

CONFÉRENCE DE CONSENSUS

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



RAPPORT SCIENTIFIQUE

**Les pratiques d'évaluation en classe :
des compétences professionnelles
pour soutenir l'apprentissage des élèves**

#CC_EVALUATION

LES 23 ET 24 NOVEMBRE 2022

En partenariat avec :



**LES PRATIQUES D'ÉVALUATION EN CLASSE :
DES COMPÉTENCES PROFESSIONNELLES POUR SOUTENIR
L'APPRENTISSAGE DES ÉLÈVES**

Annick FAGNANT
Université de Liège (Belgique)

Mars 2023

le cnam
Cnesco

Centre national d'étude des systèmes scolaires

Pour citer ce document, merci d'utiliser la référence suivante :
Fagnant, A. (2023). *Les pratiques d'évaluation en classe : des compétences professionnelles pour soutenir l'apprentissage des élèves*. 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 : **L'évaluation en classe**.

Les opinions et arguments exprimés n'engagent que l'auteure 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 75 005 Paris

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

Sommaire

Liste des encadrés « Ce que l'on peut retenir de... »	5
Liste des encadrés	5
Liste des figures.....	5
Liste des tableaux.....	5
Introduction.....	6
I. Les multiples facettes de l'évaluation en classe.....	8
A. Les fonctions de l'évaluation et les enjeux pour soutenir les apprentissages	8
B. Vers un élargissement de la vision de l'évaluation formative	13
1. Dans la littérature francophone.....	13
2. Dans la littérature anglophone	15
C. Une évaluation-soutien d'apprentissage qui conjugue évaluation formative et certains aspects de l'évaluation sommative	17
II. Quand l'évaluation en classe soutient les apprentissages des élèves	22
A. Que nous disent les articles de synthèse, les revues systématiques et les méta-analyses sur les effets de l'évaluation-soutien d'apprentissage ?	22
B. Comment les chercheurs modélisent-ils l'évaluation soutien d'apprentissage en classe ?	30
1. La modélisation proposée par l'équipe américaine.....	30
2. La modélisation proposée dans les travaux francophones.....	32
C. L'enseignant, les pairs et l'élève : trois sources importantes d'évaluation-soutien d'apprentissage en classe	35
1. Les conversations évaluatives entre l'enseignant et les élèves pendant l'apprentissage	35
2. L'évaluation mutuelle entre pairs et l'autoévaluation.....	39
III. Mettre en œuvre une évaluation-soutien d'apprentissage en classe : compétences et développement professionnel des enseignants	43
Références	48

Liste des encadrés « Ce que l'on peut retenir de... »

(1) Ce que l'on peut retenir des différentes facettes de l'évaluation en classe.....	21
(2) Ce que l'on peut retenir des articles de synthèse, revues systématiques et méta-analyses	30
(3) Ce que l'on peut retenir des modélisations de l'évaluation-soutien d'apprentissage en classe.....	34
(4) Ce que l'on peut retenir des conversations évaluatives	39
(5) Ce que l'on peut retenir des pratiques d'évaluation mutuelle entre pairs et d'autoévaluation.....	42
(6) Ce que l'on peut retenir des éléments susceptibles de favoriser la mise en œuvre d'une évaluation-soutien d'apprentissage en classe.....	47

Liste des encadrés

Encadré 1 : Quelques limites liées aux contours de ce rapport.....	7
Encadré 2 : Article de synthèse, revue systématique et méta-analyse : de quoi s'agit-il ?	22
Encadré 3 : La taille de l'effet : de quoi s'agit-il ?.....	23
Encadré 4 : Pour en savoir plus... Discours monologiques ou dialogiques	37

Liste des figures

Figure 1 : L'alignement curriculaire élargi (adapté de Pasquini, 2019, p. 63 – Extrait de Goffin, Baron, Ochelen & Fagnant, 2022, p. 6).....	12
Figure 2 : <i>Assessment for learning</i> : sa fonction formative et son interface avec l'évaluation sommative (traduit de Laveault & Allal, 2016b, p. 8 – d'après les travaux d'Allal publiés au début des années 2010)	19
Figure 3 : Modélisation de l'ESA proposée par l'équipe américaine (d'après Wiliam & Thompson, 2008 – traduit de Wiliam, 2011, p. 12)	31
Figure 4 : Modélisation du processus de co-régulation dans une situation de classe (d'après Allal, 2007b – traduit d'Allal, 2020, p. 340).....	33

Liste des tableaux

Tableau 1 : Les fonctions de l'évaluation (d'après Allal, 1991)	8
Tableau 2 : Perspective classique <i>versus</i> élargie de l'évaluation formative (inspiré de Allal & Mottier Lopez, 2005, p. 269)	13
Tableau 3 : Caractéristiques de <i>l'assessment of, for</i> et <i>as learning</i> (traduit de Earl, 2003, p. 6).....	16
Tableau 4 : Une vision plus nuancée des liens entre les buts et les types d'évaluations (traduit de Bennett, 2011, p. 8).....	17
Tableau 5 : Description de quelques éléments centraux du programme de formation développé dans l'étude de Kruze <i>et al.</i> (adapté de Kruze <i>et al.</i> , p. 114).....	45

Introduction

Les écrits scientifiques récents autour de l'évaluation en classe montrent un dépassement de l'opposition classique entre évaluations « sommative » et « formative » pour aller vers une vision qui met en évidence leur complémentarité pour soutenir les apprentissages des élèves. On retrouve une telle perspective tant dans les travaux francophones (par exemple Crahay, Mottier Lopez & Marcoux, 2019) qu'anglophones (par exemple Black & Wiliam, 2018a) ou encore dans un ouvrage au titre évocateur *Assessment for learning. Meeting the challenge of implementation* (Laveault & Allal, 2016a), qui compile une série de recherches internationales et dont le modèle proposé nous servira de point de référence.

La question qui guide ce rapport pourrait être formulée comme suit : « Comment l'évaluation en classe peut-elle soutenir les apprentissages des élèves ? ». On parlera d'évaluation formative au sens élargi (ou des fonctions formatives de l'évaluation en classe) ou d'une évaluation-soutien d'apprentissage (*assessment for learning* dans la littérature anglo-saxonne).

Pour définir les contours d'une « évaluation-soutien d'apprentissage », il conviendra d'évoquer brièvement en quoi l'évaluation peut aussi jouer d'autres rôles importants pour l'école ou la société (comme de communiquer un bilan des acquis vers les parents, de certifier la réussite d'une année scolaire ou encore de sélectionner des étudiants à l'entrée d'un cursus de formation) et en quoi elle peut parfois avoir des effets délétères sur l'apprentissage (comme les effets pervers d'une évaluation normative, par exemple). Ensuite, comme les pratiques d'évaluation qui visent à soutenir les apprentissages des élèves sont nombreuses et que les terminologies pour les qualifier sont également variées, tant dans le monde francophone que dans la littérature anglophone, il nous paraît tout aussi essentiel d'apporter quelques clarifications conceptuelles. Nous évoquerons ainsi la vision francophone de l'évaluation formative élargie (Allal & Mottier Lopez, 2005) et les travaux anglo-saxons qui distinguent notamment l'évaluation *de, pour* et *en tant qu'* apprentissage (*assessment of, for* et *as learning*, Earl, 2003). Enfin, en nous appuyant sur le modèle de l'*assessment for learning*, tel que conceptualisé par Laveault et Allal (2016a,b), nous chercherons à mettre en évidence les différentes facettes que peut comporter l'évaluation en classe pour soutenir les apprentissages des élèves.

Après cette première partie conceptuelle, la deuxième partie de ce rapport cherchera à faire le point sur les recherches empiriques qui se sont attelées à mettre en évidence l'efficacité de l'évaluation formative (ou de l'*assessment for learning*) en général, ou plus spécifiquement de diverses pratiques relevant de cette fonction formative de l'évaluation (conversations évaluatives, évaluations entre pairs ou et autoévaluation). Comme les écrits sur la question sont extrêmement nombreux, nous avons fait le choix de chercher à nous appuyer, autant que faire se peut, sur des articles de synthèse, des revues systématiques et des méta-analyses¹.

Au terme de leur revue des écrits francophones sur l'évaluation formative, Allal et Mottier Lopez (2005) pointaient les avancées conceptuelles réalisées, mais regrettaient le manque d'études empiriques sur ces questions. Si le constat n'est sans doute plus tout à fait vrai aujourd'hui, force est de constater que c'est presque exclusivement dans la littérature anglo-saxonne que nous avons trouvé des travaux de synthèse et des méta-analyses. Nous nous appuyons essentiellement sur ces travaux

¹ L'Encadré 2 (p. 22) explicite ce que sont les articles de synthèse, les revues systématiques et les méta-analyses.

parce que référencer l'ensemble des écrits sur la question constituerait une entreprise impossible dans le cadre de ce travail.

Pour que l'évaluation en classe puisse soutenir les apprentissages des élèves, les enseignants doivent disposer de compétences professionnelles en évaluation ; compétences qu'ils devront pouvoir conjuguer harmonieusement avec leurs compétences professionnelles en matière didactique notamment. Il nous paraît dès lors pertinent de terminer cette note de synthèse par une partie consacrée à ces compétences professionnelles et à leur développement. Après avoir exploré les conditions propices à la mise en œuvre d'une évaluation formative en classe (Heitink *et al.*, 2016 ; Schildkamp *et al.*, 2020 ; Yan *et al.*, 2021) et défini le concept d'« *assessment literacy* » (Mertler, 2004), nous évoquerons quelques recherches empiriques qui se sont attelées à mesurer l'efficacité de programmes de développement professionnel propices au développement des compétences évaluatives (Mertler, 2009 ; Kruse *et al.*, 2020).

Nous avons fait le choix de ne pas proposer de partie conclusive à ce rapport pour préférer développer des synthèses partielles au fil du texte (voir les parties intitulées « Ce que l'on peut retenir de... »).

Encadré 1 : Quelques limites liées aux contours de ce rapport

- Il ne s'agit pas d'une revue systématique. Certains choix ont été posés tant au niveau de l'organisation du texte que des modèles retenus et des écrits sélectionnés. Dans la partie I. du texte, nous avons tenté de conjuguer les apports théoriques des travaux francophones et anglophones. Pour développer les parties II. et III. du texte, le choix de privilégier les revues systématiques et les méta-analyses a conduit à sélectionner essentiellement des écrits anglophones.
- Ce rapport porte un regard global sur l'évaluation, sans prendre en compte les revues de la littérature qui portent sur des disciplines scolaires spécifiques. Nous verrons pourtant que la place des connaissances didactiques est pointée dans plusieurs travaux comme un élément essentiel au développement d'une évaluation-soutien d'apprentissage efficace.
- Les feedbacks font évidemment partie intégrante des éléments à prendre en compte pour que l'évaluation puisse jouer un rôle formatif et ainsi permettre une régulation des apprentissages. Nous évoquerons évidemment le rôle des feedbacks lorsque nous ferons état des recherches sur l'évaluation formative, mais nous ne consacrerons pas un point spécifique à ceux-ci dans les développements ultérieurs pour nous consacrer à d'autres éléments impliqués dans l'évaluation en classe. Pour en savoir plus sur les feedbacks eux-mêmes, nous renvoyons le lecteur intéressé au rapport réalisé par Calone et Lafontaine (2023) dans le cadre de cette même conférence de consensus 2022 du Cnesco puisqu'il s'agit là du cœur même de leur synthèse.

I. Les multiples facettes de l'évaluation en classe

Après avoir rappelé les fonctions classiques de l'évaluation et précisé quelques caractéristiques que devrait remplir l'évaluation sommative pour jouer un rôle de soutien aux apprentissages des élèves, nous développerons plus en profondeur la notion d'évaluation formative et les élargissements que ce concept a connu dans les mondes francophone et anglophone. Nous terminerons cette première partie en présentant le modèle « d'évaluation-soutien d'apprentissage » (*assessment for learning*, Allal & Laveault, 2009 ; Laveault & Allal, 2016a,b) qui sous-tendra la suite des travaux.

A. Les fonctions de l'évaluation et les enjeux pour soutenir les apprentissages

Depuis les travaux d'Allal au début des années 1990, il est courant de distinguer trois fonctions d'évaluation qui envisagent des formes de régulation et conduisent à des décisions différentes (Allal, 1991, 2007, Crahay *et al.*, 2019)². Le Tableau 1 en propose une vue d'ensemble³.

Tableau 1 : Les fonctions de l'évaluation (d'après Allal, 1991)

Fonction de l'évaluation	Moment	Formes de régulation	Décision à prendre
Formative	Pendant une période de formation	Assurer que les activités d'enseignement-apprentissage correspondent aux caractéristiques des élèves	Adaptation des activités d'enseignement-apprentissage
Sommative	À la fin d'une période de formation	Assurer que les caractéristiques des élèves répondent aux exigences du système scolaire	Certification intermédiaire ou finale
Pronostique	Au début d'un cycle de formation		Admission ou orientation

L'évaluation formative intervient pendant une période de formation, à la suite d'une unité d'enseignement/apprentissage portant sur un ensemble restreint d'objectifs. L'enjeu de l'évaluation formative est de favoriser la maîtrise des objectifs par le plus grand nombre d'élèves. Pour ce faire, elle devrait avoir une visée diagnostique en vue d'identifier les erreurs et les lacunes des élèves. La régulation consiste ici à s'assurer que les moyens de formation correspondent aux caractéristiques des élèves, ce qui va nécessiter des décisions en termes d'adaptation des activités d'enseignement et d'apprentissage (Allal, 2007a).

² Quoique courante, cette catégorisation ne fait pas consensus. Ainsi, par exemple, Merle (2018) distingue l'évaluation diagnostique de l'évaluation formative ou formatrice, d'une part, et l'évaluation sommative de l'évaluation certificative, d'autre part. Selon cet auteur, « l'évaluation diagnostique a pour objet de connaître le niveau des élèves en début d'année scolaire ou au début d'une nouvelle séquence d'apprentissage » (p. 21) alors que « l'évaluation formative est un processus d'évaluation continue ayant pour objet d'assurer la progression de chaque individu dans une démarche d'apprentissage » (p. 22). L'évaluation formative serait centrée sur l'activité de l'élève et en ayant pour objet principal « de favoriser des processus d'autoévaluation » (p. 24). Les évaluations sommatives et certificatives visent toutes deux à faire un bilan des connaissances et compétences acquises par les élèves, mais « la particularité de l'évaluation certificative *serait* d'être réalisée dans le cadre d'un examen tel que le diplôme national du brevet ou le baccalauréat pour l'enseignement secondaire » (p. 18).

³ Le tableau 1 et le descriptif des trois fonctions s'inspirent largement du syllabus de « didactique générale » de Fagnant (2021).

Deux pans complémentaires sont alors généralement convoqués :

- (a) la régulation de l'enseignement [qui] désigne le processus par lequel l'enseignant décide d'adapter et de différencier son enseignement à partir des résultats produits par l'évaluation formative (dans des modalités formelles et informelles) ;*
- (b) la régulation des apprentissages de l'élève qui, étant activement impliqué dans des évaluations formatives, développe des capacités d'autorégulation favorables à son apprentissage (Crahay et al., 2019, p. 360).*

Une autre forme de régulation consiste à s'assurer que les caractéristiques des élèves correspondent aux exigences du système scolaire ; il s'agit cette fois de prendre des décisions administratives qui vont avoir un impact sur le cursus scolaire des élèves (Crahay et al., 2019). Deux fonctions différentes de l'évaluation entrent ici en jeu : la fonction sommative et la fonction pronostique.

L'évaluation sommative intervient au terme d'une période de formation (à la fin d'une unité d'enseignement/apprentissage) et permet à l'enseignant de dresser un bilan des acquis des élèves. L'évaluation sommative conduit généralement à attribuer une note (chiffrée ou non) à la performance observée. Selon Allal (2007a), cette forme d'évaluation présente deux caractéristiques institutionnelles importantes qui la distinguent de l'évaluation formative :

- (a) elle est enregistrée dans un document officiel ayant un caractère public et permanent (un carnet, livret ou bulletin scolaire) ;*
- (b) elle est communiquée à des interlocuteurs en dehors de la relation pédagogique qui lie l'enseignant à l'élève, par exemple aux parents de l'élève, aux instances de formations ultérieures, voire aux employeurs (Allal, 2007a, p. 141).*

En théorie, la procédure qui suit l'évaluation sommative ne poursuit donc plus aucun dessein pédagogique ; la forme de régulation qui la concerne est une « régulation de l'action de certification assurée par l'école face aux attentes de la société » (Allal, 2007a, p. 142). Dans cette perspective, l'évaluation certificative est donc une forme d'évaluation sommative qui vise la délivrance d'un diplôme ou d'un certificat. On peut considérer que les évaluations sommatives réalisées en cours d'année par un enseignant dans sa classe constituent des étapes intermédiaires à cette certification dans la mesure où la délivrance du « certificat » permettant le passage dans la classe supérieure s'appuie souvent, au moins partiellement, sur ces notes.

L'évaluation pronostique intervient quant à elle lorsqu'il s'agit d'admettre les élèves dans un cursus de formation ou de les orienter vers un nouveau cursus ; elle estime leurs chances de réussite dans la formation à venir. Les examens d'entrée constituent des exemples de ce type d'évaluation. Toutefois, selon Allal (2007a),

- Dans le cadre scolaire, les décisions les plus courantes sont celles qui concernent la promotion ou le redoublement, le passage du secteur ordinaire dans une classe ou institution spécialisée, l'orientation vers des filières de l'école secondaire⁴, l'accès à des dispositifs de formation ayant des exigences particulières (Allal, 2007a, p. 142).*

⁴ L'école secondaire correspond au collège et au lycée en France.

Cette fonction de l'évaluation « porte sur l'avenir mais se base en bonne partie, au sein de la scolarité obligatoire, sur les résultats d'évaluations sommatives antérieures » (p. 142). Cette fonction pronostique est souvent gérée de manière collégiale, lors des conseils de classe notamment. Tout comme pour les évaluations sommatives, les décisions sont inscrites dans des documents officiels accessibles à divers destinataires (Allal, 2007a).

À côté de ces trois fonctions, il est aussi important de préciser les caractéristiques que peuvent rencontrer les évaluations. L'évaluation est dite à **référence normative** quand elle réfère la performance d'un apprenant aux performances des autres apprenants. Elle situe donc l'élève non pas par rapport aux objectifs à atteindre mais par rapport aux résultats des autres élèves, ce qui autorise une procédure de classement ou de sélection. Cette façon de procéder est caractéristique d'une tradition visant à mettre en exergue l'excellence scolaire dans un système méritocratique (Crahay *et al.*, 2019) ; ses effets délétères sont largement connus et détaillés dans la littérature⁵. *A contrario*, l'évaluation à référence **critériée** réfère quant à elle la performance d'un apprenant à des critères précis, établis en lien avec des objectifs de formation. Son but est ainsi de décrire ce dont un élève est capable, indépendamment de la performance des autres élèves.

Crahay *et al.* (2019) explicitent clairement comment la **référence normative** utilisée dans un système qui prône l'excellence scolaire conduit certains enseignants à conférer aux évaluations sommatives une **visée pronostique** :

*Il n'est guère étonnant que, dans la culture de l'excellence et de l'échec qui habite certains systèmes éducatifs, de nombreux enseignants conçoivent que l'évaluation des travaux d'élèves doive se traduire par un classement de ceux-ci. Se croyant tenus de produire une hiérarchisation des élèves, ces enseignants sont inéluctablement amenés à privilégier les questions discriminatives et à créer pour chaque épreuve une échelle – souvent artificielle – de valeurs qui, idéalement, débouchera sur une distribution gaussienne des notes. Cette évaluation est pronostique [...] On peut dès lors craindre qu'en fin d'année, ces enseignants ne se soucient que très partiellement d'évaluation sommative et que leurs décisions de promotion ou de redoublement ne se fondent que partiellement sur un bilan des acquisitions réalisées durant l'année écoulée. Ils s'attachent à anticiper dans quelle mesure chaque élève est apte à suivre les enseignements de l'année à venir [...] L'évaluation pronostique est par essence sélective : s'interroger sur les élèves aptes à suivre avec succès un degré d'enseignement ultérieur, c'est nécessairement suspecter que certains n'en sont pas capables (Crahay *et al.*, 2019, p. 363).*

A contrario, si l'on souhaite que l'évaluation serve réellement à soutenir les apprentissages des élèves, les **évaluations sommatives** se doivent d'être à **référence critériée** et **articulées aux évaluations formatives**. Crahay *et al.* (2019) opposent ainsi les évaluations normatives situées dans une visée d'excellence aux évaluations critériées situées dans l'optique d'une pédagogie de la maîtrise :

Les partisans de la pédagogie de maîtrise prennent le contre-pied de cette conception méritocratique de l'école. Pour eux, la mission de l'enseignant n'est pas de fabriquer des hiérarchies d'excellence, mais de susciter un maximum d'apprentissages chez un maximum d'élèves. L'évaluation sommative se doit d'être à référence critérielle ou critériée (Glaser, 1963 ; de Landsheere, 1992, p. 301), c'est-à-dire devra situer l'élève par rapport aux compétences à

⁵ Pour une revue critique des « notes » et d'une évaluation normative, on renverra le lecteur intéressé au rapport proposé par Calone & Lafontaine (2023) dans le cadre de cette même conférence de consensus du Cnesco.

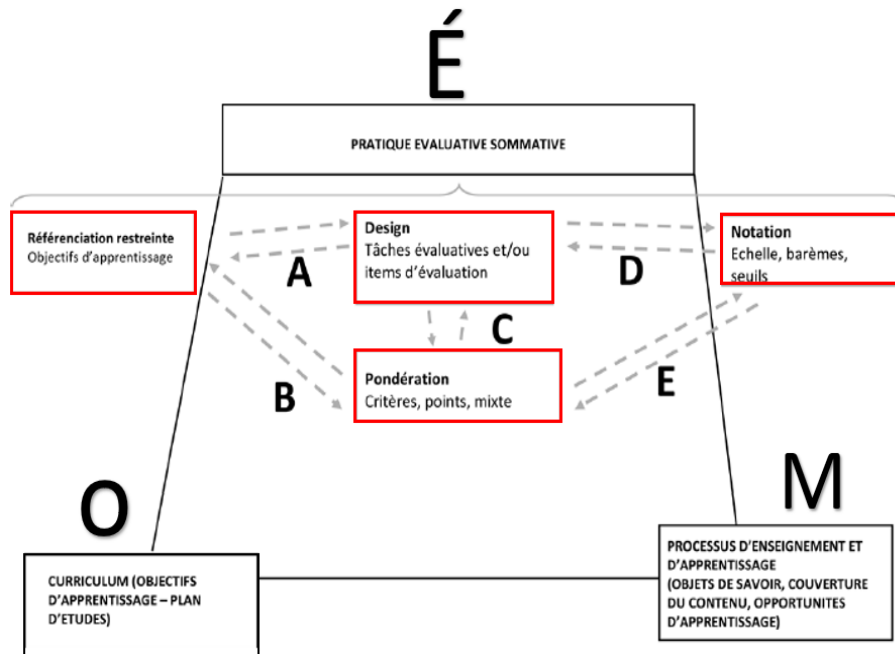
maîtriser. Plus important encore ! Pour atteindre la finalité de la pédagogie de maîtrise – c'est-à-dire pour amener la quasi-totalité des élèves à un même niveau élevé de réussite, et ce malgré leurs différences initiales –, il faut nécessairement des adaptations fréquentes du processus d'enseignement et d'apprentissage. En cours de processus, il convient dès lors d'utiliser l'évaluation à des fins formatives. En définitive, évaluation formative et évaluation sommative à référence critériée doivent être articulées (Crahay et al., 2019, pp. 363-364).

Ajoutons encore que **les évaluations sommatives peuvent aussi servir à des fins formatives**. Début des années 1990, pour pallier la critique du manque de temps souvent évoquée par les enseignants pour renoncer aux évaluations formatives, Allal (1991) proposait déjà d'insérer des activités de régulation à la suite d'un contrôle effectué d'abord dans une optique sommative. Ainsi, elle précisait que « greffer » une fonction formative sur des évaluations sommatives permet deux types de régulation : (a) un retour aux objectifs non maîtrisés de l'unité de formation que l'on vient d'évaluer (régulation qualifiée alors de rétroactive) et/ou (b) une différenciation des activités d'enseignement/apprentissage abordées dans l'unité suivante (régulation alors qualifiée de proactive). Toutefois, pour que les évaluations sommatives puissent soutenir les apprentissages et informer correctement les enseignants et les élèves, ajoutons qu'en plus d'être à référence critériée, il convient qu'elles s'inscrivent dans un processus qui respecte l'alignement curriculaire (Anderson, 2002), c'est-à-dire une triple concordance entre objectifs, méthode et évaluation (Tyler, 1949, cité par Leclercq, 2008).

S'intéressant spécifiquement à la valeur potentiellement informative de la note, Pasquini (2019) propose un modèle « élargi » de l'alignement curriculaire qui déploie quatre composantes permettant d'appréhender finement la problématique de la notation : (a) la *référenciation restreinte* qui renvoie aux objectifs d'apprentissage à évaluer ; (b) le *design* qui correspond à la manière dont les enseignants construisent leurs tâches évaluatives en fonction des contenus et des habiletés cognitives visées ; (c) la *pondération* qui représente le poids attribué aux différents objectifs évalués à l'aide de critères et points notamment et (d) la *notation* qui fait référence à la manière dont les enseignants construisent la note en s'appuyant sur des échelles ou des barèmes par exemple.

La Figure 1 ci-dessous présente une adaptation de ce modèle ; nous avons choisi de le présenter en lien avec le modèle classique de triple concordance qui met en cohérence les pratiques d'évaluation sommative (É) avec les objectifs d'apprentissage évalués (O) et les méthodes d'enseignement/apprentissage (M).

Figure 1 : L'alignement curriculaire élargi
 (adapté de Pasquini, 2019, p. 63 – Extrait de Goffin, Baron, Ochelen & Fagnant, 2022, p. 6)



Les quatre composantes développées par Pasquini (2019) sont indiquées en rouge dans le schéma et mises en relation par les doubles flèches en pointillés qui modélisent les liens réciproques les unissant (relations A-B-C-D et E). Comme le précisent Goffin *et al.* (2022) dans leur article s'intéressant aux freins et leviers pour œuvrer à la construction et à la communication d'une note à haute valeur informative,

Le degré d'information que la note véhicule découle de la cohérence (ou de l'incohérence) de ces relations, mais aussi de la façon dont l'épreuve s'insère dans le modèle plus large, prenant notamment en compte les opportunités d'apprentissages offertes aux élèves avant l'évaluation (Goffin et al., 2022, p. 7).

En accord avec ces derniers, on considèrera alors que le « modèle de l'alignement curriculaire élargi » (Pasquini, 2019) permet d'interroger la valeur informative de la note en s'intéressant à la façon dont cette dernière est construite, mais aussi à ce qui se passe en amont et en aval de l'évaluation. En amont de l'évaluation, il est important que l'enseignant communique les objectifs d'apprentissage aux élèves et que ces derniers soient informés des critères qui caractérisent la maîtrise de ces objectifs. En aval de l'évaluation, les commentaires qui accompagnent les résultats devraient aider les élèves à cerner leurs forces et leurs lacunes, tout en leur proposant des pistes pour remédier aux difficultés qu'ils rencontrent⁶.

⁶ Pour une revue critique des différents types de « feedbacks », on renverra le lecteur intéressé au rapport proposée par Calone & Lafontaine (2023) dans le cadre de cette même conférence de consensus du Cnesco.

B. Vers un élargissement de la vision de l'évaluation formative

Les travaux sur l'évaluation formative se sont déployés sous des formes diverses dans le monde francophone et dans la littérature anglo-saxonne. Les points suivants rendent compte de ces travaux.

1. Dans la littérature francophone

Le texte publié en 2005 par Allal et Mottier Lopez dans le cadre d'une conférence internationale OCDE/CERI⁷ propose une synthèse éclairante des travaux francophones et de l'élargissement du concept au cours du temps⁸. Le Tableau 2 propose une mise en perspective de la vision initiale de l'évaluation formative (EF), telle qu'imaginée par Bloom dans les années 1970, et de la vision élargie, telle qu'on la considère généralement aujourd'hui.

**Tableau 2 : Perspective classique versus élargie de l'évaluation formative
(inspiré de Allal & Mottier Lopez, 2005, p. 269)**

Perspective « classique » (Bloom)	Perspective élargie
Insertion de l'EF après chaque phase d'enseignement/apprentissage	Intégration de l'EF 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'EF par l'enseignant	Participation active des élèves dans l'EF (autoévaluation, évaluation mutuelle, co-évaluation)

La perspective actuelle propose ainsi quatre types d'ouvertures que nous allons brièvement décrire dans les lignes qui suivent. Cette version élargie de l'évaluation formative préfigure la vision de l'évaluation-soutien d'apprentissage (*Assessment for Learning*) telle qu'envisagée par Laveault et Allal (2016b) et telle que nous la développerons au point C.

a. Première ouverture : les moments durant lesquels pratiquer l'évaluation formative

Dans la perspective « classique » héritée de Bloom, l'évaluation formative se réalise au terme de chaque étape d'enseignement/apprentissage, pour s'assurer que les élèves maîtrisent les objectifs liés à une unité d'enseignement avant de passer à la suivante. La perspective « élargie » quant à elle « prône l'intégration de l'évaluation formative dans chaque activité d'enseignement/apprentissage » (Allal & Mottier Lopez, 2005, p. 268). En ce sens, plutôt que d'être une activité dissociée de l'enseignement lui-même, l'évaluation en fait partie intégrante, ce qui est indissociable de la deuxième ouverture liée aux moyens de recueillir l'information.

⁷ Le CERI est le Centre pour la Recherche et l'Innovation dans l'Enseignement (Centre for Educational Research and Innovation) de l'OCDE (Organisation de Coopération et de Développement Economique) – Pour en savoir plus, voir <https://www.oecd.org/fr/education/ceri/>

⁸ Le texte qui suit s'appuie largement sur le texte d'Allal & Mottier Lopez (2005) ainsi que sur les notes de cours de « didactique générale » de Fagnant (2021).

b. Deuxième ouverture : l'utilisation de divers moyens de recueil d'information

Puisqu'elle s'insère au sein même du processus d'enseignement/apprentissage, l'évaluation formative va nécessairement prendre des modalités diversifiées. Outre les traditionnelles épreuves « papier-cayon » (que l'on pourrait qualifier d'évaluations formatives « formelles » ou « instrumentées »), l'évaluation se fonde également sur des observations directes de l'enseignant lorsque les élèves travaillent seuls ou en petits groupes par exemple. Elle peut s'appuyer sur l'observation des démarches produites par les élèves lorsqu'ils réalisent un exercice en classe, sur l'observation des interactions entre élèves lors de travaux en duo ou petits groupes ou encore sur des « interactions collectives permettant aux élèves d'exposer différentes façons de comprendre une tâche ou d'effectuer une activité » (Allal & Mottiez Lopez, 2005, p. 269). Dans cette optique, l'évaluation formative devient davantage « informelle » (Allal & Mottier Lopez, 2005) et peut aussi être qualifiée de « non instrumentée » (Morissette, 2013). Elle va impliquer la mise en œuvre de différentes formes de régulations, notamment interactives, ce qui conduit à la troisième ouverture.

c. Troisième ouverture : du concept de remédiation à celui de régulation

Dès les années 1980, Allal (voir Allal, 2007a) a distingué trois types de régulations associées à cette perspective élargie de l'évaluation formative :

- La **régulation rétroactive** consiste à « remédier aux erreurs et aux lacunes ciblées lors de l'évaluation qui a été proposée à la fin d'une unité de formation » (p. 141). Elle correspond à l'idée de remédiation dans la perspective « classique » de Bloom.
- La **régulation proactive** est celle qui permet de guider la mise en place de nouvelles situations d'enseignement/apprentissage en différenciant les activités en fonction des besoins des apprenants. Il peut ainsi s'agir de proposer « non seulement des remédiations pour les élèves en difficulté mais aussi des activités adaptées aux besoins des élèves avancés, ou encore des tâches différentes en fonction des difficultés des élèves » (p. 141).
- La **régulation** est qualifiée d'**interactive** lorsqu'elle « résulte des interactions de l'élève avec l'enseignant, avec ses pairs, avec le matériel didactique pendant le déroulement de la situation d'enseignement/apprentissage, contribuant ainsi à l'élaboration des compétences et des connaissances en cours de construction » (p. 141). Les interactions entre l'enseignant et les élèves lors des exercices individuels ou des travaux de groupes constituent des occasions privilégiées de régulation interactive. L'enseignant peut directement repérer les démarches et les difficultés des élèves et tenter d'y faire face dès qu'elles apparaissent en les invitant à vérifier leur démarche, en leur proposant un indice, en leur donnant une explication complémentaire,...

Mottier Lopez (2012) pointe deux dangers de sur-généralisation intrinsèquement liés à cette perspective : l'un consiste à considérer que « toute intervention de l'enseignant est régulation » (p. 31) et l'autre à « considérer que l'enseignant "régule l'élève" de façon directe » alors qu'il s'agit de « sources potentielles de régulation qui ont une fonction de médiation à l'autorégulation de l'élève » (p. 31). De façon complémentaire, avec le concept de « co-régulation », Allal (2011) considère que la régulation externe peut faire office « de tremplin » à la régulation interne de l'élève. Pour le dire autrement, les interventions de l'enseignant ne constituent que des sources potentielles de régulation

et il est essentiel d'impliquer l'élève dans le processus évaluatif pour qu'il puisse, *in fine*, tirer réellement parti de ces interventions pour améliorer ses apprentissages. Ceci nous conduit dès lors vers la dernière ouverture proposée par cette perspective « élargie ».

d. Quatrième ouverture : l'implication de l'élève dans le processus d'évaluation

Dans la version « classique », l'évaluation est intrinsèquement dans les mains de l'enseignant qui construit l'évaluation, l'administre à ses élèves, analyse les résultats, puis propose des activités de remédiations adaptées. La version « élargie » implique davantage l'élève dans le processus évaluatif⁹ selon trois modalités couramment distinguées dans la littérature : « **l'autoévaluation** au sens strict, **l'évaluation mutuelle entre pairs** et la **co-évaluation** stipulant une confrontation des évaluations réalisées par l'enseignant et par l'élève » (Allal & Mottier-Lopez, 2005, p. 276). Nous développerons ces modalités d'évaluation dans la suite du texte (partie II., C.).

2. Dans la littérature anglophone

Dans un certain nombre d'écrits anglophones, on retrouve une distinction entre *assessment of learning* (évaluation de l'apprentissage), *assessment for learning* (évaluation pour l'apprentissage) et *assessment as learning* (évaluation en tant qu'apprentissage). Cette distinction est souvent proposée en référence aux travaux de Earl (2003). D'après Black et Wiliam (2018), c'est une distinction que l'on retrouve fréquemment en Australie et au Canada (ce qui explique qu'on trouve aussi cette typologie dans certains écrits francophones, québécois notamment) et qui correspond globalement aux cas de figure suivants :

Assessment of learning fait référence à l'utilisation de l'évaluation pour décider dans quelle mesure les élèves ont atteint les objectifs d'apprentissage ;

Assessment for learning correspond au processus mis en œuvre par l'enseignant pour utiliser les résultats de l'évaluation en vue d'informer [et de réguler] son enseignement ;

Assessment as learning est utilisé dans un sens très spécifique pour décrire le rôle de l'élève pour monitorer et diriger son propre apprentissage (Black & Wiliam, 2018, p. 553).

Il est assez tentant d'associer le concept d'*assessment of learning* à l'évaluation sommative ; celui d'*assessment for learning* à l'évaluation formative et celui d'*assessment as learning* à l'évaluation formatrice¹⁰. Le tableau synthétique de Earl (2003) résume ces éléments, mais l'auteure va encore plus loin en considérant l'évaluation de l'apprentissage dans une perspective normative puisque ce sont les autres élèves qui constituent le point de comparaison.

⁹ En ce sens, on peut considérer que la vision élargie de l'évaluation formative intègre en partie les éléments distingués par Merle (2018) sous le couvert d'évaluation « formatrice ».

¹⁰ La distinction proposée ici entre *assessment for* et *as learning* correspond assez bien à la distinction proposée par Merle (2018) entre évaluation formative (centrée sur l'activité de l'enseignant) et formatrice (centrée sur l'activité de l'élève) (voir note 2, p. 8). Comme pointé précédemment, une vision « élargie » de l'évaluation formative permet toutefois d'intégrer les deux concepts.

Tableau 3 : Caractéristiques de l'*assessment of, for et as learning* (traduit de Earl, 2003, p. 6)

Approche	Objectif	Point de référence	Évaluateur principal
Assessment of learning	Jugements concernant le rang (classement), le passage dans un niveau supérieur, les diplômes, etc.	Les autres élèves	L'enseignant
Assessment for learning	Informations utiles pour les prises de décisions pédagogiques par les enseignants	Critères ou attentes externes	L'enseignant
Assessment as learning	Autocontrôle et autocorrection ou ajustement	Objectifs personnels et critères externes	L'étudiant

Même si elle semble *a priori* éclairante, cette typologie est critiquée pour plusieurs raisons (Black & Wiliam, 2018) : (a) l'expression *assessment as learning* est potentiellement confusionnelle parce que l'évaluation est bien de l'évaluation et n'est pas à confondre avec l'apprentissage lui-même ; (b) la typologie ne rend pas compte du rôle des pairs dont on verra pourtant l'importance dans les différents développements ultérieurs et (c) la permutation des termes obscurcit les choses plutôt que d'apporter un réel éclairage.

Revenons sur le premier point. Black et Wiliam (2018, voir aussi Bennett, 2011, Wiliam, 2018) insistent sur l'importance de considérer l'évaluation comme une évaluation (*assessment as assessment*), c'est-à-dire, une procédure permettant de faire des inférences à propos des apprentissages effectifs des élèves. En effet, « l'obtention, l'identification et l'interprétation des informations recueillies¹¹ sont des composantes indispensables d'un enseignement efficace » (Black & Wiliam 2018, p. 561) ; ce processus pouvant intervenir aussi bien au départ de travaux écrits, que de dialogues professeur-élèves ou encore de feedbacks entre pairs dans une activité d'évaluation mutuelle. Notons par ailleurs que la définition extensive de l'évaluation formative donnée par ces mêmes auteurs (Black & Wiliam, 1998a,b) intègre également le rôle de l'élève :

On utilise le terme général d'évaluation pour désigner toutes les activités conduites par les enseignants – et par leurs élèves lorsqu'ils s'autoévaluent – qui fournissent des informations destinées à être utilisées en tant que feedback pour ajuster les activités d'enseignement et d'apprentissage. Une telle évaluation devient formative quand les informations recueillies sont effectivement utilisées pour adapter l'enseignement aux besoins des élèves (Black & Wiliam, 1998b, p. 14 ; voir aussi Wiliam, 2011).

Concernant le dernier point, Bennett (2011) critique le choix réalisé par de nombreux auteurs de remplacer « évaluation formative » par « *assessment for learning*¹² » et « évaluation sommative » par « *assessment of learning* ». Pour lui, non seulement cette permutation des termes n'apporte aucune clarification conceptuelle, mais elle pose aussi un problème de fond « car elle exempte l'évaluation sommative de toute responsabilité en matière de soutien aux apprentissages » (p. 7). En cohérence avec la vision que nous avons développée précédemment, Bennett (2011) considère que les deux fonctions sont interreliées : une évaluation sommative bien conçue et critériée devrait donner à l'enseignant des informations utiles pour cibler les besoins d'apprentissage des élèves et une

¹¹ En anglais, le terme « *evidence* » est souvent utilisé pour évoquer les « traces » sur lesquelles peut s'appuyer le jugement interprétatif de l'enseignant. Nous avons choisi ici de le traduire par « informations » ou « informations recueillies ».

¹² Notons également que puisque Earl (2003) évoque des « critères ou attentes externes » comme point de référence à l'*assessment for learning*, on peut considérer que cela fait aussi référence à l'évaluation sommative.

évaluation formative bien menée devrait « compléter les jugements informels globaux de l'enseignant concernant la réussite des élèves » (p. 7). Le tableau suivant propose une vision plus nuancée des relations entre les types d'évaluation et les buts qu'elles remplissent.

Tableau 4 : Une vision plus nuancée des liens entre les buts et les types d'évaluations (traduit de Bennett, 2011, p. 8)

Type d'évaluation	Objectif	
	Assessment <i>of</i> learning	Assessment <i>for</i> learning
Sommative	X	x
Formative	x	X

Note : X = objectif principal, x = objectif secondaire

Finalement, pour Black et Wiliam (2018), « la distinction entre évaluation formative et évaluation sommative devient une distinction dans le type d'inférences réalisées suite aux résultats d'une évaluation » (p. 553). Si l'inférence porte sur le statut de l'élève ou sur son futur (par exemple, certifier ses apprentissages ou lui donner l'accès à l'année ultérieure), c'est la visée sommative qui est privilégiée. Si au contraire le type d'inférence porte sur le type d'action (produite par l'élève ou par l'enseignant) qui aide l'élève à mieux apprendre, l'évaluation fonctionne alors de façon formative. Ce n'est donc pas l'évaluation elle-même qui est sommative ou formative, mais bien ce que l'on en fait¹³.

C. Une évaluation-soutien d'apprentissage qui conjugue évaluation formative et certains aspects de l'évaluation sommative

Laveault et Allal (2016a) sont les éditeurs scientifiques d'un ouvrage publié en langue anglaise, mais regroupant des travaux provenant d'auteurs francophones et anglophones, cherchant ainsi à faire dialoguer des écrits qui, par le passé, relevaient souvent de traditions de recherches et de conceptualisations théoriques différentes. Intitulé « *Assessment for learning. Meeting the challenge of implementation* », cet ouvrage s'appuie sur la définition proposée par l'*Assessment Reform Group* (ARG) qui a conceptualisé cette version moderne et élargie de l'évaluation formative¹⁴ :

L'évaluation-soutien d'apprentissage¹⁵ désigne le processus de collecte et d'interprétation des informations recueillies en vue d'être utilisées par les élèves et les enseignants pour déterminer où en sont les élèves dans leur apprentissage, dans quelle direction ils ont besoin d'aller et quel est le meilleur chemin pour y parvenir (Assessment Reform Group, 2002, cité par Laveault & Allal, 2016b, p. 3).

¹³ Ces considérations rejoignent assez largement la distinction proposée par Allal (1991, voir point 1) qui évoque quant à elle des formes de régulations et des décisions différentes.

¹⁴ Dans le chapitre introductif, Laveault & Allal (2016b) précisent que le terme a été introduit par l'ARG en 1999 en vue de décrire « l'orientation qui devrait être donnée aux pratiques évaluatives développés par les enseignants dans leur classe dans le but de soutenir l'apprentissage des élèves » (p. 3). Wiliam (2011) considère quant à lui que ce terme est apparu pour la première fois dans un chapitre publié par Black dans les années 1980, mais nous n'entrerons pas ici dans ces controverses. Dans ce chapitre de 2011, Wiliam présente toutefois les deux définitions de l'ARG (celles de 2002 et de 2009, telles que reprises ici au départ du texte de Laveault & Allal, 2016b).

¹⁵ Les termes « Evaluation-soutien d'apprentissage » pour traduire cette conception de l'*Assessment for learning* ont été proposés par Allal & Laveault (2009) dans un article publié dans la revue *Mesure et évaluation en éducation*.

Les auteurs précisent aussi que cette définition a été mise à jour en 2009 en vue de mettre davantage en lumière la nature dynamique de l'*assessment for learning* (AfL) et de mieux montrer encore comment l'évaluation s'insère dans le processus d'enseignement apprentissage.

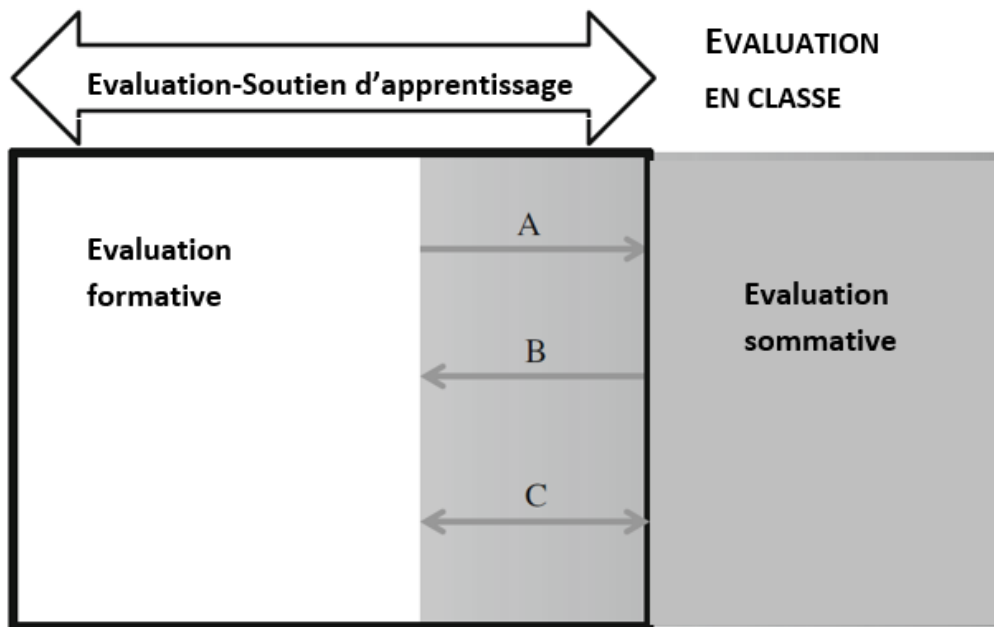
L'évaluation-soutien d'apprentissage fait partie des pratiques quotidiennes des élèves et des enseignants qui, individuellement et en interaction, recherchent, réfléchissent sur et réagissent à l'information provenant d'échanges, démonstrations et observations afin de favoriser les apprentissages en cours (Klenowski, 2009, p. 264, cité par Laveault & Allal, 2016b, p. 4)¹⁶.

En cohérence avec l'élargissement de la vision francophone de l'évaluation formative, la définition de l'AfL considère que l'évaluation s'insère au cœur des activités d'enseignement/apprentissage et implique activement l'élève dans le processus (autoévaluation et évaluation mutuelle entre pairs notamment). Laveault et Allal (2016b) insistent également sur le versant métacognitif du processus : l'AfL devrait permettre l'engagement des élèves dans une réflexion à propos de leur propre processus d'apprentissage. Ils précisent aussi que l'AfL ne se limite pas à la sphère cognitive, mais prend également en compte le versant affectif et social (motivation, attitudes par rapport à l'apprentissage et coopération dans l'apprentissage).

Tout comme les auteurs américains précédemment cités (Black & Wiliam, 2018 ; Benett, 2011), Laveault et Allal (2016b) pointent les limites des typologies distinguant « *assessment of/for/as learning* ». La Figure 2 modélise la façon dont ils conçoivent l'intégration de l'AfL dans les pratiques d'évaluation en classe pour rendre compte du fait que plusieurs travaux ont montré que les enseignants expérimentés avaient tendance à conjuguer les fonctions formative et sommative de l'évaluation. Black et Wiliam (2018) vont dans le même sens lorsqu'ils écrivent : « de notre travail avec les enseignants, nous avons retiré la conviction que toute tentative d'amélioration des pratiques de classe se concentrant sur l'évaluation doit intégrer tous les aspects de l'évaluation » (p. 552). Dans le schéma, on note que l'AfL englobe l'évaluation formative mais aussi plusieurs aspects de l'interface entre l'évaluation formative et sommative.

¹⁶ La traduction a été empruntée à Allal & Laveault (2009, p. 102).

Figure 2 : *Assessment for learning* : sa fonction formative et son interface avec l'évaluation sommative (traduit de Laveault & Allal, 2016b, p. 8 – d'après les travaux d'Allal publiés au début des années 2010)



La flèche A illustre que certains éléments issus des évaluations formatives peuvent aussi être pris en compte dans une évaluation sommative. C'est le cas, par exemple, lorsqu'un enseignant décide de la note attribuée à un élève dans son bulletin en tenant compte, non seulement des résultats issus des tests sommatifs, mais aussi d'observations et de discussions avec l'élève durant des évaluations formatives plus informelles. Les auteurs considèrent que cette façon de procéder permet un jugement plus robuste que si l'enseignant avait fait une simple moyenne des notes obtenues à des tests. On peut comprendre qu'ils considèrent que l'enseignant fait ici preuve d'un jugement professionnel (Mottier Lopez & Allal, 2008) appuyé sur une diversité de sources d'informations, mais on peut aussi craindre que cela ne conduise à brouiller un peu les cartes au niveau des élèves. Parmi les points d'attention qui concluaient leur revue de la littérature, Black et Wiliam (1998a) soulignaient en effet les tensions possibles entre les fonctions formative et sommative de l'évaluation, et ceci tant pour les enseignants que pour les élèves.

La flèche B illustre que certaines informations issues d'une évaluation sommative peuvent aussi être utilisées à des fins formatives. C'est le cas par exemple quand l'enseignant analyse les erreurs commises par les élèves lors d'une interrogation pour proposer ensuite des activités aux élèves en vue de les aider à progresser dans leur compréhension des concepts en jeu. Les auteurs précisent qu'utiliser les évaluations sommatives à des fins formatives est une pratique courante quand l'enseignant dispose de trop peu de temps pour mettre en œuvre des évaluations formatives en classe¹⁷.

¹⁷ Notons toutefois que ce « manque de temps » correspond à une vision assez classique de l'évaluation formative selon laquelle il faudrait prendre le temps de mettre en place des évaluations formelles ou instrumentées, mais ne s'applique pas nécessairement aux évaluations plus informelles qui s'intègrent pleinement au cœur des activités d'enseignement/apprentissage.

Black et Wiliam (2018) soutiennent aussi que l'évaluation sommative peut œuvrer à des fins formatives, tant pour l'enseignant que pour les élèves. On peut en effet escompter que, si l'évaluation est critériée et que le feedback fourni aux élèves en lien avec l'évaluation sommative ne se limite pas à une note, mais précise les éléments maîtrisés ou non, les élèves peuvent eux-mêmes s'en saisir pour réguler leurs apprentissages. Ce n'est certes pas suffisant et il est important que l'enseignant les soutienne dans ce processus, mais il paraît néanmoins important de préciser cette potentielle utilisation de l'évaluation sommative par l'élève. Black et Wiliam (2018) précisent aussi qu'un des éléments par lesquels l'évaluation sommative peut jouer un rôle formatif consiste « simplement » à ce que l'évaluation proposée permette de préciser les attentes de l'enseignant dans le domaine visé. En effet, nous verrons que la compréhension par l'élève des objectifs à atteindre et des critères en témoignant constituent des incontournables des pratiques d'évaluations formatives efficaces (voir notamment la synthèse de Black & Wiliam, 1998a, que nous développerons dans la partie II. de cette note).

La **flèche C** illustre l'interaction directe entre les fonctions sommatives et formatives dans des activités évaluatives complexes qui favorisent l'apprentissage des élèves tout en contribuant à rendre compte de ces apprentissages. Laveault et Allal (2016b) donnent l'exemple de la préparation d'un portfolio, qui peut se réaliser sur une période relativement longue et qui déterminera la note de l'élève pour une matière. Dans chaque phase successive de construction de ce portfolio, les élèves sont amenés à améliorer leurs productions écrites grâce aux feedbacks des pairs ou de l'enseignant ; ils sont aussi amenés à effectuer un recul réflexif (métacognitif) sur leur processus d'écriture et de révision et à mettre en œuvre progressivement une autoévaluation de leurs travaux.

Laveault et Allal (2016b) reconnaissent l'intérêt des éléments relevant de la partie grisée du schéma de l'AfL. Ils estiment en effet qu'ils traduisent 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. Ils considèrent toutefois que les formes optimales d'AfL sont celles qui s'intègrent pleinement dans les activités d'enseignement/apprentissage (cf. partie gauche du schéma) et qui permettent d'assurer des régulations interactives des apprentissages. Les auteurs précisent que la vision qu'ils qualifient d'optimale est celle qui s'accorde le mieux avec la définition révisée de l'ARG. Or, ce qui est frappant dans cette définition, c'est qu'elle semble s'appuyer essentiellement sur une évaluation informelle (cf. dialogues, observations,...) et que la régulation devrait dès lors se faire dans l'instant. Wiliam (2018) regrette que la vision défendue dans l'ouvrage soit extrêmement marquée par une approche socioculturelle somme toute limitative, tout comme la définition de l'ARG qui « limite sévèrement ce qui peut être considéré comme relevant de l'AfL » (p. 684). Ainsi, selon cet auteur, un enseignant qui se servirait des résultats d'une évaluation sommative pour réguler la façon d'organiser son cours l'année suivante n'entrerait pas dans la définition retenue. De même, on pourrait avoir l'impression que la définition exclut partiellement une évaluation formative instrumentée, qui pourrait prendre la forme d'un test proposé à la fin d'un cours, corrigé par l'enseignant par la suite en vue de cibler des activités engageant une forme de régulation rétroactive ou proactive pourtant chères

à Allal (2007a, 2020¹⁸). C'est peut-être la raison pour laquelle Laveault et Allal (2016b) proposent une définition opérationnelle de l'AfL qui ne semble exclure aucune forme 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, 2016b, p. 7).

(1) Ce que l'on peut retenir des différentes facettes de l'évaluation en classe

En classe, les enseignants doivent nécessairement jongler avec les différentes fonctions de l'évaluation (Black & Wiliam, 2018) et il semblerait que les enseignants expérimentés tendent à conjuguer les fonctions sommative et formative de l'évaluation (Laveault & Allal, 2016b).

Pour que les évaluations sommatives puissent se conjuguer harmonieusement aux évaluations formatives en vue d'œuvrer de concert au soutien des apprentissages des élèves, la recherche suggère d'éviter toute référence normative pour privilégier une référence critériée (Crahay *et al.*, 2019) ; de respecter l'alignement curriculaire tout en réfléchissant aux éléments constitutifs de la note (Pasquini, 2019) et d'accompagner les résultats de commentaires constructifs (Calone & Lafontaine, 2023).

Depuis les travaux de Bloom, tant les travaux francophones qu'anglophones ont proposé des élargissements à la vision initiale de l'évaluation formative (Allal & Mottier Lopez, 2005 ; Black & William, 2018). Aujourd'hui, les différents auteurs consultés semblent s'accorder sur le fait que l'évaluation formative devrait s'insérer pleinement au cœur des activités d'enseignement/apprentissage et qu'elle devrait chercher à impliquer activement l'élève dans le processus (*via* l'autoévaluation et l'évaluation mutuelle entre pairs notamment, que nous développerons au point II.C.2.).

Quoi que séduisantes au premier abord, les typologies qui cherchent à distinguer l'évaluation « de », « pour » et « en tant que » apprentissage (*Assessment OF, FOR ou AS learning*) – que l'on serait tenté d'associer aux évaluations sommative, formative et formatrice – risquent de concourir à brouiller davantage les choses et à conférer une vision étroite tant à l'évaluation sommative (qui ne jouerait alors aucun rôle de soutien à l'apprentissage) qu'à l'évaluation formative (qui n'intégrerait pas suffisamment l'élève dans le processus) (Black & William, 2018 ; Bennett, 2011).

Finalement, c'est donc la vision large de *l'assessment for Learning* (Laveault & Allal, 2016b) – concept traduit par les termes d'« Évaluation-soutien d'apprentissage » (ESA) (Allal & Laveault, 2009) – que nous retiendrons dans ce rapport. Ainsi conceptualisée, l'ESA implique l'élève dans le processus d'évaluation et intègre des évaluations formatives formelles et informelles, ainsi que les différents types de régulations qui peuvent découler d'une évaluation sommative critériée.

¹⁸ Allal (2020) considère en effet que « pour être efficace, l'évaluation formative nécessite la coordination de modalités de régulation interactives, rétroactives et proactives » p. 337).

II. Quand l'évaluation en classe soutient les apprentissages des élèves

Comme la précédente, cette partie est structurée en trois points : le premier (A.) synthétise les grands constats que l'on peut tirer sur les effets de l'évaluation-soutien d'apprentissage (ESA) ; le deuxième (B.) propose deux modélisations de la façon dont l'ESA opère en classe ; enfin, le troisième point (C.) propose un focus sur trois modalités d'ESA (les conversations évaluatives, les évaluations mutuelles entre pairs et l'autoévaluation) en cohérence avec les modélisations présentées au point B.

A. Que nous disent les articles de synthèse, les revues systématiques et les méta-analyses sur les effets de l'évaluation-soutien d'apprentissage ?

Pour aborder la question des effets de l'évaluation-soutien d'apprentissage, un incontournable est l'article de synthèse publié fin des années 1990 par Black et Wiliam. Cet article de plus de 60 pages (Black & Wiliam, 1998a) est complété par un autre nettement plus bref (10 pages) qui reprend les résultats principaux (Black & Wiliam, 1998b). Leur article de synthèse porte sur l'évaluation formative, mais la définition large qu'ils en donnent colle assez bien à la définition de l'ESA que nous avons retenue. Pour compléter ces travaux, nous ferons aussi référence à un article plus récent, rédigé par Wiliam (2011) et intitulé « *What is assessment for learning ?* ». Cet article fait le point sur les nouvelles synthèses portant sur l'évaluation en classe et explique comment le concept d'évaluation formative a évolué dans la littérature anglo-saxonne, probablement sous l'influence des travaux d'Allal. Enfin, nous avons aussi trouvé quelques méta-analyses ou synthèses plus récentes (Benett, 2011 ; Kingston & Nash, 2011 ; Briggs *et al.*, 2011 ; Klute *et al.*, 2017 ; Lee *et al.*, 2020) qui nous permettront de mettre à jour les résultats tirés des premiers écrits. Il est à noter que de nombreux travaux récents n'envisagent plus l'évaluation formative, l'AfL ou l'ESA au sens large, mais proposent des revues systématiques ou des méta-analyses portant sur des modalités spécifiques (comme l'évaluation entre pairs par exemple), ce qui montre en soi l'ampleur que ce champ de recherche a prise au cours des vingt dernières années. Nous ferons référence à ces travaux au point C.

Encadré 2 : Article de synthèse, revue systématique et méta-analyse : de quoi s'agit-il ?

« L'article scientifique de type article de synthèse (*review paper*) fait le bilan des recherches et des connaissances sur un sujet précis à partir d'une sélection de publications et propose de nouvelles pistes de recherche. Sur la base d'une analyse bibliographique approfondie, les auteurs décortiquent les concepts et les questions, les méthodologies et les résultats des équipes de recherche travaillant sur ce sujet dans le monde ».

La revue systématique et la méta-analyse sont deux types particuliers d'articles de synthèse.

- « La revue systématique (*systematic review*) recense, analyse et synthétise de façon standardisée et objective des données scientifiques probantes et pertinentes afin de répondre à une question précise et clairement formulée. Pour réduire le plus possible les biais, elle utilise une méthode rigoureuse de recherche et de sélection d'information bibliographique. »

- « La méta-analyse (*meta-analysis*) est une revue systématique à laquelle s’ajoute une analyse statistique combinant les données de toutes les publications recensées. Elle aboutit à un résultat original, comme révéler des tendances qui ne pouvaient pas être vues à partir de chaque article de recherche pris séparément »

Source : <https://coop-ist.cirad.fr/rediger/article-de-synthese/5-cas-particuliers-revue-systematique-meta-analyse>

Le point de départ de la synthèse de Black et Wiliam (1998a) est l’intérêt, considéré comme déjà croissant à l’époque, des recherches qui relient « évaluation » et « apprentissage ». Leur article de synthèse poursuit deux objectifs : (a) vérifier si les espoirs associés à cette perspective permettent réellement d’améliorer les apprentissages et, si oui, (b) clarifier comment et dans quelles conditions. Pour réaliser ce travail, ils prennent pour point de départ la définition suivante de l’évaluation formative :

Toutes les activités entreprises par les enseignants et/ou par leurs élèves qui fournissent des informations destinées à être utilisées en tant que feedbacks en vue de modifier les activités d’enseignement et d’apprentissage dans lesquelles ils sont engagés (Black & Wiliam, 1998a, pp. 7-8).

Concernant le premier objectif, ils concluent que l’évaluation formative améliore l’apprentissage : les progrès observés sont notables et se situent parmi les plus importants observés pour des interventions éducatives. Les auteurs n’ont pas procédé à une méta-analyse permettant de calculer une taille d’effet moyen, mais ils concluent que l’effet des interventions varie entre 0,4 et 0,7, ce qui est considérable, mais qui sera remis en cause par des travaux ultérieurs.

Encadré 3 : La taille de l’effet : de quoi s’agit-il ?

La « taille de l’effet » est un indice statistique qui permet de qualifier l’ampleur des différences observées entre deux groupes d’individus comparables, l’un – qualifié de *groupe expérimental* – ayant bénéficié d’une intervention innovante (ici, la mise en œuvre de pratiques d’évaluation formative) et l’autre – qualifié de *groupe contrôle* – n’ayant pas participé à ce dispositif innovant.

Un indicateur couramment utilisé dans les recherches en sciences de l’éducation est le « d » de Cohen (d_{Cohen}) qui se calcule comme suit :

$$d_{Cohen} = (\text{moyenne du groupe expérimental} - \text{moyenne contrôle}) / \text{écart-type combiné des deux groupes}^{19}$$

Classiquement, en référence aux travaux de Cohen (1988, cité par Hill *et al.*, 2008), il était courant de considérer une taille de l’effet de 0,20 comme petite ; de 0,50 comme moyenne et de 0,80 comme importante. Suite aux travaux de Hattie (2009), c’est généralement une taille de l’effet de 0,40 qui est considérée comme « faisant la différence » par rapport à une approche d’enseignement classique

¹⁹ Lehanrd et Lenhard (2016) précisent toutefois qu’il peut être recommandé de diviser par l’écart-type du groupe contrôle plutôt que par l’écart-type combiné lorsque les dispersions sont très différentes entre les deux groupes. Par ailleurs, ils précisent aussi que d’autres formules peuvent être appliquées lorsque les deux groupes sont de taille distinctes (g_{Hedges}) ou encore pour prendre en compte un plan de recherche qui s’appuie sur un dispositif pré / post-test (d_{ppc2} sensu Morris). Le lecteur intéressé pourra trouver les informations détaillées sur le site https://www.psychometrica.de/effect_size.html qui propose par ailleurs des « calculateurs en ligne » permettant de calculer aisément ces différents indices statistiques.

(c'est-à-dire sans intervention innovante ou spécifique). Toutefois, certains auteurs recommandent d'interpréter ces tailles de l'effet en référence « à des repères empiriques pertinents pour l'intervention, la population cible et le type de résultats considérés » (Hill *et al.*, 2008, p. 172). Ainsi par exemple, une taille de l'effet de 0,40 n'aura pas la même signification s'il s'agit d'une intervention réalisée au préscolaire ou en fin de secondaire. Les méta-analyses mentionnées dans ce rapport regroupant une diversité de recherches menées à des niveaux scolaires différents, on comprend à quel point il est difficile d'interpréter la réelle signification des tailles de l'effet agrégées.

Concernant le deuxième objectif (la clarification des conditions dans lesquelles l'AfL peut améliorer les apprentissages), Black et Wiliam (1998a) concluent qu'ils n'ont pas pu déceler un modèle optimal d'évaluation formative, mais qu'il est possible de dégager un ensemble de principes et de précautions. Nous reprenons ici les éléments principaux :

- préciser les **objectifs** d'apprentissage et **lier les évaluations et les feedbacks (FB)** ;
- réaliser des **évaluations fréquentes** ;
- fournir des **FB (diagnostiques)** et donner l'opportunité aux élèves de les **utiliser** ;
 - Les FB adressés à chaque élève doivent porter sur « les qualités particulières du travail de l'élève, donner des conseils sur ce qu'il ou elle peut faire pour progresser et éviter toute comparaison avec les autres élèves » (Black & Wiliam, 98b, p. 7).
 - Il est important de ne pas se limiter aux FB proposés, mais de prendre en compte les réactions de l'élève face à ces FB, ainsi que l'environnement d'enseignement/apprentissage dans lequel les FB opèrent.
- **impliquer les élèves dans le processus évaluatif** : l'élève doit constater un écart entre un but désiré et son état actuel ; il doit entreprendre une action pour combler/réduire cet écart ;
 - Il est important de prendre en compte le rôle de l'enseignant, des pairs et de l'élève lui-même.
 - Il faut éviter de considérer l'élève comme un récepteur passif : il est nécessaire de questionner la façon dont le message est réceptionné et d'interroger les actions menées ou à mener en conséquence.
 - Il est intéressant de développer **l'évaluation entre pairs** et **l'autoévaluation**. Pour ce faire, les élèves doivent comprendre les objectifs d'apprentissage visés ; ils doivent comprendre les critères d'évaluation et, enfin, ils doivent avoir l'opportunité de réfléchir sur leur travail et de le réguler. En procédant de la sorte « l'évaluation doit être considérée comme un moment d'apprentissage » (Black & Wiliam, 1998a, p. 29).
 - Les élèves doivent aussi avoir l'occasion d'être impliqués dans des dialogues constructifs²⁰ avec leur enseignant et leurs pairs.
- tenir compte des **perceptions** (buts poursuivis, concept de soi, sentiment d'efficacité personnelle, attributions causales, perception de l'effort,...) et des croyances (relatives à l'apprentissage, à l'enseignement, à l'intelligence,...) des élèves et des enseignants. Ces

²⁰ En lien avec les travaux de Ruiz-Primo (2011), nous qualifierons ces dialogues de « conversations évaluatives ».

derniers doivent avoir des attentes élevées et être convaincus que les élèves peuvent atteindre les objectifs visés par les programmes scolaires ;

- prendre en compte le **climat de classe** dans lequel opèrent les processus d'évaluation formative et, le cas échéant, revoir le **contrat didactique**²¹ avec les élèves.

Au terme de leur article de synthèse, Black et Wiliam (1998a) pointent deux regrets : (a) la synthèse qu'ils ont réalisée n'est pas une méta-analyse et, à leur connaissance, il n'y en pas encore sur le sujet ; (b) le concept d'évaluation formative est flou et mal défini dans de nombreuses études ; il recouvre ainsi une série de pratiques diversifiées tels que la pédagogie de la maîtrise, les évaluations en classe, les feedbacks,...

Ils pointent aussi quelques points d'attention qu'il nous paraît utile de souligner ici : (1) les effets observés semblent particulièrement bénéfiques pour les apprenants « à risque » ou peu performants ; (2) plusieurs travaux pointent des tensions possibles entre les fonctions formative et sommative de l'évaluation, tant pour les enseignants que pour les élèves ; (3) pour être efficace, l'évaluation formative nécessite bien plus qu'un changement marginal ; elle doit s'intégrer pleinement dans les pratiques de classe (cf. importance de prendre en compte les perceptions, les croyances, le climat de classe et le contrat didactique).

Le premier point (effets observés particulièrement bénéfiques pour les apprenants « à risque » ou peu performants) est important en matière d'équité, mais nous verrons que, malheureusement, les synthèses et méta-analyses ultérieures semblent accorder peu d'attention à cet aspect. Nous avons déjà évoqué le deuxième élément en présentant le modèle d'*assessment for learning* développé par Laveault et Allal (2016a,b). S'il semble indéniable aujourd'hui que les fonctions formative et sommative de l'évaluation doivent être combinées, il semble tout aussi essentiel d'être clair et explicite envers les élèves quant aux enjeux des différentes formes d'évaluation pratiquées en classe. Il convient aussi d'intégrer l'implémentation de l'évaluation formative à une réflexion plus large sur les pratiques de classe, ce que pointe d'ailleurs explicitement le dernier point d'attention. L'extrait suivant est particulièrement éclairant à cet égard :

Tout enseignant qui veut pratiquer l'évaluation formative doit reconstruire le contrat didactique de manière à contrecarrer les habitudes acquises par ses élèves. C'est l'élève qui est l'utilisateur final des informations issues d'évaluations et destinées à améliorer l'apprentissage. Ce point présente des aspects négatifs et positifs. [...] Lorsque la culture de la classe est axée sur les récompenses, les "bons points", les notes ou le classement, les élèves cherchent à obtenir les meilleures notes plutôt qu'à améliorer leur apprentissage. L'une des conséquences observées est que, lorsqu'ils ont le choix, les élèves évitent les tâches difficiles. Ils consacrent également du temps et de l'énergie à chercher des indices de la "bonne réponse". Beaucoup hésitent à poser des questions par peur de l'échec. Les élèves qui rencontrent des difficultés sont amenés à croire qu'ils manquent de capacités, et cette croyance les conduit à attribuer leurs difficultés à un défaut intrinsèque auquel ils ne peuvent pas grand-chose. Ils évitent ainsi d'investir des efforts dans les apprentissages qui ne peuvent mener qu'à des déceptions, et tentent de renforcer leur estime d'eux-mêmes par d'autres moyens. [...] Ce qu'il faut, c'est une culture de la réussite,

²¹ Le contrat didactique est un terme qui a été défini par Brousseau dans les années 1980 comme étant « l'ensemble des comportements (spécifiques [des connaissances enseignées]) du maître qui sont attendus de l'élève et l'ensemble des comportements de l'élève qui sont attendus du maître (Brousseau 1980, p. 127) », cité par Sansevy, 1995).

soutenue par la conviction que tous les élèves peuvent atteindre le objectifs d'apprentissage. À cet égard, l'évaluation formative peut être une arme puissante si elle est adéquatement communiquée (Black & Wiliam, 1998b, p. 143).

Quoi de neuf depuis ces travaux datant de la fin des années 1990 ?

Wiliam (2011) pointe trois éléments importants. (1) Tout d'abord, entre 2002 et 2004, des études ont été lancées par l'OCDE sous l'intitulé « *What works in innovation in education ?* ». C'est de ces travaux que découle notamment la revue des travaux en langue française de Allal et Mottier Lopez (2005) que nous avons évoquée précédemment pour expliciter l'élargissement du concept d'évaluation formative dans le monde francophone. En conclusion de leur article, on peut retenir que les travaux francophones ont conduit à des développements conceptuels importants, mais qu'il manquait à ce jour des travaux empiriques solides pour étayer les effets et les conditions d'efficacité des pratiques d'évaluation formative²². (2) Wiliam (2011) a aussi pu repérer trois nouvelles revues de la littérature, mais deux sont centrées explicitement sur les feedbacks (Hattie & Timperley, 2007 et Shute, 2008, cités par Wiliam, 2011²³) et la dernière sur les apprentissages mathématiques (Wiliam, 2007, cité par Wiliam, 2011). (3) Le dernier élément qui a retenu notre attention concerne l'évolution du concept dans le monde anglophone et la possible influence des travaux francophones. L'extrait suivant, faisant référence à la synthèse précitée sur les mathématiques, est éclairant à cet égard :

Wiliam a dégagé de cette recherche certaines implications pour l'enseignement des mathématiques et a souligné comment les idées centrales de l'évaluation formative pouvaient être intégrées dans le concept plus large de 'régulation des processus d'apprentissage' tel que développé à partir de la littérature francophone (Wiliam, 2011, p. 8).

Deux méta-analyses (Kingston & Nash, 2011 ; Klute *et al.*, 2017) et deux revues critiques (Bennett, 2011 ; Briggs *et al.*, 2012) paraissent à la même époque ou un peu plus tard. La méta-analyse de Kingston et Nash (2011) conduit à relativiser la taille de l'effet observé, tout en confirmant l'efficacité de l'évaluation formative pour améliorer les apprentissages des élèves (taille de l'effet proche de 0,20 alors que Black & Wiliam, 1998a,b ciblaient une taille de l'effet variant entre 0,40 et 0,70). Elle permet aussi de mettre en évidence que des différences notables apparaissent selon les études et selon les disciplines scolaires envisagées ; l'année scolaire concernée²⁴ ne semblant quant à elle pas être un facteur discriminant. Ils ont aussi pu mettre en évidence deux types d'approches qui semblaient particulièrement efficaces : celles basées sur le développement professionnel des enseignants, d'une part, et celles basées sur un système informatique permettant de rendre des FB rapidement, d'autre part. Enfin, bien que leur travail ait porté sur plus de 300 études, ils terminent leur article en faisant appel à davantage de recherches empiriques de qualité dans la mesure où ils n'ont finalement pu retenir que 13 études qui présentaient un design de recherche suffisamment solide.

²² Le même type de synthèse réalisée par Köller (2005, cité par Wiliam, 2011) pour faire le point sur les travaux allemands conduit à un constat similaire en regrettant le peu d'études empiriques sur les effets des évaluations formatives.

²³ Pour les travaux sur les feedbacks, nous renvoyons le lecteur intéressé au rapport de Calone et Lafontaine (2023) réalisé dans le cadre de cette même conférence de consensus du Cnesco de 2022.

²⁴ Kingston & Nash (2011) ont ciblé les recherches portant sur la fin de l'enseignement préscolaire jusqu'à la fin de l'enseignement secondaire – grades K-12).

La revue critique de Briggs *et al.* (2012) apporte un certain nombre de critiques méthodologiques à la méta-analyse de Kingston et Nash (2011) et conclut que « des incertitudes persistent concernant l'effet des pratiques d'évaluation formative sur la réussite des élèves » (p. 13)²⁵.

Quelques années plus tard, la méta-analyse de Klute *et al.* (2017) s'appuie quant à elle sur 19 études et conclut à une taille de l'effet moyenne de 0,26, tout en pointant à nouveau des différences en fonction des disciplines scolaires²⁶.

Finalement, malgré un certain nombre de critiques émises, les diverses études susmentionnées permettent de conclure à une certaine efficacité de l'ESA, mais il semble important d'aller encore un pas plus loin pour mieux comprendre les conditions dans lesquelles cette ESA peut s'avérer efficace. À cet égard, la revue critique de de Bennett (2011) souligne un certain nombre de points d'attention qui nous paraissent pertinents. L'auteur développe six axes que nous synthétisons brièvement ici :

- 1. La faiblesse conceptuelle des définitions de l'évaluation formative ou de l'AfL.** Cette faiblesse conceptuelle rend toute entreprise de synthèse systématique ou de méta-analyse particulièrement complexe « car si nous ne pouvons pas clairement définir une innovation, il n'est pas possible de documenter son efficacité de façon sensée » (p. 8).
- 2. Les doutes quant aux preuves d'efficacité.** Après les travaux de Black et Wiliam (1988a,b), il a été courant de considérer que la taille de l'effet des pratiques d'évaluation formative variait entre 0,40 et 0,70. Les travaux de ces auteurs ne permettent toutefois pas de tirer de telles conclusions. Non seulement, il ne s'agit pas d'une méta-analyse, mais Bennett (2011) estime aussi que cela a peu de sens de vouloir agréger l'effet de pratiques qui s'avèrent finalement très disparates²⁷.
- 3. Le manque d'attention accordée aux domaines de contenus.** Rien ne prouve que l'ESA se situe à un niveau général, non dépendant d'un domaine spécifique. Si des principes généraux sont essentiels, il est raisonnable de considérer qu'une compréhension approfondie de la discipline enseignée est nécessaire pour mettre en place efficacement des pratiques d'évaluation formative. En effet, si l'enseignant ne dispose pas de connaissances didactiques spécifiques (notion de *pedagogical content knowledge*, Shulman, 1986, 2007), il ne sera pas à même de poser des questions pertinentes aux élèves, de réaliser les inférences nécessaires pour interpréter leurs réponses et de dégager des pistes didactiques adéquates pour remédier aux difficultés mises à jour. Par ailleurs, il est aussi possible « que l'instrumentation et l'outillage intellectuels que l'on apporte aux enseignants diffèrent significativement d'une

²⁵ Les critiques adressées par Briggs *et al.* (2012) portent sur (a) les mots-clés utilisés qui restreignent trop le champ (Kingston & Nash avaient fait le choix de ne retenir que les études qui mentionnent *formative assessment* ou AfL dans la description de l'intervention) ; (b) l'application rigoureuse des critères définis (selon eux, deux études auraient dû être retenues et ont été rejetées) ; (c) la façon même de calculer la taille de l'effet moyenne et (d) le manque de recul quant à l'effet induit par le type de test utilisé (quand il s'agit d'un test construit par l'enseignant, il y a un risque d'enseigner par rapport au test - *teaching to test* - et d'ainsi augmenter artificiellement l'effet de l'intervention).

²⁶ Klute *et al.* (2017) ont ciblé les recherches portant sur l'enseignement primaire (grades 1-6), mais quelques études comprennent aussi des élèves de début secondaire (grades 7-8 voire grade 9 exceptionnellement).

²⁷ Si l'on s'accorde à cette critique, on comprend mieux pourquoi les travaux plus récents ont porté sur l'analyse d'éléments singuliers, comme l'évaluation mutuelle entre pairs par exemple, tels que ceux que nous présenterons au point C.

discipline à l'autre parce qu'ils devraient justement être sphériquement calibrés en fonction de la discipline en question » (Bennett, 2011, p. 14-15).

4. **Les questionnements quant aux « mesures » permettant de poser le diagnostic.** L'évaluation formative est un processus inférentiel qui nécessite de pouvoir interpréter adéquatement les éléments d'observation recueillis pour pouvoir agir en conséquence. La difficulté de ce processus inférentiel n'est pas suffisamment soulignée dans les travaux (par exemple, une erreur observée est-elle une faute de distraction, un manque de compréhension ou le signe d'une conception erronée ?). Pourtant, chaque type d'erreurs donnera lieu à des actions différentes (un simple FB pour une erreur de distraction, des explications complémentaires pour une mauvaise compréhension et une approche didactique plus complexe s'il s'agit de bousculer et de restructurer des conceptions erronées). Dans de nombreux cas, pour réaliser des inférences correctes, il est nécessaire de procéder à des prises d'informations complémentaires : poser de nouvelles questions aux élèves, leur demander d'expliquer leur raisonnement, leur soumettre une série d'exercices bien pensés en vue de dégager des patterns d'erreurs, ... Tous ces éléments qui semblent essentiels pour que l'ASE puisse s'avérer efficace ne sont pas toujours développés dans les travaux analysés et mériteraient des recherches spécifiques.
5. **La nécessaire attention à porter au développement professionnel des enseignants.** La mise en place de pratiques efficaces d'évaluation formative nécessite des connaissances pédagogiques générales en évaluation, mais aussi des connaissances didactiques (Shulman, 1986, 2007). Ces deux facettes devraient être prises en compte dans les programmes de développement professionnel.
6. **L'importance de prendre en compte le système (cohérence interne et externe).** La cohérence interne fait référence à la façon d'articuler évaluation formative et sommative. La cohérence externe fait référence aux finalités du système éducatif, mais aussi aux liens avec les évaluations externes (sont-elles toutes deux en cohérence avec les intentions de l'ESA ?).

Nous terminerons ce point en faisant référence à l'étude de Lee *et al.* (2020) qui présente le double intérêt de proposer une revue systématique récente (appuyée sur 33 études réalisées de la dernière année de l'enseignement maternel à la fin de l'enseignement secondaire - grades k-12) et d'utiliser des modèles statistiques complexes (des régressions multi-niveaux leur permettant d'analyser l'effet propre d'une modalité en neutralisant les autres variables) pour chercher à dégager l'effet différentiel de différentes caractéristiques de l'ESA. Tout d'abord, ils mettent en évidence un taille de l'effet moyenne de 0,29 (c'est-à-dire proche de ceux mis en évidence par Kingston & Nash, 2011 et par Klute *et al.*, 2017) et un effet différentiel en fonction des disciplines concernées (les différences varient toutefois en fonction des études et permettent difficilement de tirer une conclusion à cet égard²⁸). Leur revue de la littérature les a aussi conduits à dégager huit variables qui peuvent prendre des modalités diverses :

²⁸ Kingston & Nash (2011) ont observé une taille d'effet supérieure pour les langues (*English language art*, 0,32) par rapport aux mathématiques (0,17) et aux sciences (0,17). C'est en mathématiques que Klute *et al.* (2017) observent quant à eux l'effet le plus important (0,36, contre 0,22 en lecture et 0,21 en écriture). Lee *et al.* (2020) pointent quant à eux des effets plus faibles en sciences (0,13) comparativement aux autres disciplines (0,34 en math, 0,33 en littérature et 0,29 en art).

- la source du FB : l'enseignant et/ou l'élève (autoévaluation) ;
- le caractère formel ou informel de l'évaluation formative ;
- la durée des cycles de FB (courts ou longs) ;
- la participation des enseignants à un programme de développement professionnel plus ou moins long (une séance ou un soutien en cours de recherche) ou absent ;
- le fait que le FB soit informatisé ou non ;
- la planification ou non de la régulation de l'enseignement ;
- le niveau scolaire dans lequel se déroule l'intervention (grades K-6 ou 7-12) ;
- le type d'enseignement dans lequel se déroule l'intervention (ordinaire ou spécialisé).

La seule variable pour laquelle ils observent une différence significative concerne la source du FB : toutes choses égales par ailleurs, l'effet est nettement plus important lorsque le point de départ est l'autoévaluation de l'élève que lorsqu'il s'agit d'une pratique initiée par l'enseignant voire d'une approche mixte. Ce résultat met en évidence l'importance de considérer l'implication active de l'élève dans le processus évaluatif ; ce dernier doit notamment « comprendre les objectifs d'apprentissage et les critères de réussite » et « devenir acteur de son propre apprentissage », deux éléments centraux du modèle de Wiliam et Thompson (2008, cités par Klute *et al.*, 2017, p. 135) que nous développerons au point suivant.

Quoique non-significatives, des différences ont également été mises en évidence en faveur du caractère formel de l'évaluation, de la participation à un programme de développement professionnel au cours de la recherche ou encore de la planification des actions de régulation. Ces éléments les conduisent à dégager quelques caractéristiques d'une pratique efficace d'ESA.

D'une façon générale, il s'agit de mettre en place un processus itératif pour contrôler constamment l'atteinte des objectifs d'apprentissage. Pour ce faire, il convient de collecter fréquemment des « traces » d'apprentissage, d'analyser les activités d'enseignement et les apprentissages réalisés par les élèves en s'appuyant sur ces traces, puis de prendre des décisions adéquates, basées sur ces analyses. Finalement, l'ESA est une pratique qui permet d'informer constamment l'enseignant et les élèves pour prendre des décisions, éclairées par les informations tirées de ces évaluations, sur les prochaines étapes d'apprentissage. Nous verrons au point suivant que ces éléments sont concordants avec la définition de Black et Wiliam et avec la modélisation proposée par l'équipe américaine (Wiliam, 2011).

(2) Ce que l'on peut retenir des articles de synthèse, revues systématiques et méta-analyses

Les diverses études susmentionnées permettent de conclure à une certaine efficacité de l'ESA et de dégager certaines caractéristiques qui semblent œuvrer à l'efficacité de cette pratique. À ce stade, nous proposons de retenir quatre éléments importants :

(1) l'implémentation des pratiques d'ESA devrait s'envisager en cohérence avec les pratiques d'enseignement et le climat de classe instauré (Black & Wiliam, 1998) ;

(2) l'implication de l'élève dans les pratiques d'ESA constitue un élément essentiel (Black & Wiliam, 1998 ; Lee *et al.*, 2020) ;

(3) si la diversité des pratiques d'ESA rend complexe le calcul d'un effet moyen (Bennett *et al.*, 2011) et rend difficile la mise en évidence des variables qui font la différence (Lee *et al.*, 2020), les divers travaux ont néanmoins permis aux chercheurs de proposer des modélisations qui explicitent la façon dont l'ESA pourrait être pensée en classe. C'est ce que nous allons développer au point B. Ils ont aussi permis de mettre en évidence l'efficacité potentielle de diverses pratiques qui impliquent l'élève dans l'évaluation (conversations évaluatives, évaluations entre pairs et autoévaluation) ; c'est ce que nous développerons au point C.

(4) Enfin, même si les résultats concernant l'efficacité de l'ESA selon les disciplines scolaires ne vont pas tous dans le même sens (King & Nash, 2011 ; Klute, 2017 ; Lee *et al.*, 2020), ne perdons pas de vue qu'il conviendrait de prendre en compte conjointement les connaissances générales en évaluation et les connaissances didactiques (*pedagogical content knowledge*, PCK) dans les programmes de développement professionnel des enseignants (Bennett, 2011).

B. Comment les chercheurs modélisent-ils l'évaluation soutien d'apprentissage en classe ?

Nous avons choisi de retenir ici deux modélisations : l'une produite par l'équipe américaine de chercheurs (Wiliam & Thompson, 2008, cités par Wiliam, 2011) ; l'autre issue des travaux provenant du monde francophone (Allal, 2007b, 2020).

1. La modélisation proposée par l'équipe américaine

Pour présenter ce modèle, partons de la définition retenue par les auteurs :

Une pratique de classe est formative dans la mesure où les informations concernant la réussite des élèves sont recueillies, interprétées et utilisées par les enseignants, les élèves ou leurs pairs pour prendre des décisions concernant les prochaines étapes d'apprentissage qui semblent être meilleures (ou mieux fondées) que celles qu'ils auraient prises en l'absence des informations recueillies (Black & Wiliam, 2009, p. 9, cité par Wiliam, 2011, p. 11).

Cette définition souligne que l'évaluation formative peut concerner l'ensemble des pratiques de classe plutôt que de constituer des événements isolés et ponctuels. La

Figure 3 illustre la mise en œuvre du processus d’ESA en classe en cohérence avec la définition susmentionnée.

Figure 3 : Modélisation de l’ESA proposée par l’équipe américaine (d’après Wiliam & Thompson, 2008 – traduit de Wiliam, 2011, p. 12)

	Ce vers quoi l’élève doit aller (les objectifs d’apprentissage)	Où l’élève en est actuellement	Comment atteindre les objectifs d’apprentissage
Enseignant	Clarifier les objectifs d’apprentissage et partager les critères de réussite	Mettre en place des échanges, des activités et des tâches qui permettent de recueillir des traces de l’apprentissage réalisé	Fournir des feedbacks qui font avancer les élèves
Pair	Comprendre et partager les objectifs d’apprentissage et les critères de réussite	Mobiliser les élèves de façon à ce qu’ils constituent des ressources pédagogiques les un pour les autres	
Élève	Comprendre les objectifs d’apprentissage et les critères de réussite	Mobiliser les élèves de façon à ce qu’ils soient acteurs de leur propre apprentissage	

Les colonnes mettent en évidence les **trois processus clés** d’un enseignement qui s’appuie sur l’ESA : **établir où les apprenants doivent aller** (les objectifs d’apprentissage visés) ; **établir où les apprenants en sont dans leur apprentissage** et **établir ce qui doit être fait pour atteindre les objectifs visés**. En plus de reconnaître l’importance d’être clair sur les objectifs d’apprentissage visés, le modèle met ainsi en évidence l’importance de recueillir des traces d’apprentissage et de les interpréter (pour savoir où les élèves en sont dans leur apprentissage), ainsi que l’importance d’utiliser ces informations pour prendre des décisions sur les prochaines étapes de l’enseignement ou de l’apprentissage afin d’amener les apprenants vers l’atteinte des objectifs visés (Anderson & Palm, 2017). Wiliam (2011) précise aussi que les décisions doivent être meilleures grâce aux informations récoltées. Il précise aussi que les prochaines étapes d’enseignement/apprentissage peuvent être entreprises par l’enseignant, l’élève, les pairs ou une combinaison des trois. En cohérence, les lignes du tableau représentent **les trois agents qui interviennent dans le processus**. Ainsi, l’enseignant et les élèves travaillent de concert et interagissent pour soutenir l’apprentissage au cours des trois processus clés.

Le **premier processus** (établir où les apprenants doivent aller) nécessite une compréhension mutuelle des objectifs d’apprentissage et des critères de réussite. **L’enseignant** a un rôle à jouer pour clarifier ces objectifs et partager avec les élèves les critères témoignant de leur atteinte. Il peut également mettre en place des activités invitant les élèves à partager leur compréhension des objectifs visés, favorisant ainsi les **interactions entre pairs** de façon à permettre *in fine* à **chaque apprenant** d’en avoir une compréhension solide.

Dans le **deuxième processus** (établir où les apprenants en sont dans leur apprentissage), **l’enseignant** met en place des dialogues constructifs (des « conversations évaluatives » selon Ruiz-Primo, 2011) et propose des tâches ou des activités diverses permettent d’obtenir des « traces » de la compréhension et des performances des élèves. Wiliam (2011) précise que ces « traces » doivent être « pédagogiquement traitables » (p. 11), c’est dire qu’elles ne doivent pas se contenter de pointer le fossé entre les performances actuelles et celles attendues, mais elles doivent aussi fournir des informations précises quant au type d’activités à mettre en place pour améliorer les apprentissages.

Ces éléments peuvent être mis en relation avec les propos de Bennett (2011) concernant le processus inférentiel impliqué dans l'ESA et la nécessité de disposer à cette fin de connaissances didactiques approfondies.

Dans le **troisième processus** (établir ce qui doit être fait pour arriver à l'atteinte des objectifs visés), **l'enseignant** peut fournir un FB aux élèves et ajuster les activités pédagogiques pour mieux répondre aux besoins d'apprentissage identifiés dans le deuxième processus (Anderson & Palm, 2017). Pour que ces actions soient efficaces, il est alors nécessaire que l'élève s'engage dans des actions lui permettant d'améliorer ses apprentissages (réaliser les activités proposées par l'enseignant, demander de l'aide à ses pairs ou encore réfléchir sur ses propres démarches dans une perspective métacognitive). Comme le rappelle Wiliam (2011, p. 12), « après tout, le meilleur feedback est inutile s'il n'est pas réutilisé ensuite » (p. 12).

L'implication de l'élève et des pairs est explicitement prise en compte dans le modèle. Pour les deux derniers processus, l'enseignant peut ainsi mettre en place des activités qui sollicitent les **interactions entre pairs** (sous la forme d'évaluations mutuelles ou de différentes modalités de travaux de groupe) de façon à leur permettre de s'entraider pour apprendre (cf. « Mobiliser les élèves de façon à ce qu'ils constituent des ressources pédagogiques les un pour les autres » dans la figure 3). Au final, l'objectif est que chaque élève apprenne et se sente responsable de ses apprentissages. Ainsi l'enseignant doit susciter cette responsabilisation et cet engagement auprès de chaque élève (cf. « Mobiliser les élèves de façon à ce qu'ils soient acteurs de leur propre apprentissage » dans la figure 3). Comme le précise Wiliam (2011), ce dernier élément relie explicitement le modèle au champ de recherche de l'apprentissage autorégulé (Andrade & Brookhart, 2016).

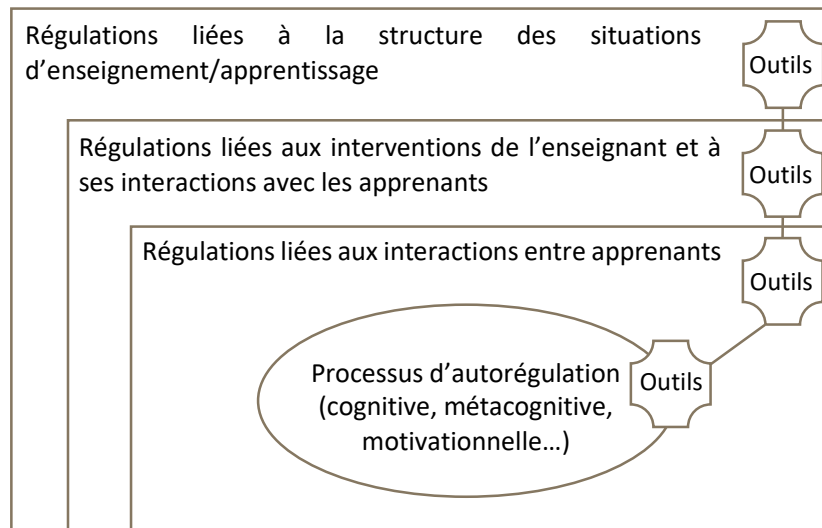
2. La modélisation proposée dans les travaux francophones

Dans un ouvrage consacré à la régulation des apprentissages, Allal (2007b) propose un modèle illustrant le processus de co-régulation soutenant les apprentissages en classe. L'originalité du modèle est de proposer d'analyser les différentes formes d'interactions en les situant au cœur d'une situation d'enseignement/apprentissage définie par le curriculum ou par l'enseignant. L'idée sous-jacente est que la situation va orienter la façon dont ces interactions vont se dérouler (Allal, 2020).

Le processus d'autorégulation (qui peut porter notamment sur les versants cognitif, métacognitif, et motivationnel) se situe au cœur du modèle ; il découle de la participation de l'apprenant à la situation d'enseignement/apprentissage et est soutenu par trois sources de régulations emboîtées : les régulations liées à la structure des situations d'enseignement apprentissage ; les régulations liées aux interventions de l'enseignant et à ses interactions avec les apprenants et les régulations liées aux interactions entre apprenants. Les « outils » constituent une quatrième source de régulation qui peut intervenir à chacun des niveaux, mais aussi aider à relier les niveaux entre eux (Allal, 2007b, 2020). La figure 4 illustre ce modèle²⁹.

²⁹ L'explication du modèle s'inspire largement des notes de cours de « didactique générale » de Fagnant (2021).

Figure 4 : Modélisation du processus de co-régulation dans une situation de classe (d'après Allal, 2007b – traduit d'Allal, 2020, p. 340)



Le premier niveau porte sur **la structure des situations d'enseignement/apprentissage**. Cette « structure » inclut les objectifs visés, les activités et le matériel didactique en lien avec le curriculum (Allal, 2020). Allal (2007b) considère en effet que la régulation des apprentissages est « influencée par la structure des situations et des tâches proposées aux apprenants, c'est-à-dire par les contenus choisis, leur formulation et leur formatage, par l'agencement des situations dans le temps et dans l'espace, par les consignes et le matériel fournis, par les modalités d'interactions sociales encouragées, permises, tolérées » (p. 15).

Au deuxième niveau, les régulations sont liées aux **interventions de l'enseignant et à ses interactions avec les apprenants** (avec la classe entière, avec un petit groupe d'élèves ou encore avec un seul élève). C'est dans cette catégorie que l'on peut situer les nombreuses recherches qui ont étudié les caractéristiques du feedback fourni par l'enseignant et ses effets sur les apprentissages³⁰. C'est aussi à ce niveau que l'on peut situer les gestes de régulation orchestrés par l'enseignant ; autrement dit, l'instauration d'un « processus d'étayage (*scaffolding*), tel que défini par Bruner dans son analyse des interactions régulatrices qui se produisent dans la zone proximale de développement de l'enfant » (Allal, 2007a, p. 16). C'est aussi ici que l'on situera les « conversations évaluatives » (Ruiz-Primo, 2011) qui peuvent être menées en groupe classe lors de phases de mise en commun ou de correction notamment. Allal (2020) précise aussi que les interventions de l'enseignant consistant à ajuster les situations d'apprentissage en fonction des besoins identifiés auprès des apprenants se situent également à ce niveau.

Au troisième niveau, les régulations sont liées aux **interactions entre apprenants**. Ces interactions constituent une « autre forme de médiation sociale présente dans presque toutes les situations d'apprentissage à l'école et en formation. (...) Ces interactions constituent une source de régulation liée aux relations entre la structure des situations et la dynamique interactive instaurée par les

³⁰ Une fois encore, nous renvoyons le lecteur intéressé au rapport de Calone et Lafontaine (2023).

participants » (Allal, 2007b, p. 17). Il peut s'agir d'interactions opérant dans le cadre de travaux de groupes, de phases de tutorat ou encore d'un travail d'évaluation mutuelle entre pairs.

Les **outils** constituent une quatrième source de régulation. Allal (2020) précise tout d'abord qu'ils assurent les liens entre les différents niveaux. L'enseignant peut par exemple sélectionner une grille d'évaluation (spécifiant les objectifs d'apprentissage et les critères) et animer une discussion avec toute la classe pour aider les élèves à bien comprendre les enjeux d'apprentissage ; cette grille peut ensuite être utilisée lors d'évaluations mutuelles entre pairs ou pour soutenir le processus d'autoévaluation de l'élève. L'auteure précise que les outils peuvent aussi soutenir les régulations interactives en rendant les objectifs plus explicites et le suivi des progrès plus systématique que s'il s'agissait d'interactions purement informelles. Enfin, elle mentionne que les outils permettent souvent d'enregistrer des traces d'évaluation qui peuvent être utilisées pour réguler les activités ultérieures. L'enseignant peut par exemple utiliser des grilles remplies par les élèves pour repérer des difficultés courantes et planifier de nouvelles activités d'apprentissage.

Au cœur du modèle, le **processus d'autorégulation** correspond à la régulation interne qui se produit au niveau de l'apprenant. Tout comme dans le modèle américain présenté précédemment (Wiliam, 2011), les liens avec le champ de recherche de l'apprentissage autorégulé (Andrade & Brookhart, 2016) sont indéniables. Toutefois, dans ce modèle situé, Allal (2020) considère que tout apprentissage en classe est co-régulé 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 co-régulation 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).

(3) Ce que l'on peut retenir des modélisations de l'évaluation-soutien d'apprentissage en classe

En cohérence avec les travaux développés précédemment, les deux modèles montrent que l'implication de l'élève dans le processus évaluatif est essentielle. Cette implication peut se matérialiser au travers des interactions entre l'enseignant et les élèves (notamment sous forme de « conversations évaluatives ») et lors des interactions entre pairs (notamment via des évaluations mutuelles) voire d'autoévaluation. Ce sont dès lors ces éléments spécifiques que nous développerons au point suivant (C.).

Les deux modèles envisagent également des liens entre le processus d'ESA et l'apprentissage autorégulé. De nombreux travaux s'attachent en effet à analyser ces liens, mais nous n'aurons pas l'occasion de les développer ici dans la mesure où ils pourraient à eux seuls faire l'objet d'un rapport ou d'une note de synthèse spécifiques. Nous renvoyons le lecteur intéressé vers quelques références qui sont en lien avec le modèle d'*assessment for Learning* développé par Laveault et Allal (2016b) :

- *Dans l'ouvrage de Laveault & Allal (2016a)*

Andrade, H. & Brookhart, S.M. (2016) The role of classroom assessment in supporting self-regulated learning. In Laveault, D. & Allal, L. (Eds). *Assessment for learning. Meeting the challenge of implementation* (pp. 293-309). Springer.

- *Un numéro thématique sur les liens entre évaluations en classes et apprentissage autorégulé*

[Présentation du numéro thématique] Brandmo, C., Panadero, E. & Hopfenbeck, T.N. (2020) Bridging classroom assessment and self-regulated learning, *Assessment in Education: Principles, Policy & Practice*, 27(4), 319-331.

[Article faisant explicitement des liens avec le modèle de corégulation d'Allal] Andrade, H. & Brookhart, S.M. (2020) Classroom assessment as the co-regulation of learning, *Assessment in Education: Principles, Policy & Practice*, 27(4), 350-372,

- *Une revue systématique récente :*

Andrade, H., Brookhart, S.M. & Yu, E.C. (2021). Classroom assessment as co-regulated learning: A systematic review. *Frontiers in Education*, 6, 1-18.

C. L'enseignant, les pairs et l'élève : trois sources importantes d'évaluation-soutien d'apprentissage en classe

Dans cette partie nous allons développer les résultats de recherche propres aux pratiques mentionnées précédemment : les conversations évaluatives entre l'enseignant et les élèves pendant l'apprentissage d'une part (1) ; l'évaluation mutuelle entre pairs et l'autoévaluation d'autre part (2).

1. Les conversations évaluatives entre l'enseignant et les élèves pendant l'apprentissage

Les interactions verbales en classe constituent des moments fréquents et privilégiés d'évaluation informelle. On parlera de « conversations évaluatives » à partir du moment où l'enjeu des interactions est de vérifier la compréhension des élèves et d'agir en conséquence pour soutenir leurs apprentissages (Primo-Ruiz, 2011).

Plus précisément, le concept d'évaluation formative informelle s'appuie sur l'idée que **la plupart des choses que les élèves et l'enseignant font en classe peuvent être considérées comme des opportunités évaluatives pour collecter des éléments à propos des apprentissages des élèves et de leur compréhension**. Cette évaluation informelle peut se réaliser sous différentes modalités permettant de recueillir différents types de traces d'apprentissage (Primo-Ruiz, 2011, p. 15) :

- (a) *des traces orales (ex. les questions posées par les élèves ou les réponses qu'ils fournissent aux questions qui leur sont posées, les dialogues entre élèves lors de travaux de groupes ou encore un dialogue entre l'enseignant et un élève singulier) ;*
- (b) *des traces écrites (ex. les notes prises par les élèves dans leur cahier d'exercices) ;*
- (c) *des traces graphiques (ex. des dessins, des schémas, des cartes mentales,...) ;*
- (d) *des traces non-verbales (ex. le langage corporel, la façon de se comporter en classe,...).*

Pour que l'on puisse parler d'évaluation, ces « traces » doivent être interprétées et utilisées pour orienter le cours de la leçon, en tenant compte du contexte dans lequel elles interviennent et des objectifs d'apprentissage visés.

Comme nous l'avons vu dans le modèle d'Allal (2007b), les régulations liées aux interactions enseignant-élèves peuvent se produire en groupe classe, avec un petit groupe d'élèves ou avec un seul élève. Lorsque l'enseignant interagit avec un élève en particulier, ces moments d'échanges peuvent prendre la forme d'un étayage qui soutient efficacement les apprentissages de cet élève, comme l'illustrent les propos d'Allal (2020) ci-dessous :

Lorsqu'il interagit avec un élève en particulier, l'enseignant peut tenter de fournir un étayage étroitement adapté à ce que cet élève dit et fait et ainsi promouvoir une co-régulation située dans la zone proximale de développement de l'élève. Dans une perspective vygotskienne, l'étayage est efficace si l'enseignant s'approprie des éléments des réponses de l'apprenant et que, parallèlement, l'apprenant progresse en s'appropriant des éléments des interventions de l'enseignant (Allal, 2020, p. 243).

Même si l'efficacité de ce type d'approche a pu être démontré lorsque l'enseignant interagit **individuellement avec un élève ou avec un petit groupe d'élèves** (Allal, 2020), le format le plus fréquent que l'on observe en classe reste celui des **interactions menées avec le groupe-classe entier**, lors des phases de mise en commun d'un travail ou lors de corrections collectives d'exercices notamment. S'appuyant sur les travaux de Black et Wiliam, Allal (2020) précise que les interactions verbales (ou les « conversations évaluatives ») durant les discussions en groupe-classe semblent constituer « l'un des moyens les plus puissants par lesquels les enseignants peuvent contribuer à la régulation de l'apprentissage des élèves » (p. 344). Interrogeant alors les conditions à remplir pour que ces conversations évaluatives puissent s'avérer efficaces, elle fait référence à la revue publiée par Ruiz-Primo (2011) « qui met en lumière l'évaluation formative informelle telle qu'elle s'incarne dans les dialogues pédagogiques organisés en groupe-classe » (p. 344).

Étant donné la complexité et la lourdeur du processus de recherche nécessaire pour analyser l'effet de ces évaluations informelles (enregistrements vidéos, retranscriptions et analyse des dialogues et de leurs effets au cours du temps), peu d'études ont pu mettre en évidence les effets positifs des liens entre la qualité des interactions et les apprentissages des élèves (Primo-Ruiz, 2011). La plupart des études répertoriées ont été menées sur de petits effectifs ou constituent des études de cas proposant des analyses fines des interactions observées³¹. Même si ces études ne permettent pas de généraliser les résultats, on s'accordera avec Ruiz-Primo (2011) pour reconnaître « que ces études apportent non seulement un éclairage sur le processus d'interaction, mais qu'elles peuvent aussi, lorsqu'on les considère toutes ensemble, conforter les résultats observés dans diverses études isolées » (p. 21) Toutes les études mentionnées ici montrent « un effet positif, sur l'apprentissage des élèves, des modèles d'interaction considérés comme souhaitables pour l'évaluation en classe (p. 21) . Même s'il ne s'agit pas d'une méta-analyse ou d'une revue systématique, nous nous appuyerons donc sur ce texte pour dégager un certain nombre de caractéristiques de ces pratiques d'évaluations informelles (ou de ces « conversations évaluatives ») qui semblent pouvoir porter leurs fruits³².

³¹ On trouve notamment de telles études dans le monde francophone (voir par exemple Fagnant, Dupont & Demonty, 2016 ; Mottier Lopez, 2007 ; Mercier-Brunel, 2019, pour ne citer que quelques exemples provenant respectivement de Belgique, de Suisse et de France).

³² Précisons aussi que, même si le texte date de 2011, c'est la seule synthèse de ce type mentionnée dans l'écrit plus récent d'Allal 2020 et que nous n'avons pas trouvé de revue plus récente.

Tout d'abord, nous retiendrons que, pour que ces pratiques d'évaluations informelles puissent impacter le processus d'enseignement et, à terme, l'apprentissage des élèves, il est essentiel que l'enseignant adopte un état d'esprit caractérisé par la volonté (Primo-Ruiz, 2011, p. 16) :

- *de se renseigner continuellement sur l'apprentissage des élèves (par ex., examiner les variations dans les réponses des élèves, la façon dont ils résolvent les problèmes ou les types d'erreurs qu'ils commettent) ;*
- *d'utiliser régulièrement les informations récoltées pour décider que faire ensuite (par ex., comment répondre à l'incompréhension des élèves, quelle démonstration pourrait aider les élèves à tester leurs hypothèses ou quel exercice proposer ensuite ?) ;*
- *d'examiner comment les pratiques pédagogiques soutiennent l'apprentissage des élèves (par ex., comment les tâches sélectionnées soutiennent-elles la compréhension des élèves et les guident-elles sur ce qu'il est important d'apprendre ?).*

Pour que les « conversations évaluatives » puissent s'avérer efficaces (Primo-Ruiz, 2011), il est nécessaire qu'elles soient **guidées par les objectifs d'apprentissage**. Ces objectifs doivent donc être précis et explicités auprès des élèves ; ces derniers doivent aussi comprendre les critères qui correspondent à l'atteinte de ces objectifs.

Les conversations évaluatives sont des **outils d'étayage qui soutiennent les apprentissages individuels des élèves** (grâce aux interactions avec l'enseignant et/ou avec des pairs plus avancés) ; ce sont aussi des **outils qui soutiennent les interactions sociales et la construction collective des connaissances**. Ce sont également des **outils d'enculturation** qui permettent d'immerger les élèves dans le langage et les habitudes culturelles d'une discipline. Pour être efficaces, ces conversations devraient être *dialogiques et interactives*. L'enseignant devrait encourager et prendre en compte une variété de réponses de façon à se faire une idée claire de la compréhension des élèves ; ces derniers devant se sentir autorisés à partager leurs points de vue et à discuter celui des autres.

Enfin, pour être efficaces, ces conversations évaluatives nécessitent aussi que l'enseignant dispose de bonnes connaissances didactiques (PCK ou connaissances pédagogiques de contenu selon Shulman, 1986, 2007) de façon à pouvoir interpréter les réponses des élèves et y réagir de façon rapide et contingente (Primo-Ruiz, 2011).

Encadré 4 : Pour en savoir plus... Discours monologiques ou dialogiques³³

Les recherches ont montré que l'enseignant parle entre 70 % et 80 % du temps scolaire ; elles indiquent également qu'il parle de plus en plus au fur et à mesure que l'on progresse dans la scolarité. De façon un peu caricaturale, on pourrait dire que, dans l'enseignement secondaire, « les élèves semblent venir à l'école pour regarder l'enseignant travailler » (Hattie, 2013).

L'enseignant parle ainsi bien plus que les élèves. Si cette « configuration » est en cohérence avec une méthode transmissive, on constate que le « flot de paroles » de l'enseignant semble prendre le pas dans les méthodes interrogatives. L'idée n'est évidemment pas de dire à l'enseignant de se taire pour lancer les élèves dans des exercices papier-crayon qui pourraient rapidement devenir lassants, mais plutôt de voir comment l'enseignant pourrait initier une « conversation productive à propos de

³³ Cet encadré est largement inspiré du syllabus du cours de « didactique générale » (Fagnant, 2021).

l'apprentissage » en vue de favoriser les apprentissages de tous les élèves (Hattie, 2013). La distinction entre discours « monologique » et « dialogique » apporte en ce sens un éclairage intéressant.

Hattie (2013) décrit l'enseignant de type « **monologique** » comme essentiellement concerné par la transmission de connaissances ; il se sent responsable de l'atteinte des objectifs fixés ; il utilise fréquemment une forme de discussion de type « récitation/question/réponse » de façon à vérifier qu'au moins quelques étudiants ont acquis une compréhension (généralement superficielle) des connaissances transmises. Dans cette configuration, c'est le flot de paroles de l'enseignant qui domine nettement et les questions posées nécessitent généralement des réponses composées de deux ou trois mots, guère plus. L'erreur n'a pas vraiment sa place dans le processus et les élèves ont donc généralement peur de ne pas bien répondre. Le discours « **dialogique** », quant à lui, vise à promouvoir une réelle communication entre l'enseignant et ses élèves aussi bien qu'entre élèves. L'enseignant cherche à faire prendre conscience à l'élève de l'importance de son point de vue ainsi que de celui de ses pairs ; il essaie d'aider les élèves à construire collectivement les connaissances visées et à leur donner du sens. Dans cette configuration, les élèves posent fréquemment des questions ; ils n'hésitent pas à donner leur point de vue et à proposer des réponses partielles, voire erronées, puisque celles-ci trouvent légitimement leur place dans le processus collectif.

Mercer et Littleton (2007, cités par Hattie, 2013) dressent les caractéristiques de ces deux types de configurations interactives. Les classes où le discours est essentiellement « **monologique** » se caractérisent par une dominance de *récapitulations* (les élèves sont invités à répéter ou à synthétiser ce qui a été vu juste avant), de *rappels factuels* (l'enseignant pose des questions brèves qui font appel à une information factuelle précise), de *reformulations* (l'enseignant ou un élève paraphrase la réponse d'un autre élève en vue de la préciser pour le reste de la classe) et d'*incitations indéterminées* (l'enseignant invite les élèves à réfléchir à ce qui s'est dit précédemment). Au contraire, les classes où le discours est essentiellement « **dialogique** » sont décrites comme *collectives* (les tâches d'apprentissage sont réalisées ensemble), *soutenant les élèves* (la classe cherche à explorer diverses idées, sans crainte de commettre des erreurs), *cumulatives* (les connaissances sont construites en s'appuyant sur les diverses idées partagées) et *orientées* (la leçon est guidée par des objectifs d'apprentissages précis).

Même si les deux types de discours peuvent avoir leur intérêt et trouver leur place dans les classes selon les moments de la leçon et les objectifs visés, on peut regretter que les premiers occupent très majoritairement le temps de parole dans de nombreuses classes alors que les seconds semblent davantage porteurs d'apprentissage pour les élèves (Hattie, 2013).

D'après Primo-Ruiz (2011), les « conversations évaluatives » sont mises en place selon un modèle cyclique qui débute par la *clarification des objectifs d'apprentissage* (rappeler les objectifs d'apprentissage et les buts visés par les activités réalisées ; faire des liens avec ces objectifs durant la conversation) et se poursuit par la *collecte d'informations* (via des questionnements pouvant faire appel à des connaissances déclaratives, procédurales ou conditionnelles ; ces questionnements pouvant demander des clarifications, des confirmations, des justifications,...), *leur interprétation* et *les actions à mener en conséquence*.

Comme il n'est pas facile d'interpréter les réponses des élèves et d'y réagir dans le feu de l'action, Primo-Ruiz (2011) identifie une série de stratégies que l'enseignant peut employer pour impliquer les élèves dans la conversation évaluative, sans porter directement un jugement sur leurs réponses. Il

peut ainsi inviter les élèves à reformuler, clarifier, synthétiser ou répéter leur propre réponse ; faire des liens entre leur réponse et celles fournies par d'autres élèves ; confronter les réponses fournies par plusieurs élèves (par exemple en les organisant sous une forme schématique pour en débattre ensuite). Il peut aussi reposer des questions à l'élève pour l'inviter à poursuivre son raisonnement (lui renvoyer la balle) ou inviter d'autres élèves à réagir (renvoyer la balle d'un élève à l'autre).

Au niveau des actions que peut entreprendre l'enseignant en fonction de l'interprétation qu'il peut faire des réponses des élèves, l'auteure mentionne quatre stratégies : donner des explications complémentaires ; fournir des feedbacks³⁴ ; comparer les réponses fournies par plusieurs élèves en identifiant les forces et les faiblesses, les points communs et divergences,... et faire des liens explicites entre des éléments de la discussion et les objectifs d'apprentissage visés.

(4) Ce que l'on peut retenir des conversations évaluatives

Une série de travaux semblent converger pour pointer l'importance des évaluations formatives informelles et des « conversations évaluatives » (Primo-Ruiz, 2011). D'après les travaux de recherche, ces « conversations évaluatives » doivent être ciblées sur les objectifs d'apprentissage ; elles doivent permettre de vérifier la compréhension des élèves et de soutenir leurs apprentissages. Elles gagnent en efficacité si elles s'apparentent à un « discours dialogique » (Hattie, 2013) qui fait réfléchir les élèves et intègre leurs réflexions dans la conversation. Le climat de classe est important pour que les élèves se sentent autorisés à partager leur point de vue et à avouer leurs incompréhensions.

On s'accordera toutefois avec Allal (2020) pour reconnaître que ces évaluations formatives informelles, qui s'accompagnent de régulations interactives orchestrées par l'enseignant « dans le feu de l'action », ne sont pas suffisantes pour assurer un enseignement de qualité. Elles doivent en effet s'accompagner d'évaluations formatives plus formelles (sous forme de tests ou de tâches spécifiquement pensées à cette fin) de façon à pouvoir déployer aussi des formes de régulations rétroactives (par exemple des remédiations ciblées sur les difficultés identifiées au terme d'une séquence) et proactives (par exemple en anticipant différentes formes de soutien qui pourraient être mises en place dans une nouvelle activité).

2. L'évaluation mutuelle entre pairs et l'autoévaluation

L'autoévaluation et l'évaluation mutuelle entre pairs (ou évaluation par les pairs) impliquent que les élèves portent des jugements sur leurs propres performances scolaires ou sur celles des autres : ils évaluent dans quelle mesure les critères ont été atteints et fournissent un FB critérié (sous la forme de notes et/ou de commentaires) aux autres élèves ou à eux-mêmes (Sanchez *et al.*, 2017). Dans le processus d'autoévaluation, l'élève doit porter un jugement sur son propre travail en vue de générer un feedback lui permettant de l'améliorer (Andrade, 2019) ; dans l'évaluation par les pairs, c'est sur le travail d'autrui que portent ces jugements, mais l'objectif formatif reste le même (Li *et al.*, 2020).

Les notions de **feedback** et de possibilité **d'amélioration** du travail (de mettre en œuvre des procédures de régulation) sont essentielles (Andrade, 2019) pour considérer réellement qu'il s'agit

³⁴ L'auteur fait ici référence explicitement aux différentes synthèses réalisées sur les différents types de feedbacks et leur condition d'efficacité. Nous renvoyons pour notre part au rapport de Calone & Lafontaine (2023).

d'autoévaluation ou d'évaluation mutuelle entre pairs. *A contrario*, inviter simplement l'élève à évaluer son propre travail ou celui d'autrui relève davantage de procédures d'auto-notation (Rey & Feyfant, 2014) ou de simples corrections (par exemple, noter sa propre copie ou celle d'autrui en s'appuyant sur la grille de critères ou sur le correctif fournis par l'enseignant) que de réelles pratiques d'autoévaluation ou d'évaluation mutuelle. En effet, dans son acception ambitieuse, l'autoévaluation constitue une réelle activité d'apprentissage : « c'est une manière d'encourager les élèves à réfléchir sur ce qu'ils ont appris, à chercher les moyens d'améliorer leur apprentissage, et à planifier ce qui leur permet de progresser en tant qu'apprenant et d'atteindre leurs objectifs » (Rey & Feyfant, 2014, p. 28). L'autoévaluation implique des réflexions métacognitives (Mottier Lopez, 2015, p. 85) mobilisant des questionnements du type : « Qu'est-ce que je sais ? Qu'est-ce que je sais faire ? Comment je m'y prends ? Qu'est-ce que je peux modifier ? ». Le même type d'attentes devrait se transférer à l'évaluation mutuelle entre pairs dans la mesure où l'un des enjeux de cette pratique est que les élèves développent des stratégies métacognitives et deviennent davantage autonomes (Li *et al.*, 2020 ; Sanchez *et al.*, 2017).

Malgré l'intérêt théorique que rencontrent ces deux types de pratiques, les résultats quant à leur efficacité respective sont variables, sans doute parce que les recherches analysées s'appuient sur des pratiques trop diversifiées (Andrade, 2019 ; Double *et al.*, 2020 ; Li *et al.*, 2020 ; Sanchez *et al.*, 2017). Ainsi Andrade (2019) explique par exemple que, sous le couvert d'autoévaluation, on rencontre des pratiques invitant simplement les élèves à coller un émoticône sur leur propre travail, à estimer le nombre de réponses correctes produites dans un test de mathématiques ou encore à utiliser une grille critériée pour repérer les forces et les faiblesses d'un premier jet d'écriture. Dans sa revue critique, Andrade (2019) fait alors appel à des clarifications conceptuelles et renvoie notamment au champ de l'apprentissage autorégulé dans la mesure où l'autoévaluation met en œuvre une série de processus qui relèvent de ce champ de recherches (définition du but, monitoring, métacognition et régulation notamment). Les trois méta-analyses répertoriées ici (Double *et al.*, 2020 ; Li *et al.*, 2020 ; Sanchez *et al.*, 2017) montrent quant à elles un effet bénéfique des pratiques d'évaluation par les pairs ou d'autoévaluation (taille de l'effet de l'ordre de 0,30³⁵) : mettre en place des évaluations par les pairs serait non seulement plus efficace que de ne pas procéder à des évaluations, mais aussi plus efficace que lorsque c'est l'enseignant qui mène les évaluations (Double *et al.*, 2020 ; Li *et al.*, 2020). Les résultats de Sanchez *et al.* (2017) vont dans le même sens, tout en pointant un effet légèrement plus important pour l'autoévaluation comparativement à l'évaluation par les pairs. Cette différence ne se retrouve pas dans les deux autres études (Double *et al.*, 2020 ; Li *et al.*, 2020) qui concluent à l'absence de différence significative entre les deux pratiques.

Parmi ces trois méta-analyses, les deux plus récentes (Double *et al.*, 2020 ; Li *et al.*, 2020) cherchent aussi à identifier les conditions d'efficacité de l'évaluation par les pairs. Nous synthétisons ici les résultats principaux :

³⁵ L'étude de Double *et al.* (2020) s'appuie sur 54 études menées en primaire, en secondaire et dans le supérieur et obtient une taille de l'effet moyenne de 0,31. L'étude de Li *et al.* (2020) porte sur le même public (grades K-12 et supérieur) ; elle inclut 58 études et conduit à une taille de l'effet moyenne de 0,29. L'étude de Sanchez *et al.* (2017) porte quant à elle uniquement sur le primaire et le secondaire (grades 3-12) ; elle intègre 12 études et conduit à une taille de l'effet de 0,29 pour les évaluations par les pairs et de 0,33 pour l'autoévaluation comparativement à des élèves qui n'ont mis en place aucune de ces deux procédures.

- aucune différence n'apparaît en fonction du **niveau scolaire concerné** (primaire ou secondaire), hormis éventuellement en ce qui concerne l'enseignement supérieur (mais les résultats des deux études se contredisent, Double *et al.*, 2020 ; Li *et al.*, 2020) ;
- aucune différence significative n'est mise en évidence en fonction du **domaine de contenu (de la discipline) investigué(e)** (Double *et al.*, 2020 ; Li *et al.*, 2020).
- l'effet semble plus important lorsque **les élèves ont reçu une formation** pour évaluer les travaux de leurs pairs (Li *et al.*, 2020) et lorsqu'ils **disposent de critères clairs** pour effectuer cette évaluation (Li *et al.*, 2020) ;
- le **caractère informatisé ou non des FB** reste sujet à débat (ce serait plus efficace *via* ordinateur que papier-crayon d'après Li *et al.*, 2020, mais ce constat n'est pas vérifié par Double *et al.*, 2020 qui ne notent pas de différences significatives entre les deux modalités) ;
- la possibilité de **jouer tour à tour le rôle d'évaluateur et d'évalué** semble constituer un atout (d'après Li *et al.*, 2020 mais Double *et al.*, 2020 pointent que c'est une de limites de leur étude que ne pas avoir pu distinguer ces deux modalités, ce qui selon eux ne permet pas de conclure à ce qui est le plus important entre donner des FB et en recevoir) ;
- le **caractère obligatoire ou non de la participation** à l'évaluation mutuelle ne semble pas avoir d'effet (Double *et al.*, 2020 ; Li *et al.*, 2020) ;
- le fait que **le FB prenne la forme de commentaires ou de notes** ne conduit pas à des différences significatives (Li *et al.*, 2020 ; Double *et al.*, 2020). Pour Li *et al.* (2020), il semblerait même qu'attribuer des scores soit plus efficace que de proposer des commentaires (plus précisément, les résultats de « scores seuls » sont proches de ceux de la modalité « scores + commentaires » et ces deux modalités semblent plus efficaces que la modalité « commentaires seuls »). Ces résultats sont surprenants et vont à l'encontre de ce que défendent d'autres auteurs (voir par exemple Bourgeois & Laveault, 2015). Li *et al.* (2020) estiment qu'une hypothèse explicative à ces résultats serait que « lorsque les élèves fournissent uniquement des commentaires, il se pourrait que la procédure d'évaluation mise en place soit peu structurée et que les critères d'évaluation ne soient pas suffisamment explicités » (p. 203).
- le **caractère oral ou écrit du FB** ne semble pas non plus engendrer de différences significatives (Li *et al.*, 2020 ; Double *et al.*, 2020) : même s'il semble cohérent de combiner les deux modalités de façon à proposer un FB plus élaboré (Li *et al.*, 2020), le fait que le FB soit ou non accompagné d'un dialogue entre élèves ne révèle pas d'effet significatif (Double *et al.*, 2020) ;
- il semble que le processus est plus efficace lorsque **plusieurs sessions d'évaluation par les pairs** sont mises en place (Li *et al.*, 2020) ; même si les différences ne sont pas significatives, l'effet est assez faible lorsque les élèves n'ont qu'une occasion de mettre en œuvre le processus (Double *et al.*, 2020) ;
- le fait que les **tâches de réinvestissement ne nécessitent pas de transfert, nécessitent un transfert proche ou un transfert éloigné** ne fait pas non plus l'objet de différences significatives, mais les résultats semblent montrer que les gains sont plus importants quand les tâches ne nécessitent pas de transfert (Double *et al.*, 2020) ;
- il n'a pas été possible de distinguer les études selon que l'évaluation par les pairs relevait d'un **caractère formatif ou sommatif** (ce que Double *et al.*, 2020 soulignent comme étant une limite de leur étude).

Finalement, les seules variables qui semblent avoir un effet important sont le fait que les élèves soient formés aux pratiques d'évaluation mutuelle entre pairs et qu'ils disposent de critères clairs à cette fin (Li *et al.*, 2020). Sanchez *et al.* (2017) appuient cette idée en notant qu'il serait important que les enseignants reçoivent une formation pour les aider à mettre en place ce type de pratiques. Ces formations pourraient s'appuyer sur les lignes directrices proposées par Ross (2006, cité par Sanchez *et al.*, 2017) pour implémenter avec succès des pratiques d'autoévaluation ou d'évaluation par les pairs en classe :

|| Définir les critères par rapport auxquels les élèves évaluent leur propre travail ; leur enseigner comment utiliser ces critères ; leur donner un feedback sur leur note et leur permettre de suivre leurs progrès pour améliorer leurs performances (Sanchez *et al.*, 2017, p. 1050).

L'auteur précise aussi que d'autres chercheurs ont suggéré d'ajouter les éléments suivants : prévoir suffisamment de temps pour permettre aux élèves de revoir leur travail après la procédure d'évaluation mutuelle et utiliser ces pratiques à des fins formatives plutôt que sommatives (Brown & Harris, 2013, cités par Sanchez *et al.*, 2017). Avoir l'occasion d'améliorer son travail (de le réguler) après avoir obtenu un feedback par les pairs ou *via* sa propre autoévaluation sont deux éléments en cohérence avec les éléments soulignés dans la revue critique d'Andrade (2019), tout comme la nécessité que ce type de pratiques s'ancre dans des évaluations-soutien d'apprentissage.

(5) Ce que l'on peut retenir des pratiques d'évaluation mutuelle entre pairs et d'autoévaluation

Les travaux mettent en évidence des effets positifs des pratiques d'évaluation mutuelle entre pairs et des pratiques d'autoévaluation³⁶.

Pour être efficaces, ces pratiques doivent s'appuyer sur des critères clairs et fournir des feedbacks précis aux élèves. Il est dès lors tout d'abord important de s'assurer que les critères soient bien compris des élèves. Il semble ensuite nécessaire d'aider les élèves à structurer leurs commentaires pour que ces derniers apportent une plus-value à la simple notation. En ce sens, il semble important que les élèves reçoivent une formation pour évaluer les travaux de leurs pairs (Li *et al.*, 2020) et qu'ils aient plusieurs occasions de mettre en œuvre ce type de pratiques (Double *et al.*, 2020). Enfin, il est tout aussi important que les élèves soient incités à réellement prendre en compte ces feedbacks et de prévoir du temps pour qu'ils puissent les utiliser pour réguler leur travail (Sanchez *et al.*, 2017), et ceci que les feedbacks proviennent des pairs ou d'un processus d'autoévaluation (Andrade, 2019).

³⁶ Comme le soulignent Double *et al.* (2020), ces résultats rejoignent ceux de synthèses plus narratives ou d'études plus qualitatives qui appuient l'intérêt de ces pratiques. C'est en se basant sur ces études que les auteurs avaient ciblé les variables dont ils ont tenté de mesurer l'effet différentiel selon les modalités comparées. Les résultats de leurs travaux, tout comme ceux de Li *et al.* (2020), ne permettent toutefois pas de mettre en évidence de différences significatives. Une raison pourrait être que « les caractéristiques des feedbacks retenues dans les synthèses narratives sont davantage motivées par des considérations théoriques que par des données empiriques quantitatives » Double *et al.*, 2020, p. 500). Une autre explication est que ces variables sont complexes et peuvent prendre des modalités très diversifiées (par ex., il existe une multitude de modalités de grilles d'évaluation et de procédures pour rendre des commentaires appuyés sur ces grilles) rendant difficile toute agrégation quantitative des résultats (Double *et al.*, 2020). Ce type de critique avait déjà été formulé par Bennett (2011) quant aux questionnements que posaient les méta-analyses sur des pratiques diversifiées de mise en œuvre de l'ESA. En accord avec les auteurs, on plaidera alors pour des études complémentaires qui cherchent à manipuler expérimentalement certaines variables pour en mesurer les effets, et ceci dans des champs disciplinaires contrastés.

III. Mettre en œuvre une évaluation-soutien d'apprentissage en classe : compétences et développement professionnel des enseignants

Plusieurs revues systématiques (Heitink *et al.*, 2016 ; Schildkamp *et al.*, 2020 ; Yan *et al.*, 2021) ont ciblé les facteurs qui influencent la mise en œuvre d'une évaluation-soutien d'apprentissage en classe (ESA). Ces études mettent en évidence l'influence d'un certain nombre de **variables contextuelles**, comme une culture d'école qui encourage la collaboration entre enseignants (Heitink *et al.*, 2016 ; Schildkamp *et al.*, 2020) ou encore le soutien fourni par l'école, les conditions de travail et les caractéristiques des élèves (Yan *et al.*, 2021). Ces études ont aussi mis en évidence l'influence d'un certain nombre de **variables personnelles** telles que les attitudes envers l'ESA (Schildkamp *et al.*, 2020), le sentiment d'efficacité personnelle (Yan *et al.*, 2021) ou encore les connaissances et compétences en matière d'évaluation (Heitink *et al.*, 2016 ; Schildkamp *et al.*, 2020 ; Yan *et al.*, 2021).

Dès leurs premiers travaux, Black et Wiliam (1998a) avaient déjà souligné l'importance de la culture de classe ; les études précitées élargissent les facteurs contextuels à prendre en compte en envisageant la culture de l'école voire les attentes du système éducatif. On peut ici faire des liens avec les propos de Crahay *et al.* (2019), qui soulignent qu'un système éducatif ancré dans des valeurs compétitives et d'excellence scolaire, procédant selon une logique méritocratique, pourrait s'avérer en contradiction avec les valeurs de l'ESA. Toujours en référence aux premiers travaux de Black et Wiliam (1998a) qui précisaient que la mise en place de l'ESA ne pouvait se limiter à quelques adaptations de surface, on comprend la complexité que requiert la modification des pratiques d'enseignement et l'importance que les enseignants soient soutenus par leur direction (Yan *et al.*, 2021) et soient invités à collaborer entre eux (Schildkamp *et al.*, 2020).

En ce qui concerne les variables personnelles, l'étude de Schildkamp *et al.* (2020) a permis de mettre au jour un ensemble de connaissances et compétences qui semblent influencer la mise en œuvre de l'ESA. Ils identifient ainsi deux éléments déjà pointés par Bennett (2011) dans sa note critique : les compétences en matière d'évaluation (*assessment literacy*, Mertler, 2004) et les connaissances didactiques (*Pedagogical Content Knowledge*, Shulman, 1986, 2007), ainsi que d'autres facteurs que l'on retrouve dans la plupart des travaux depuis ceux de Black et Wiliam (1998a), comme la capacité à fixer des objectifs précis, à fournir des feedbacks de qualité ou encore à orchestrer des « conversations évaluatives » de qualité (Ruiz-Primo, 2017).

Le concept d'*assessment literacy* est très présent dans de nombreux travaux américains (voir DeLuca *et al.*, 2016 ; Gotch & French, 2014 ; Khalid *et al.*, 2021). Les enseignants qui disposent de bonnes compétences évaluatives favoriseraient ainsi la motivation des élèves, leur engagement dans l'apprentissage et l'atteinte d'objectifs de haut niveau (DeLuca *et al.*, 2019).

Mais que recouvre ce concept ? Nous reprendrons ici la définition proposée par Mertler (2004) qui, d'après Pasquini (2018) est assez typique des recherches menées dans ce domaine :

Le concept d'assessment literacy a été défini comme "la possession de connaissances sur les principes de base d'une bonne pratique d'évaluation ; ce qui comprend la connaissance des terminologies, le développement et l'utilisation de méthodologies et de techniques d'évaluation, la familiarité avec les critères de qualité en matière d'évaluation... et la familiarité avec des alternatives quant aux façon traditionnelles d'évaluer l'apprentissage" (Paterno, 2001, cité par Mertler, 2004, pp. 50-51).

Mertler (2004) précise aussi que certaines études s'appuient sur une description des caractéristiques d'un enseignant qui dispose de bonnes compétences évaluatives, plutôt que de proposer une définition formelle.

Les enseignants que l'on peut qualifier comme ayant de bonnes compétences évaluatives connaissent les bonnes pratiques d'évaluation et de communication :

- *ils comprennent quelles méthodes d'évaluation utiliser pour recueillir des informations fiables et apprécier où les élèves en sont dans leur apprentissage ;*
- *ils communiquent efficacement les résultats des évaluations, que cela passe par des bulletins scolaires comprenant des notes, par la restitution des résultats à des contrôles, par un portfolio ou encore par des rencontres ;*
- *ils savent utiliser l'évaluation pour maximiser la motivation et l'apprentissage des élèves en les impliquant comme partenaires à part entière dans l'évaluation, aussi bien dans le recueil d'informations que dans leur communication (Center for School Improvement and Policy Studies, Boise State University, n.d., cité par Mertler, 2004, p. 51).*

De nombreux travaux se sont attelés à développer des « outils » permettant d'évaluer ces compétences évaluatives chez les enseignants (voir De Luca *et al.*, 2016 pour une revue de ces outils). La plupart de ces « outils » s'appuient sur les recommandations de l'*American Federation of Teachers* (AFT, 1990) et opérationnalisent des mises en situation et des questionnements invitant les enseignants à positionner leurs connaissances et pratiques évaluatives vis-à-vis des différentes dimensions recouvertes par les recommandations de l'AFT (De Luca *et al.*, 2016). Kruze *et al.* (2020, p. 110) énumèrent ces recommandations comme suit :

- 1. Choisir les méthodes d'évaluation appropriées ;*
- 2. Développer des méthodes d'évaluation ;*
- 3. Faire passer une évaluation, la noter et interpréter les résultats ;*
- 4. Utiliser les résultats d'évaluations pour prendre des décisions ;*
- 5. Développer des procédures de notation ;*
- 6. Communiquer les résultats d'évaluations ;*
- 7. Identifier des pratiques d'évaluation non éthiques.*

Le questionnaire qui semble le plus fréquemment utilisé dans les études (voir DeLuca *et al.*, 2016 pour une synthèse) est le questionnaire « ALI » (*Assessment Literacy Inventory*) développé par Mertler et Campbell (2005) ou une de ses variantes. Il se compose de scénarios décrivant des situations d'enseignement, chacun composé de plusieurs questions à choix multiples testant les connaissances des enseignants à propos de leurs pratiques évaluatives.

S'appuyant sur ces outils, des travaux montrent que les jeunes enseignants semblent se sentir particulièrement mal préparés à l'évaluation (DeLuca *et al.*, 2019) et, même s'il semble que l'expérience du métier fasse en partie la différence (Mertler, 2004), la plupart des travaux continuent à pointer des lacunes auprès des enseignants plus expérimentés (DeLuca *et al.*, 2016). Plusieurs études se sont dès lors attelées à mettre en place des programmes de développement professionnel qui visent à combler ces lacunes.

Nous n'avons pas trouvé de méta-analyse, de revue systématique ou d'article de synthèse qui font le point sur l'efficacité des programmes de développement professionnel, mais on peut mentionner ici quelques travaux qui ont montré une certaine efficacité de ce type de programme.

Par exemple, Mertler (2009) a mis en place un programme de formation à *l'assessment literacy* auprès d'enseignants du primaire (N=7) provenant de différents établissements. La formation visait à aider les enseignants à développer une compréhension plus profonde des différents processus d'évaluation et à améliorer leurs compétences à évaluer les apprentissages de leurs élèves. Ils étaient amenés à travailler sur la construction et l'utilisation d'instruments d'évaluation, ainsi que sur l'interprétation des résultats à des fins sommatives, mais aussi diagnostiques et formatives. La comparaison des scores obtenus par les enseignants au questionnaire ALI (Mertler & Campbell, 2005) ainsi que l'analyse des journaux de bord remplis par les enseignants tout au long de la formation a conduit l'auteur à conclure que la formation était efficace, « comme l'indique l'augmentation spectaculaire des résultats au post-test par rapport à ceux du pré-test, et comme le montre peut-être encore davantage l'examen critique des journaux de bord » (Mertler, 2009, p. 101).

L'étude de Mertler (2009) porte sur un très petit nombre d'enseignants et ne décrit pas précisément le programme de formation mis en place. Dans une perspective assez proche, l'étude de Kruze *et al.* (2020) implique quant à elle 96 futurs enseignants du primaire et du secondaire (grades 4-9 et 7-12). Le programme s'étalait sur une période de 17 semaines, incluant 11 semaines de cours à l'université, 4 semaines de stage dans les classes et 2 semaines durant lesquelles ils ont été amenés à participer à des séminaires de pratiques réflexives à l'université. Les cours portaient sur six éléments centraux de l'évaluation, en lien les recommandations de l'AFT présentées précédemment : (1) conception d'une évaluation sommative classique ; (2) conception d'une grille critériée pour noter une évaluation sommative ; (3) conception et mise en œuvre d'une évaluation formative ; (4) réflexion sur la valeur ajoutée d'une évaluation ; (5) implémentation d'une évaluation et réflexion sur le processus mis en œuvre ; (6) interprétation des résultats d'une évaluation et communication de ceux-ci. À titre illustratif, le Tableau 5 détaille trois éléments centraux de ce programme de formation.

Tableau 5 : Description de quelques éléments centraux du programme de formation développé dans l'étude de Kruze *et al.* (adapté de Kruze *et al.*, p. 114)

Éléments principaux	Description
Élément 1. Construction d'une évaluation sommative classique	Les futurs enseignants ont dû concevoir un test écrit cohérent avec les objectifs et les contenus pédagogiques, ainsi qu'avec les habiletés cognitives visées (en lien avec la taxonomie de Bloom). Le test devait comprendre les consignes de passation, les questions et un guide ou une grille de correction. Il devait aussi comprendre au moins trois types de questions différentes parmi les modalités suivantes : questions binaires de type vrai-faux ; questions proposant des éléments à associer, questions à choix multiple ; phrases à compléter ; questions ouvertes à réponse brève et questions ouvertes demandant un développement.
Élément 3. Conception et mise en œuvre d'une évaluation formative	Les futurs enseignants ont planifié et mis en œuvre en œuvre une évaluation formative au cours de leur stage pratique. Ils ont également dû porter un recul réflexif sur l'expérience menée.
Élément 6. interprétation des résultats d'une évaluation et communication de ceux-ci	Les futurs enseignants ont utilisé les résultats obtenus par un élève à une évaluation pour écrire une lettre à un parent en vue d'expliquer les résultats de leur enfant. La lettre devait être écrite dans un langage clair et non technique ; elle devait expliquer les forces et les faiblesses de l'élève telles que révélées par les résultats du test. Les étudiants devaient aussi formuler des recommandations pédagogiques cohérentes avec les résultats du test.

Tout comme dans l'étude de Mertler (2009), l'efficacité de l'approche a été mesurée en comparant les scores obtenus au questionnaire ALI (Mertler & Campbell, 2005) avant et après la formation. Un questionnaire centré sur le sentiment d'efficacité personnelle des enseignants a également été administré. Kruze *et al.* (2020) concluent que l'impact de la formation est révélé par l'augmentation des scores des participants aux deux types de questionnaires.

Si ces deux études ont montré qu'une formation pouvait impacter les connaissances des enseignants en matière d'évaluation, on pointera toutefois qu'elles ne comparent par les résultats à ceux de groupes contrôle qui auraient reçu un autre type de formation, ce qui en limite d'une certaine façon la portée. Plus fondamentalement, on s'accordera avec Pasquini (2018) pour reconnaître que les mesures effectuées *via* le questionnaire ALI correspondent à une mesure assez théorique des connaissances en jeu et que les questionnements restent ouverts quant à la façon dont ces connaissances peuvent être réellement mises en œuvre dans les pratiques de classes en tenant compte des contextes spécifiques de leur implémentation. Autrement dit, si ces études mesurent des connaissances évaluatives (et si l'on peut s'accorder sur l'importance de disposer de telles connaissances), des études complémentaires seraient nécessaires pour réellement parler de « compétences évaluatives » qui affectent positivement les pratiques réelles et ont un impact positif sur les apprentissages des élèves³⁷.

Notons par ailleurs que les recherches sur l'*assessment literacy* ne sont pas ancrées directement sur les modèles de l'ESA que nous avons développés tout au long de cette synthèse. Même si l'évaluation formative est prise en compte, les auteurs s'intéressent largement à l'évaluation sommative. On s'accordera d'ailleurs avec Brookhart (2011) pour regretter que cette approche ne prend pas suffisamment en compte les travaux récents en matière d'*assessment for learning*³⁸. De plus, en cohérence avec les points d'attention stipulés par Bennett (2011), Brookhart (2011) regrette aussi le manque de place accordée aux connaissances didactiques et aux contenus disciplinaires³⁹.

Enfin, malgré les critiques précédemment émises, certains auteurs ont néanmoins tenté de proposer un certain nombre de recommandations issues d'une analyse de la littérature existante. Ainsi, dans leur note de synthèse sur les programmes de développement professionnel en évaluation formative, Trunbull et Gerzon (2013) établissent une série de caractéristiques que devraient satisfaire les

³⁷ On retrouve le même type de débat dans les travaux portant sur le PCK. Même si plusieurs travaux ont pu montrer des liens entre ces PCK et les performances des élèves (voir Depaepe *et al.*, 2013 pour une synthèse en mathématiques), la question de savoir comment ces connaissances s'opérationnalisent dans les pratiques de classes et comment ces pratiques affectent les apprentissages des élèves est aussi au cœur des travaux récents dans ce domaine (voir Depaepe *et al.*, 2020 pour une synthèse).

³⁸ D'autres études proposent des programmes de développement professionnel qui ne sont pas en lien avec les recherches sur l'*assessment literacy*. Nous n'aurons pas l'occasion de les répertorier ici parce que cela pourrait faire l'objet d'une note de synthèse complémentaire. Mentionnons toutefois, à titre illustratif, l'étude d'Anderson et Palm (2017) qui mesure l'effet d'un programme de développement professionnel ancré dans le modèle d'AfL que nous avons présenté précédemment (Thompson & Wiliam, 2008, cités par Wiliam, 2011). Les auteurs ont pu mettre en évidence non seulement des croyances plus positives, de la part des enseignants, quant aux bienfaits de l'AfL, mais aussi des progrès plus élevés, mesurés via un test de mathématiques, observés auprès des élèves issus des classes des enseignants du primaire (grade 4) qui ont suivi la formation comparativement à un groupe contrôle qui n'a pas reçu de formation spécifique au développement de l'AfL.

³⁹ Dans le monde francophone, de nombreux travaux ancrent la problématique du développement professionnel des enseignants dans des recherches collaboratives. Parmi ces travaux diversifiés, certains cherchent à soutenir le développement d'une ESA en classe, mais on peut faire le même type de constat quant au peu de place accordée aux didactiques et aux entrées disciplinaires (voir Fagnant & Goffin, 2021 pour une synthèse).

programmes de développement professionnel. Selon les auteurs, les programmes efficaces devraient notamment 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.*

(6) Ce que l'on peut retenir des éléments susceptibles de favoriser la mise en œuvre d'une évaluation-soutien d'apprentissage en classe

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).

Références

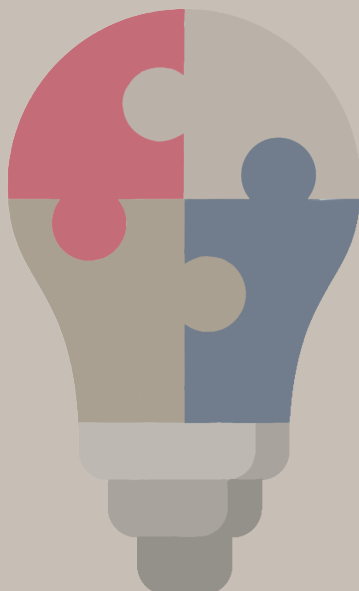
- Allal, L. (1991). Vers une pratique de l'évaluation formative. De Boeck.
- Allal, L. (2007a). Evaluation : lien entre enseignement et apprentissage. In V. Dupriez & G. Chapelle (Eds.). Enseigner (pp. 139-149). Presses Universitaires de France.
- Allal, L. (2007b). Evaluation : lien entre enseignement et apprentissage. In V. Dupriez & G. Chapelle (Eds.). Enseigner (pp. 139-149). Presses Universitaire de France.
- Allal, L. (2011). Pedagogy, didactics and the co-regulation of learning: a perspective from the French-language world of educational research. *Research Papers in Education*, 26(3), 329-336.
- 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: Evaluation-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 Centre pour la recherche et l'innovation dans l'enseignement (Ed.). *L'évaluation formative pour un meilleur apprentissage dans les classes secondaires* (265-290).Edition OCDE.
- Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory Into Practice*, 41, 255–260.
- Anderson, C. & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction*, 49, 92-102.
- Andrade, H.L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 1-13.
- Andrade, H. & Brookhart, S.M. (2016) The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L., Allal (Eds.). *Assessment for learning. Meeting the challenge of implementation* (pp. 293-309). Springer.
- Andrade, H. & Brookhart, S.M. (2020) Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice*, 27(4), 350-372,
- Andrade, H., Brookhart, S.M. & Yu, E.C. (2021). Classroom assessment as co-regulated learning: A systematic review. *Frontiers in Education*, 6, 1-18.
- Bennett, E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.

- Black, P., & Wiliam, D. (2018) Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575.
- Bourgeois, L. & Laveault, D. (2015). Evaluation par les pairs à l'écrit : qualité des rétroactions pour soutenir la phase de révision. In P.-F. Coen et L. M. Bélair (dir.), *Évaluation et autoévaluation. Quels espaces de formation ?* (pp. 101-117). De Boeck.
- Brandmo, C., Panadero, E. & Hopfenbeck, T.N. (2020) Bridging classroom assessment and self-regulated learning. *Assessment in Education: Principles, Policy & Practice*, 27(4), 319-331.
- Briggs, D., Ruiz-Primo, M. A., Furtak, E. & Shepard, L. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13–17.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, 30(1), 3-12
- Calone, A. & Lafontaine, D. (2023). L'impact des différents types de feedbacks en milieu scolaire. Rapport réalisé pour le Cnesco dans le cadre de la conférence de consensus sur l'évaluation scolaire de 2022.
- Crahay, M., Mottier Lopez, L., & Marcoux, G. (2019). L'évaluation des élèves : Docteur Jekyll and Mister Hyde de l'enseignement. In M. Crahay (Ed.), *Peut-on lutter contre l'échec scolaire ?* (pp. 358-425). De Boeck.
- DeLuca, C., Coombs, A., MacGregor, S., & Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education*, 4, 1-10.
- Deluca, C., Lapointe-Mcewan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272.
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12-25.
- Depaepe, F., Verschaffel, L. & Star, J. (2020). Expertise in developing students' expertise in mathematics: Bridging teachers' professional knowledge and instructional quality. *ZDM Mathematics Education*, 52, 179-192.
- Double, K.S.; Mc Grane, J.A. & Hopfenbeck, T.N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32, 481-509.
- Earl, L. (2003). Assessment of Learning, for Learning, and as Learning (Chapter 3). In *Assessment as learning: Using classroom assessment to maximise student learning*. Thousand Oaks, CA: Corwin Press.
- Fagnant, A. (2021). Syllabus de didactique générale. Document non publié : Université de Liège.

- Fagnant, A., Dupont, V., & Demonty, I. (2016). Régulation interactive et résolution de tâches complexes en mathématiques. In L. Mottier Lopez & W. Tessaro (Eds.), *Le jugement professionnel, au coeur de l'évaluation et de la régulation des apprentissages* (pp. 229-251). Peter Lang.
- Fagnant, A., & Goffin, C. (2021). Note de synthèse – Comment se conjuguent « évaluation » et « développement professionnel des enseignants » dans les écrits scientifiques de l'Admee. *Évaluer. Journal international de recherche en éducation et formation*, 7(3), 5-26.
- Goffin, C., Baron, J., Ochelen, J.-P. & Fagnant, A. (2022). Freins et leviers pour œuvrer à la construction et à la communication d'une note à haute valeur informative : une étude exploratoire dans l'enseignement secondaire en Belgique francophone. *La Revue LEeE*, 6, 1-22.
- Gotch, C.M. & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33, 14-18
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New-York : Routledge.
- Hattie, J. (2013). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- 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.
- Hill, C.J., Bloom, H.S., Black, A.R. and Lipsey, M.W. (2008), Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2, 172-177.
- Khalid, N.H.M., Latif, A.A. & Yusof, I.J. (2021). Research focusing on assessment literacy: A systematic literature review. *Studies of applied economics*, 39(10), 1-25.
- Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37.
- Klute, M., Aphthorp, H., Harlacher, J., & Reale, M. (2017). Formative assessment and elementary school student academic achievement: A review of the evidence (REL 2017-259). Regional Educational Laboratory Central.
- Kruse, L., Impellizeri, W., Witherel, C. E., & Sondergeld, T. A. (2020). Evaluating the Impact of an Assessment Course on Preservice Teachers' Classroom Assessment Literacy and Self-Efficacy. *Mid-Western Educational Researcher*, 32(2), 107-132.
- Laveault, D. & Allal, L. (2016a). *Assessment for learning. Meeting the challenge of implementation*. Springer.
- Laveault, D. & Allal, L. (2016b). *Implementing Assessment for Learning: Theoretical and Practical Issues*. In D. Laveault & L., Allal (Eds.). *Assessment for learning. Meeting the challenge of implementation* (pp. 1-18). Springer.
- Leclercq, D. (2008) *A la recherche de la Triple concordance en Education. Illustration sur un cours de 1° Bac universitaire en grand groupe*. Document non publié. Université de Liège.

- 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.
- Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. DOI: 10.13140/RG.2.1.3478.4245
- Li, H., Xiong, Y., Hunter, C.V., Guo, X. & Tywoniw, R. (2020) Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193-211.
- Mercier-Brunel, Y. (2019). Les feedbacks des enseignants pour soutenir les processus d'autorégulation lors de corrections collectives. *Évaluer. Journal international de recherche en éducation et formation*, 5(2), 47-64.
- Merle, P. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives*. PUF.
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33 (1), 49-64.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113.
- Mertler, C. A., & Campbell, C. (2005, avril). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the "Assessment Literacy Inventory." Communication presented at the Annual Meeting of the American Educational Research Association, Montreal, Québec.
- Morrisette, J. (2013). Des modes d'interaction au coeur de la mise en œuvre d'une évaluation formative non instrumentée. *Nouveaux cahiers de la recherche en éducation*, 16(2), 88-111.
- Mottier Lopez, L. (2007). Régulations interactives situées dans des dynamiques de microculture de classe. *Mesure et évaluation en éducation*, 30(2), 23-47
- Mottier-Lopez, L. (2012). *La régulation des apprentissages en classe*. De Boeck.
- Mottier-Lopez, L. (2015). *Évaluation formative et certificative des apprentissages : enjeux pour l'enseignement*. De Boeck.
- Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation: un acte cognitif et une pratique sociale située. *Revue Suisse des Sciences de l'Éducation*, 30(3), 465-482.
- Pasquini, R. (2018). *Le modèle de l'alignement curriculaire élargi pour étudier les pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire: enjeux conceptuels et pragmatiques*. Thèse de doctorat : Université de Genève, no. FPSE 704 ;
- Pasquini, R. (2019). Élargir conceptuellement le modèle de l'alignement curriculaire pour comprendre la cohérence des pratiques évaluatives sommatives notées des enseignants : enjeux et perspectives. *Mesure et évaluation en éducation*, 42(1), 63-92.

- Rey, O. & Feyfant, A. (2014). Evaluer pour (mieux) faire apprendre. Dossier de veille de l'Institut Français de l'Éducation, 94, 1-44.
- Ruiz-Primo, M.A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37, 15–24.
- Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H. & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A Meta-Analysis. *Journal of Educational Psychology* 109(8), 1049–1066.
- Sansevy, B. (1995). Le contrat didactique. Note de synthèse. *Revue française de pédagogie*, 112,85-118.
- 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.
- Shulman, L. S. (1986). Those Who Understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Shulman, L. S. (2007). Ceux qui comprennent : Le développement de la connaissance dans l'enseignement. *Éducation et Didactique*, 1(1), 97-114.
- 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/>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.
- Wiliam, D. (2018). Assessment for learning: meeting the challenge of implementation. *Assessment in Education: Principles, Policy & Practice*, 25(6), 682-685,
- 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.



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](#)