation (Crahay *et al.*, 2019 ; Merle, 2018).

Comme le souligne Butera (2023), « les évaluations normative et formative ne sont pas seulement des méthodes ou des pratiques, mais elles s'insèrent dans un ensemble cohérent de représentations sur les fonctions et les valeurs de justice de l'école. » (Butera, 2023, p. 6). L'école française est fortement imprégnée d'une idéologie méritocratique, où il apparaît comme « juste » que les meilleurs élèves ou du moins les plus méritants (ceux qui font des efforts) soient récompensés par de meilleures notes, qui leur ouvriront les portes des meilleurs cursus et les mettront sur la voie de la réussite sociale. Ce modèle méritocratique, étroitement couplé au modèle de l'égalité des chances (voir 0, p. 33 et suivantes) entre en confrontation directe avec un autre modèle, le modèle de l'égalité des acquis ou des résultats, qui est au principe même de l'évaluation soutien d'apprentissage. Cette évaluation vise bien à soutenir chaque élève dans ses apprentissages et à amener un maximum d'élèves à la réussite plutôt qu'à sélectionner les meilleurs ou à trier et orienter les élèves vers des parcours différenciés. Sur ce sujet, Austin, Batruch et Butera (2015) ont réalisé une enquête par questionnaire auprès d'étudiants du supérieur et mis en évidence que « plus les participants percevaient le système éducatif comme ayant un rôle de sélection, plus ils percevaient l'évaluation normative comme une méthode utile et à utiliser » (Butera, 2023, p. 6).

Amener les enseignants à évoluer vers une évaluation au service de l'apprentissage n'est donc pas un simple changement de pratique ou de méthode. C'est aussi un changement de la conception que l'on se fait de la justice, de l'égalité, de l'équité et de la capacité des élèves à apprendre. Loin de nous cependant l'idée de défendre qu'il convient dès lors de poser comme préalable au changement de pratiques une modification des conceptions. Néanmoins, ignorer cette intrication entre pratiques, fonctions de l'évaluation d'une part, modèles de l'égalité, conceptions de la justice et de l'apprentissage d'autre part, serait une grave erreur. **Renforcer l'évaluation au service de l'apprentissage nécessite un double changement, de posture et de culture.**

B. Le développement d'une littératie de l'évaluation

Ceci étant posé, il ne fait aucun doute que la formation initiale des enseignants, tant en France que dans d'autres systèmes éducatifs, prépare insuffisamment les enseignants à l'évaluation (Fontaine, 2023), selon leurs propres dires. De surcroît, « l'évaluation des apprentissages serait une difficulté qui persiste dans le temps et qui serait associée à un désir de quitter la profession » (Fontaine *et al.*, 2012).

Parmi les raisons qui expliquent ce sentiment d'incompétence vécu par les enseignants, Fontaine (2023) souligne une « dualité quant à la responsabilité de l'évaluation » (Fontaine, 2023, p. 180),

source de possibles tensions chez les enseignants. Par dualité, elle entend le fait que les enseignants se sentent pris voire tiraillés entre deux mouvements contradictoires. D'un côté, ils perçoivent bien la montée en puissance de l'évaluation au service de l'apprentissage, évaluation dont la responsabilité leur incombe, et pour laquelle ils se sentent insuffisamment préparés. D'un autre côté, l'externalisation des évaluations de type sommatif les dépossède d'une partie de l'évaluation qui leur incombait traditionnellement.

Des recherches et des programmes ont pour objet le développement des compétences en matière d'évaluation des professionnels de l'enseignement. De nos jours, on parle à ce propos de *littératie de l'évaluation* (*assessment literacy* ; voir Fagnant, 2023).

« La littératie en évaluation des apprentissages peut se définir de la façon suivante : une bonne compréhension de l'évaluation ; des connaissances sur les démarches, concepts, etc. en évaluation ; la capacité de mettre en œuvre ces connaissances et cette compréhension dans des tâches d'évaluation des apprentissages » (Fontaine, 2023, p. 179)³².

Nous empruntons à Fontaine (2023) la liste des compétences pour le développement d'une littératie en évaluation élaborée par un comité américain formé de professionnels de l'éducation (AFT, NCME, & NEA, 1990).

Tableau 3 : Compétences pour le développement d'une littératie en évaluation (Fontaine, 2023)

Les compétences en évaluation des apprentissages	
1	Choisir la méthode d'évaluation en lien avec les décisions éducatives
2	Développer la méthode d'évaluation en lien avec les décisions éducatives
3	Administrer, noter et interpréter les résultats d'évaluation obtenus à des évaluations internes et externes
4	Utiliser les résultats de l'évaluation lors de la prise de décisions concernant les élèves, la planification de l'enseignement, le curriculum et le progrès dans l'école
5	Développer des méthodes de notation valides
6	Communiquer les résultats de l'évaluation aux diverses parties prenantes
7	Reconnaitre l'intégrité, la légalité et la pertinence des méthodes et des informations entourant l'évaluation
8	Prendre en compte les développements récents concernant les politiques et les pratiques en évaluation (ex. l'évaluation formative et la reddition de compte)

³² Pour plus de détails à ce propos, on peut se reporter au rapport scientifique de Fagnant (2023), ainsi qu'à la note de Fontaine (2023).

C. Programmes de développement professionnel : effets et points d'attention

Comme le constate Fagnant (2023), il n'existe pas à ce jour de méta-analyse, de revue systématique ou d'article de synthèse qui permettrait de dégager des critères d'efficacité des programmes de développement professionnel consacrés à l'évaluation³³. On ne pourra donc mentionner ici que les résultats encourageants de quelques études consacrées à ces questions, sans toutefois entrer dans le détail des programmes de formation qui ont chacun leurs spécificités³⁴.

Certains auteurs ont néanmoins tenté de proposer un certain nombre de recommandations issues d'une analyse de la littérature existante. Ainsi, selon Trunbull et Guerzon (2013), les programmes de développement professionnel (portant sur l'évaluation formative) efficaces³⁵ devraient être :

- *intensifs, continus et liés à la pratique ;*
- *intégrés dans des communautés d'apprentissage professionnelles qui soutiennent l'amélioration continue de la pratique enseignante ;*
- *axés sur les contenus ciblés ;*
- *prendre en compte le contexte local dans lequel les enseignants exercent ;*
- *engager les enseignants dans un apprentissage actif, avec des allers-retours vers la pratique.*

Pasquini (2023) et Fontaine (2023) vont dans le même sens, en insistant sur l'importance du travail en équipe ou en réseau *via* des dispositifs collaboratifs (Pasquini) ou des « communautés d'apprentissage professionnelles ». Enfin, Fontaine (2023) souligne l'importance du suivi et de l'accompagnement post-formation des enseignants. En effet,

« même s'ils se disent prêts à mettre en œuvre leurs nouvelles compétences, le retour dans la salle de classe donne souvent lieu au retour des "vieilles pratiques" ! Donc, un accompagnement qui va au-delà de la formation ponctuelle est nécessaire [...]. De plus, [...] les enseignants en formation continue sont aussi responsables de vérifier si les changements qu'ils apportent à leurs pratiques évaluatives font une différence pour les élèves. C'est là l'objectif ultime de ces formations, former les participants afin qu'ils puissent utiliser l'évaluation pour soutenir l'apprentissage des élèves » (Fontaine, 2023, p. 183).

En guise de conclusion à ce chapitre, nous empruntons les points saillants du chapitre qu'a consacré Fagnant (2023) à cette question des compétences et du développement professionnel des enseignants.

³³ Concernant les facteurs favorisant un développement professionnel efficace en général, voir les ressources issues de la conférence de comparaison internationale organisée par le Cnesco en novembre 2020, disponibles gratuitement en ligne : <https://www.cnesco.fr/conference-de-comparaisons-internationales-2020-la-formation-continue-et-le-developpement-professionnel-des-personnels-education/>

³⁴ On trouvera plus d'information sur des recherches visant le développement professionnel des enseignants dans les notes de Fontaine (2023) et de Pasquini (2023), ainsi que dans les rapports scientifiques de Calone & Lafontaine (2023) et de Fagnant (2023).

³⁵ De surcroît, les travaux sur le développement de la littératie de l'évaluation portent sur l'évaluation au sens large. Même si l'évaluation formative y est abordée, les programmes de développement professionnel abordent d'autres types et fonctions de l'évaluation. La notion même d'évaluation comme soutien d'apprentissage y est peu présente.

« Outre certaines variables contextuelles susceptibles de favoriser la mise en œuvre d'une évaluation-soutien d'apprentissage (ESA) en classe, les enseignants semblent aussi devoir disposer d'attitudes favorables à l'ESA ainsi que de compétences professionnelles en évaluation (Heitink et al., 2016 ; Schildkamp et al., 2020 ; Yan et al., 2021) et en didactique (Schildkamp et al., 2020).

Dans la littérature anglo-saxonne, ces compétences didactiques relèvent du concept de pedagogical content knowledge (PCK) mis en évidence par Shulman (1987, 2007) et déjà pointées par Bennett (2011) comme essentielles à prendre en compte. En ce qui concerne les compétences évaluatives, c'est le concept d'assessment literacy (Mertler, 2004) qui a retenu notre attention. Plusieurs travaux montrent qu'il est possible d'améliorer les connaissances évaluatives des enseignants via des programmes de développement professionnel adaptés, mais des recherches complémentaires devraient être menées pour mieux comprendre comment ces connaissances affectent les pratiques de classe et in fine les apprentissages des élèves.

Dans ces programmes de formation, il semblerait qu'une place plus importante mériterait d'être accordée aux travaux récents en matière d'assessment for learning ainsi qu'aux connaissances didactiques (PCK) et aux spécificités des contenus disciplinaires impliqués (Brookhart, 2011) » (Fagnant, 2023, p. 49).

C'est en raison de cette intrication entre les compétences en évaluation et les compétences et connaissances des didactiques disciplinaires qu'un des thèmes de la conférence de consensus y a été dédié³⁶. S'il est difficile, dans une note de synthèse à caractère général, de rendre compte de l'apport de ces travaux à la question de l'évaluation soutien d'apprentissage qui, par définition, ont un caractère non généralisable, ces savoirs et compétences en didactiques disciplinaires sont incontournables pour développer une véritable évaluation soutien d'apprentissage, rendre des feedbacks formatifs axés sur les processus et les démarches d'apprentissage, analyser les erreurs des élèves et en tirer parti, construire une évaluation critériée.

³⁶ Voir notes de Gottsman (2023), de Horoks et Pilet (2023), et de Roussel (2023).

Références

- American Federation of Teachers, National Council on Measurement in Education & National Education Association (AFT, NCME, & NEA) (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Allal, L. (2007). Régulations des apprentissages : orientations conceptuelles pour la recherche et la pratique en éducation. In L. Allal & L. Mottier Lopez (Éds.), *Régulation des apprentissages en situation scolaire et en formation* (pp. 7-23). De Boeck Université.
- Allal, L. (2008). Évaluation des apprentissages. In A. van Zanten (Éd.), *Dictionnaire de l'éducation* (pp. 311-314). Presses Universitaires de France.
- Allal, L. (2020) Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice*, 27(4), 332-349.
- Allal, L., & Laveault, D. (2009). Assessment for learning: évaluation-soutien d'apprentissage. *Mesure et évaluation en éducation*, 32(2), 99-106.
- Allal, L., & Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. In OCDE (Éd.), *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires* (pp. 265-290). Éditions OCDE/CERI.
- Ames, C. (1992). Classroom: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261-271. <https://doi.org/10.1037/0022-0663.84.3.261>
- Archambault, J., & Chouinard, R. (2006) Doit-on récompenser les élèves pour les motiver à apprendre ? In Galand, B. & Bourgeois, E. (Éds.), *(Se) motiver à apprendre* (pp. 135-144). Presses universitaires de France.
- 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, 717-735. <https://doi.org/10.1037/edu0000307>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1982). The self and mechanisms of agency; In J. Suls (Éd.), *Psychological perspectives on the self* (Vol. 1, p.3-40). Erlbaum.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- Beauvois, J.L., Dubois, N., & Doise, W. (1999). *La construction sociale de la personne*. PUG.
- Bénit, S. (2023). Quelles sont les perceptions par les parents et les élèves des pratiques évaluatives dans les classes ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 62-74). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

- Bennett, E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-74
- Black, P., & Wiliam, D. (2018) Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575
- Bloom, B. S., Hasting, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>
- Briki, W., & Gernigon, C. (2009). Momentum psychologique : le pouvoir de l'élan. In Paquet, Y. (Éd.), *Psychologie du contrôle : théories et applications* (pp. 229-246). De Boeck.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, 30(1), 3-12.
- Brown, P.C., Roediger, H.L., & McDaniel, M.A. (2016). *Mets-toi ça dans la tête : Les stratégies d'apprentissage à la lumière des sciences cognitives*. Markus Haller.
- Brunot, S., Huguet, P., & Monteil, J.-M. (2000). Performance feedback and self-focused attention in the classroom: When past and present interact. *Social Psychology of Education*, 3, 277-293.
- Butera, F. (2011). La menace des notes. In Butera, F., Buchs, C. & Darnon, C. (Éds.), *L'évaluation, une menace ?* (pp. 45-55). Presses universitaires de France.
- Butler, A.C., Karpicke, J.D., & Roediger III, H.L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13(4), 273-281. <https://doi.org/10.1037/1076-898x.13.4.273>
- Butler, R., & Nisan, M. (1986). Effects of no-feedback, task-related comments and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78(3), 210–216. <https://doi.org/10.1037/0022-0663.78.3.210>
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474-482. <https://doi.org/10.1037/0022-0663.79.4.474>
- Butler, R. (2006). Are mastery and ability goals both adaptive? Evaluation, initial goal construction and the quality of task engagement. *British Journal of Educational Psychology*, 76(3), 595-611. <https://doi.org/10.1348/000709905x52319>
- Butera, F. (2023) Quels sont les effets psychosociaux de l'évaluation normative ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 5-12). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

- Calone, A., & Lafontaine, D. (2018). Feedback normatif vs feedback élaboré : quel impact sur la performance et le sentiment de contrôlabilité des élèves ? *Évaluer. Journal international de recherche en éducation et formation*, 4(2), 47-76. <https://hdl.handle.net/2268/231134>
- Calone, A., & Lafontaine, D. (2023). *L'impact des différents types de feedbacks en contexte de classe*. Cnesco-Cnam.
- Carroll, J. B. (1963). A Model of School Learning. *Teachers College Record*, 64(8), 1-9. <https://doi.org/10.1177/016146816306400801>
- Chervel, A. (1992). Devoirs et travaux écrits des élèves dans l'enseignement secondaire du XX^e siècle. *Histoire de l'éducation*, 54, 13-38.
- Chesné, J.-F., Piedfer-Quênay, L. & Jeanneau, F. (2023). *Limites et biais de l'évaluation : synthèse des travaux de recherche*. Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_CHESNE-PIEDFER-JEANNEAU.pdf
- Coertjens, L. (2023). Comment l'évaluation peut-elle soutenir un suivi des progrès qui permette à tous les élèves d'apprendre ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 101-114). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf
- Cosnefroy, L., & Fenouillet, F. (2009). Motivation et apprentissages scolaires. In P., Carré & F., Fenouillet (Éds.), *Traité de psychologie de la motivation*. Dunod.
- Crahay, M. (2007). Feedback de l'enseignant et apprentissage des élèves : revue critique de la littérature de recherche. In L. Allal, L. & L. Mottier-Lopez (Éds.), *Régulation des apprentissages et situation scolaire et en formation*. (pp. 45-70). De Boeck.
- Crahay, M., & Lafontaine, D. (2012). Pistes pour une école juste et efficace. In M. Crahay (Éd.) *L'école peut-elle être juste et efficace ? De l'égalité des chances à l'égalité des acquis*. (pp. 455-480). De Boeck.
- Crahay, M., Mottier Lopez, L., & Marcoux, G. (2019). L'évaluation des élèves : Docteur Jekyll and Mister Hyde de l'enseignement. In M. Crahay (Éd.), *Peut-on lutter contre l'échec scolaire ?* (pp. 358-425). De Boeck.
- Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24, 588-594. <https://doi.org/10.1177/0146167298246003>
- Croizet, J.-C., Dutrévis, M., & Désert, M. (2002). Why do students holding non-prestigious high school degrees under-achieve at the university? *Swiss Journal of Psychology*, 61, 167-175.
- Darnon, C. (2014). *Faire sa place dans le système : Enjeux relationnels, statutaires et idéologiques de la motivation en contexte éducatif*. Habilitation à Diriger des travaux de Recherches soutenue publiquement le 26 novembre 2014, Université Blaise-Pascal, Clermont-Ferrand.

Darnon, C., Smeding, A., Toczec-Capelle, M.-C., & Souchal, C. (2011). L'évaluation comme outil de formation et/ou de sélection. In Butera, F., Buchs, C. & Darnon, C. (Éds.), *L'évaluation, une menace ?* (pp. 117-126). Presses universitaires de France.

de Boer, H., Donker, A. S., Kostons, D. D., & van der Werf, G. p. (2018). Long-term effects of metacognitive strategy instruction on student academic performance: A meta-analysis. *Educational Research Review*, 24, 98-115. Flavell, J.H. (1971). What is memory development the development of? *Human Development*, 14(4), 272-278. <https://doi.org/10.1159/000271221>

Deci, E.L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6), 627-668. <https://doi.org/10.1037/0033-2909.125.6.627>

Deci, E. L. & Ryan, R. M. (2000) The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, XV(1), 25-37. doi :103917/rfia.151.0025

Depp (2020). Premiers résultats de l'enquête sur les pratiques d'enseignement, EPODE, en 2018 au collège. *Note d'information de la Depp*, 20.23, 4.

Depp (2020). Premiers résultats de l'enquête sur les pratiques d'enseignement des professeurs des écoles, EPODE, en 2018. *Note d'information de la Depp*, 20.32, 4.

Désert, M., Croizet, J.-C., & Leyens, J.-P. (2002). La menace du stéréotype : Une interaction entre situation et identité. *L'Année Psychologique*, 102, 555-576.

Désert, M., Préaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *European Journal of Psychology of Education*, 24, 207-218.

Deubel, P., Huart, J.-M., Montoussé, M., & Vin-Datiche, D. (2007). *100 Fiches pour comprendre le système éducatif*. Bréal.

Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *The Journal of Experimental Education*, 77(1), 3-18. <https://doi.org/10.3200/JEXE.77.1.3-20>

Dubet, F. (2004). *L'école des chances. Qu'est-ce qu'une école juste ?* Seuil.

Dupeyrat, C., Excribe, C., & Mariné, C. (2006). Buts d'accomplissement et qualité de l'engagement dans l'apprentissage : le coût de la compétition. In Galand, B. & Bourgeois, E. (Éds.). *(Se) motiver à apprendre* (pp. 63-74). Presses universitaires de France.

Dweck, C. S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology*, 31, 674-685.

Dweck C. S. (2002). Messages that Motivate: How Praise Molds Students' Beliefs, Motivation, and Performance (in Surprising Ways). In Aronson, J. (Ed.), *Improving academic achievement: impact of psychological factors on education* (pp. 37-60). Academic Press.

Elliot, A. J. (1997). Integrating the "classic" and "contemporary" approaches to achievement motivation: A hierarchical model of approach and avoidance motivation. In M. Maehr & p. Pintrich (Eds). *Advances in motivation and achievement* (pp. 143-179). Emerald.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34, 169-189.

Elliot, A.J., & Church, M.A. (1997). A Hierarchical Model of Approach and Avoidance Achievement Motivation. *Journal of Personality and Social Psychology*, 72(1), 218-232. <https://doi.org/10.1037/0022-3514.72.1.218>

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: a mediational analysis. *Journal of Personality and Social Psychology*, 66, 968-980.

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70, 461-475.

Elliot, A. J., & Mcgregor, H. A. (2001). A 2*2 Achievement Goal Framework. *Journal of Personality and Social Psychology*, 80, 501-519.

Fagnant, A. (2023). *Les pratiques d'évaluation en classe : des compétences professionnelles pour soutenir l'apprentissage des élèves*. Cnesco-Cnam.

Felouzis, G. (2014). *Les inégalités scolaires*. PUF, Que sais-je ?

Fenouillet, F. (2023). Quelle est la relation entre évaluation et motivation dans le cadre des apprentissages scolaires ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 75-83). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Fontaine, S., Kane, R. G., Duquette, O., & Savoie-Zajc, L. (2012). New teachers' career intentions: Factors influencing new teachers' decision to stay or to leave the profession. *Alberta Journal of Educational Research*, 57(4), 353-378.

Fontaine, S. (2023). Comment la formation des personnels peut-elle favoriser le développement d'une évaluation au service de l'apprentissage ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 219-228). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Galand, B. (2006). Pratiques d'enseignement et adaptation scolaire des élèves. In Galand, B. & Bourgeois, E. (Éds.). *(Se) motiver à apprendre* (pp. 145-158). Presses universitaires de France.

Gaonac'h, D. (2022). *Les élèves et la mémoire*. Retz, Mythes et Réalités.

Gaonac'h, D. (2023). Quels sont les apports de la psychologie cognitive pour articuler évaluation et apprentissage ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 148-157). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Garcia, S. (2018). *Le goût de l'effort. La construction familiale des dispositions scolaires*. PUF.

Genelot, S. (2023). *Observer les pratiques évaluatives des enseignants*. Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_GENELOT.pdf

Genelot, S. (2023). Quelles sont les pratiques et les représentations de l'évaluation chez les enseignants ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 46-61). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Gibbs, G., & Simpson, C. (2004). Does your assessment support your students' learning. *Journal of Teaching and learning in Higher Education*, 1(1), 1-30.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24, 645-662. <https://doi.org/10.1016/j.appdev.2003.09.002>

Gottsmann, L. (2023). Comment peut-on définir les attentes évaluatives en classes ? Une synthèse des réflexions en EPS. dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 174-181). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Goudeau, S., & Croizet, J. C. (2017). Hidden Advantages and Disadvantages of Social Class: How Classroom Settings Reproduce Social Inequality by Staging Unfair Comparison. *Psychological Science*, 28, 162-170. <https://doi.org/10.1177/0956797616676600>

Goudeau, S. (2023). En classe, quels effets peuvent avoir les interactions à dimension évaluative sur les élèves en termes d'inégalités ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 93-100). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*. 34(3), 269-290. <https://doi.org/10.1080/01443410.2013.785384>

Harrison, L. A., Stevens, C. M., Monty, A. N., & Coakley, C. A. (2006). The consequences of stereotype threat on the academic performance of white and non-white lower income college students. *Social Psychology of Education*, 9, 341-357.

Harter, S. (1999). *Distinguished contributions in psychology. The construction of the self: A developmental perspective*. Guilford Press.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>

Hazard, B. (2023). La politique éducative de la France : quelles évolutions engagées depuis 2014 et quels impacts dans les classes ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 23-45). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Huguet, p. (2006). Apprendre en groupe : la classe dans sa réalité sociale et émotionnelle (pp. 153-167). In E. Bourgeois & G. Chapelle (Eds.), *Apprendre et faire apprendre*. Paris : PUF.

Heitink, M.C., Van der Kleij, F.M., Veldkamp, B.P., Schildkamp, K., & Kippers, W.B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50-62.

Horoks, J. et Pilet, J. (2023). Quels apports de la didactique pour penser l'évaluation ? L'exemple des mathématiques. dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 182-192). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

IGÉSR. (2022). *L'enseignement en cours moyen : État des lieux et besoins* (No 2022-048 ; p. 81). <https://www.education.gouv.fr/media/113766/download>

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, 137, 691-705. <https://doi.org/10.1037/a0013834>

Joigneaux, C. (2023). Comment peut-on rendre compte des progrès des élèves aux familles ? L'exemple de l'école maternelle. dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 167-173). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Jury, M. (2023). Peut-on évaluer différemment des enfants différents ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 115-125). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Karpicke, J.D., & Roediger, H.L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968. <https://doi.org/10.1126/science.1152408>

Kim, S.-I., Lee, M.-J., Chung, Y., & Bong, M. (2010). Comparison of brain activation during norm-referenced versus criterion-referenced feedback: the role of perceived competence and performance approach goals. *Contemporary Educational Psychology*, 35(2), 141-152. <https://doi.org/10.1016/j.cedpsych.2010.04.002>

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: Historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>

Laveault, D., & Allal, L. (2016). *Assessment for learning. Meeting the challenge of implementation*. Springer.

Lee, H., Chung, H.Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, 33(2), 124-140.

Lelièvre, C. (2014). L'évaluation : une vraie question politique. Blog Educ Pros. En ligne <http://blog.educpros.fr/claudelelievre/2014/11/14/evaluation-une-vraie-question-politique/>

Lockwood, p. & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, 73, 91-103. <https://doi.org/10.1037/0022-3514.73.1.91>

Markus, H.R., & Wurf, E. (1987). The dynamic self-concept: a social psychological perspective. *Annual Review of Psychology*, 42, 38-50.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133-163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>

Martinot, D. (2001). Connaissance de soi et estime de soi : ingrédients pour la réussite scolaire. *Revue des sciences de l'éducation*, 27 (3), 483–502. <https://doi.org/10.7202/009961ar>

Martinot, D. (2004). Connaître le soi de l'élève et ses stratégies de protection face à l'échec. In Marie-Christine Toczek (Éd.), *Le défi éducatif: Des situations pour réussir* (pp. 83-116). Armand Colin. <https://doi.org/10.3917/arco.tocze.2004.01.0083>

Martinot, D. (2008). *Le soi, les autres et la société*. Presses Universitaires de Grenoble.

Merle, p. (2005). Les explications extrascolaires de la mobilisation des élèves. Contribution à une sociologie de l'expérience subjective, *Les sciences de l'éducation pour l'ère nouvelle*, 38, 2, pp.89-114.

Merle, p. (2015). L'école française et l'invention de la note. Un éclairage historique sur les polémiques contemporaines. *Revue Française de Pédagogie*, 193(4), 77-88. <https://doi.org/10.4000/rfp.4899>.

Merle, p. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives*. Presses universitaires de France.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33 (1), 49-64.

Miller, W. R., & Seligman, M. E. (1975). Depression and learned helplessness in man. *Journal of Abnormal Psychology*, 84(3), 228-238. <https://doi.org/10.1037/h0076720>

Monteil, J.-M., & Hugué, p. (2002). *Réussir ou échouer à l'école : une question de contexte ?* Presses Universitaires de Grenoble.

Morales Villabona, F. (2023). Comment les enseignants peuvent-ils partager la responsabilité de l'évaluation avec élèves ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 209-218). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Mottier Lopez, L. (2012). *La régulation des apprentissages en classe*. De Boeck.

Mottier Lopez, L. (2023). Quelles sont les nouvelles conceptions de l'évaluation en classe ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 13-22). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Müller & Arridiaux (2023). *Enquête auprès de parents d'élèves portant sur l'évaluation de leurs enfants dans l'enseignement primaire et secondaire. Résultats de l'enquête réalisée par le Crédoc à la demande du Cnesco*. Cnesco-Cnam

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development*, 49, 800-814.

Nicholls, J. G. (1979). Quality and equality in intellectual development. The role of motivation in education. *American Psychologist*, 34, 1071-1084.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328-346.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>

Normand, A., & Croizet, J.-C. (2013). Upward Social Comparison Generates Attentional Focusing When the Dimension of Comparison is Self-Threatening. *Social Cognition*, 31, 336-348. <https://doi.org/10.1521/soco.2013.31.3.336>

OECD (2019). *Résultats de TALIS 2018 (Volume I) Des enseignants et chefs d'établissement en formation à vie*. OECD Publishing. <https://doi.org/10.1787/5bb21b3a-fr>

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of formative feedback when using student derived marking criteria in peer and self-assessment, *Assessment & Evaluation in Higher Education*, 27(4), 309-323. <https://doi.org/10.1080/0260293022000001337>

Pansu, P., de Place, A. L., Bouffard, T., Blaise, F., Boissicat, N., Insel, H., Jamain, L., Leroy, N., Lima, L., Pigière, D., Pouille, J., Py, J., Schmidt- Lainé, C., Verkampt, F., & Vezeau, C (2021). *Le biais d'auto-évaluation de compétence scolaire (SchoolBias)*. ANR-16-CE28-0011.

Pansu, p. (2023). Quels sont les liens entre l'évaluation, la perception que les élèves ont d'eux-mêmes et leurs progrès ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 84-92). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Paquet, Y. (2009). Les différents construits de la notion de contrôle. In Paquet, Y. (Éd.). *Psychologie du contrôle : théories et applications*. (pp. 7-22). De Boeck.

Pasquini, R. (2023). Quels sont les enjeux et leviers pour des pratiques de notation constructives ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 136-147). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Perrenoud, p. (1999). *L'évaluation des élèves : de la fabrication de l'excellence à la régulation des apprentissages. Entre deux logiques*. Bruxelles.

Pintrich, p. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92, 544-555.

Pulfrey, C. (2011). Promotion de soi et triche. In Butera, F., Buchs, C. & Darnon, C. (Éds.), *L'évaluation, une menace ?* (pp. 177-186). Presses universitaires de France.

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why Grades engender Performance-Avoidance Goals: The Mediating Role of Autonomous Motivation. *Journal of Educational Psychology*, 103(3), 683-700. <https://doi.org/10.1037/a0023911>

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019) Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, 60, 154-165. <https://doi.org/10.1016/j.learninstruc.2018.01.004>

Rawls, J. (1987). *Théorie de la justice*. Seuil.

Roussel, S. (2023). Comment l'évaluation peut-elle favoriser la compréhension de l'oral ? L'exemple des langues vivantes étrangères. dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 193-208). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Rowland, C.A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463. <https://doi.org/10.1037/a0037559>

Sadler, D. R. (1998) Formative assessment: revisiting the territory, *Assessment in Education*, 5(1), 77–84. <https://doi.org/10.1080/0969595980050104>

Scallon, G. (1988). *L'évaluation formative des apprentissages*. Tome I : *La réflexion*. P.U.L.

Schildkamp, K., van der Kleij, F.M., Heitink, M.C., Kippers, W.B., & Veldkamp, B.P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 1-16.

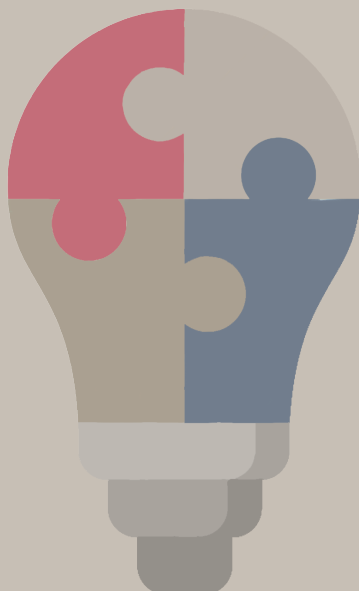
Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–356. <https://doi.org/10.1037/0033-295X.115.2.336>

Scriven, M. (1967). *Perspectives of curriculum evaluation*. Rand McNally.

- Shulman, L. S. (2007). Ceux qui comprennent : Le développement de la connaissance dans l'enseignement. *Éducation et Didactique*, 1(1), 97-114.
- Souchal, C., Toczek, M.-C., Darnon, C., Smeding, A., Butera, F., & Martinot, D. (2014). Assessing does not mean threatening: the purpose of assessment as a key determinant of girls' and boys' performance in a science class. *The British Psychological Society*. 84(1), 125-136. <https://doi.org/10.1111/bjep.12012>
- Spencer, B. & Castano, E. (2007). Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. *Social Justice Research*, 20, 418-432.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Toczek, M-C, & Souchal, C. (2017). Le pouvoir des contextes évaluatifs. *Evaluer. Journal International de Recherche en Education et Formation, e-JIREF*, 3(1-2), 21-35.
- Trunbull, E., & Gerzon, N. (2013). *Professional development on formative assessment. Insights from Research and Practice*. Retrieved from <https://www.wested.org/resources/professional-development-on-formative-assessment-insights-from-research-and-practice/>
- Vaillancourt, M. È., & Bouffard, T. (2009). Illusion d'incompétence, attitudes dysfonctionnelles et distorsions cognitives chez des élèves du primaire. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 41(3), 151-160. <https://doi.org/10.1037/a0014890>
- Vallerand, R. J., & Thill, E. E. (1993). *Introduction à la psychologie de la motivation*. Editions Etudes Vivantes.
- Viau, R. (1994). *La motivation en contexte scolaire*. De Boeck.
- Walzer, M. (2013). *Sphères de justice. Une défense du pluralisme et de l'égalité*. Seuil.
- Wiesniewski, B., Ziérer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 1-14. <https://doi.org/10.3389/fpsyg.2019.03087>
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.
- Wollenschläger M., Hattie J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubric effective in teacher feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology*, 44(45), 1-11. <https://doi.org/10.1016/j.cedpsych.2015.11.003>
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228-260.

Yerly, G. (2023). Comment la planification peut-elle permettre de passer d'une évaluation de l'apprentissage à une évaluation *pour* l'apprentissage ? dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 126-135). Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2023/03/Cnesco-CC-Eval_Notes-des-experts.pdf

Yorke, M (2003). Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice, *Higher Education*, 45(4), 477–501. <https://doi.org/10.1023/A:1023967026413>



le **cnam**
Cnesco

Centre national d'étude des systèmes scolaires

CENTRE NATIONAL D'ÉTUDE DES SYSTÈMES SCOLAIRES
CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS

41 rue Gay-Lussac - 75005 PARIS

06 98 51 82 75 - cnesco@lecnam.net

www.cnesco.fr



INSPE Institut national
supérieur du professorat
et de l'éducation
Clermont Auvergne

UNIVERSITÉ CLERMONT AUVERGNE

INSPE DE L'ACADÉMIE DE CLERMONT-FERRAND

36 avenue Jean-Jaurès - 63407 CHAMALIÈRES

04 73 31 71 50

<https://inspe.uca.fr>

RETROUVEZ LES DERNIÈRES ACTUALITÉS DU CNESCO :



www.cnesco.fr



[Cnesco](#)



[Cnesco](#)



[Cnesco](#)