

Le travail collectif des enseignants pour l'évaluation des apprentissages comme norme professionnelle ? Une revue de la littérature pour interroger cette tendance émergente

The collective work of teachers for the assessment of learning as a professional norm? A literature review to examine this emerging trend

O trabalho coletivo dos professores para a avaliação das aprendizagens como norma profissional? Uma revisão da literatura para questionar essa tendência emergente

Jean-Pascal Ochelen , Gonzague Yerly  and Lucie Mottier Lopez 

Volume 47, Number 1, 2024

URI: <https://id.erudit.org/iderudit/1114565ar>

DOI: <https://doi.org/10.7202/1114565ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Ochelen, J.-P., Yerly, G. & Mottier Lopez, L. (2024). Le travail collectif des enseignants pour l'évaluation des apprentissages comme norme professionnelle ? Une revue de la littérature pour interroger cette tendance émergente. *Mesure et évaluation en éducation*, 47(1), 27–65.
<https://doi.org/10.7202/1114565ar>

Article abstract

This paper presents a rapid review of the Anglophone and Francophone literature on collaboration among teachers to assess learning. This kind of collective work tends to become a norm in education systems, raising major issues in the field of educational assessment. The literature already contains summaries of empirical research dealing with teacher collaboration in general, but none dealing specifically with collective assessment practices. A systematic literature search led to the selection of 30 articles, in primary and secondary education. Analysis of these articles allows us to 1) identify the dynamics at work in the schools and 2) observe the consequences for all those involved in the school. The results allow us to discuss certain tensions linked to this trend, from the point of view of steering policies, but also about individual and collective evaluation practices.

© Jean-Pascal Ochelen, Gonzague Yerly et Lucie Mottier Lopez, 2024



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

**Le travail collectif des enseignants pour l'évaluation
des apprentissages comme norme professionnelle?
Une revue de la littérature pour interroger
cette tendance émergente**

**The collective work of teachers for the assessment
of learning as a professional norm? A literature review
to examine this emerging trend**

**O trabalho coletivo dos professores para a avaliação
das aprendizagens como norma profissional?
Uma revisão da literatura para questionar essa
tendência emergente**

Jean-Pascal Ochelen

ID ORCID : 0009-0004-9860-1142

HEP Fribourg & Université de Genève

Gonzague Yerly

ID ORCID : 0000-0002-5264-9720

HEP Fribourg

Lucie Mottier Lopez

ID ORCID : 0000-0002-9374-7300

Université de Genève



MOTS CLÉS : enseignants, évaluation des apprentissages, revue de littérature, travail collectif

L'article présente une rapid review de la littérature anglophone et francophone sur le travail collectif des enseignants pour l'évaluation des apprentissages. Ce genre de travail collectif tend à devenir une norme dans les systèmes éducatifs, soulevant des enjeux majeurs dans le champ de l'évaluation en éducation. La littérature compte déjà des synthèses et des recherches empiriques qui traitent de la collaboration des enseignants de manière générale, mais aucune ne traite spécifiquement des pratiques évaluatives collectives. Cette recherche documentaire dans la littérature conduit à la sélection de 30 articles, aux niveaux primaire et secondaire. L'analyse de ces articles permet 1) d'identifier les dynamiques mises en place dans les établissements et 2) d'observer les conséquences pour tous les acteurs de l'école. Les résultats permettent de discuter de certaines tensions liées à cette tendance, au point de vue des politiques de pilotage mais également sur le plan des pratiques évaluatives individuelles et collectives.

KEY WORDS: assessment, collaboration, review of literature, teachers

This paper presents a rapid review of the Anglophone and Francophone literature on collaboration among teachers to assess learning. This kind of collective work tends to become a norm in education systems, raising major issues in the field of educational assessment. The literature already contains summaries of empirical research dealing with teacher collaboration in general, but none dealing specifically with collective assessment practices. A systematic literature search led to the selection of 30 articles, in primary and secondary education. Analysis of these articles allows us to 1) identify the dynamics at work in the schools and 2) observe the consequences for all those involved in the school. The results allow us to discuss certain tensions linked to this trend, from the point of view of steering policies, but also about individual and collective evaluation practices.

Note de l'auteur : Toute correspondance liée à cet article doit être adressée à : jean-pascal.ochelen@edufr.ch

Cet article contribue à la réalisation d'une thèse de doctorat menée dans le cadre d'un projet de recherche plus vaste financé par le Fonds National Suisse de la recherche scientifique (projet FNS n° 100019_207804).

PALAVRAS CHAVE: professores, avaliação das aprendizagens, revisão de literatura, trabalho coletivo

O artigo apresenta uma rapid review da literatura anglófona e francófona sobre o trabalho coletivo dos professores para a avaliação das aprendizagens. Este tipo de trabalho coletivo tende a tornar-se uma norma nos sistemas educativos, levantando questões importantes no campo da avaliação em educação. A literatura já inclui sínteses e investigações empíricas que abordam a colaboração entre professores de forma geral, mas nenhuma trata especificamente das práticas avaliativas coletivas. Esta investigação documental na literatura resulta na seleção de 30 artigos, nos níveis primário e secundário. A análise destes artigos permite 1) identificar as dinâmicas implementadas nas instituições e 2) observar as consequências para todos os atores da escola. Os resultados possibilitam discutir algumas tensões associadas a esta tendência, tanto do ponto de vista das políticas de gestão como no que respeita às práticas avaliativas individuais e coletivas.

Introduction

Alors que depuis une trentaine d'années le travail collectif est devenu une norme en éducation (Borges & Lessard, 2007; Vangrieken et al., 2015), le concept d'évaluation collective émerge dans la littérature anglosaxonne, d'abord avec Hargreaves (2007) qui explore des collaborations entre enseignants et étudiants anglais, puis, dans les écrits australiens abordant des dispositifs collaboratifs de modération sociale entre enseignants (Adie et al., 2012; Wyatt-Smith et al., 2010). Si recourir au travail collectif pour l'évaluation n'est pas nouveau (Adie et al., 2012; Black et al., 2011; Laveault & Yerly, 2017; Mottier Lopez et al., 2012), son institutionnalisation plus officielle l'est davantage. L'objectif de ces nouvelles politiques est d'améliorer la qualité des résultats d'évaluation en harmonisant les pratiques évaluatives des enseignants au sein de leur établissement (Laveault & Yerly, 2017).

Depuis plus de 30 ans, de nombreux résultats de recherche portent sur le travail collectif des enseignants (Hargreaves, 1994, 2021; Vangrieken et al., 2015). Ces recherches indiquent que nous avons une bonne connaissance des effets du travail collectif de manière générale. Cependant, Yerly (2021) signale que c'est moins le cas lorsque ce type de travail se focalise sur les pratiques évaluatives. Le sujet de cet article est dès lors le suivant : quelles sont les connaissances scientifiques actuelles à propos du travail

collectif des enseignants quand celui-ci porte sur l'évaluation des apprentissages des élèves? Quelles sont les conclusions des recherches empiriques de ces 20 dernières années à ce propos? Pour répondre à ces deux questions, nous effectuons une revue de littérature de type *rapid review* qui conjugue certains avantages des revues systématiques tout en permettant d'accélérer le processus de sélection des articles (Grant & Booth, 2009, cités dans Dachet, 2024). La recherche est menée dans la littérature anglophone et francophone en éducation. Cette étude est inédite et vise à éclaircir un point aveugle de la littérature scientifique et des travaux en évaluation.

Contexte et cadre conceptuel

Le travail collectif des enseignants, une norme professionnelle contemporaine

Depuis les années 1980 et la volonté des politiques de décentraliser l'école, nous constatons de plus en plus l'émergence de l'établissement scolaire en tant qu'entité à part entière et comme une composante importante du système éducatif (Maroy, 2005). Selon Meirieu (2023), cette nouvelle centration sur les établissements scolaires s'est assortie d'une autonomie accrue de ceux-ci. Les écoles ont alors plus de libertés pour définir leurs projets d'établissements et leurs objectifs institutionnels. Dans cette configuration, ce sont généralement les résultats (*outputs*), et surtout les résultats des élèves à des épreuves standardisées, qui servent d'outils de pilotage du système.

Aujourd'hui, que ce soit à l'échelle internationale (OCDE, Communauté européenne, UNESCO) ou nationale (circulaires, décrets), l'appel au travail collectif des enseignants apparaît unanime (Gilbert, 2018). Les enseignants doivent désormais être de bons collaborateurs. Si le travail d'équipe prend de plus en plus de place dans la société en général, le travail collectif des enseignants devient également une norme en éducation (Decuyper et al., 2010). Cependant, les recherches visant à identifier et à comprendre la manière dont les enseignants traduisent les injonctions au travail collectif montrent une grande variabilité sur le plan de l'application dans les écoles. Spillane et al. (2016) expliquent que le degré de couplage entre les directives, les pratiques collectives et les pratiques enseignantes est plus important dans les pays ayant adopté des politiques de reddition

de comptes (*accountability*) fortes, par exemple, les États-Unis ou l'Angleterre (Maroy & Voisin, 2014). Quoi qu'il en soit, la norme liée au travail en équipe reste assez floue et les modèles nombreux (Gilbert, 2018). Les pratiques collectives peuvent prendre différentes formes.

Les modalités du travail collectif, de l'informel vers des communautés d'apprentissage

Dans leur ouvrage, Marcel et al. (2007) distinguent trois modalités de travail collectif selon l'intensité de partage du travail (coordination, collaboration, coopération). Pour Grangeat (2011), les modalités du travail collectif (collaboration, coopération distribuée, coaction) sont plutôt fonction des objectifs. Hargreaves (2021) distingue les dispositifs flexibles et démocratiques (cultures collaboratives) des dispositifs formels et fortement institutionnalisés (collégialité artificielle). La revue de la littérature de Vangrieken et al. (2015) synthétise les modalités documentées dans la littérature anglophone et clarifie les terminologies.

Pour Barrère et al. (2008), l'informel (niveau minimal de collaboration) est continuellement utilisé, même inconsciemment. Des échanges autour de la machine à café ou lors des repas ne sont pas anodins, par exemple. Ce faisant, les enseignants créent des liens, même si ceux-ci ne servent pas directement leur travail (Lator, 2010). Ces liens favorisent l'appartenance collective et sont les prémisses à une future coopération professionnelle.

Devoir réaliser un travail en équipe représente un niveau plus formel. Pour Cohen et Bailey (1997), dans ce type de travail, il y a interdépendance quant à la tâche. Les enseignants partagent la responsabilité des résultats. Salas et al. (2000) ajoutent que les équipes se caractérisent par une structure qui leur est propre et par une durée déterminée. Parfois, les équipes d'enseignants formalisent davantage le travail collectif. La littérature développe alors différents modèles pour appréhender des collectifs d'enseignants, tout à la fois dynamiques et structurés. Nous en citons brièvement trois ci-dessous : la communauté d'apprentissage professionnelle (CAP), la communauté d'apprentissage enseignante (CAE) et la communauté de pratique (CDP). Selon Gilbert (2018), la CAP désigne « un groupe de personnes qui partagent un processus d'apprentissage collaboratif dans le cadre d'une pratique réflexive pour un but commun circonscrit dans le temps » (p. 5). Le niveau d'échange et les apprentissages s'effectuent sur un plan horizontal. Les membres partagent un ensemble de normes, de valeurs et d'orientations en matière de pratiques enseignantes (Vangrieken

et al., 2015). Pour Wenger (2005), une CDP se construit au sein de groupes de personnes travaillant ensemble pour solutionner des problèmes rencontrés dans leurs pratiques professionnelles. Ce sont des praticiens qui développent un répertoire commun de ressources dans une pratique partagée. Quand ils créent une communauté de pratique, les enseignants partagent des idées et mettent les mêmes mots sur des phénomènes qu'ils jugeaient ou identifiaient alors différemment. Ils s'engagent mutuellement face à l'entreprise commune qui les rassemble. Dans une CAE, les enseignants travaillent collectivement dans un domaine de contenu unique, pour réfléchir sur leurs pratiques, pour examiner les résultats des élèves et pour apporter des changements visant à améliorer la qualité de l'enseignement et de l'apprentissage au sein de leur école (Mc Laughlin & Talbert, 2006). Dans ces différentes communautés, l'apprentissage ne se limite pas au seul individu. Influencés par l'environnement, les apprentissages sont vus comme situés en ce sens qu'ils sont ancrés dans une pratique sociale et culturelle (Mottier Lopez, 2008, 2021). Les interactions et les relations de collaboration sont considérées comme cruciales pour le développement des connaissances individuelles et collectives.

Le rôle de la négociation

Pour Wenger (2005), la pratique au sein des communautés de pratique est sans cesse négociée, notamment pour la construction d'un répertoire partagé. De nombreux auteurs se sont intéressés au rôle de la négociation dans les dispositifs de travail collaboratif. Pour Strauss (1992), il existe deux types de négociation : la négociation conflictuelle et la négociation intégrative. La première implique une confrontation de points de vue et des tactiques agressives, tandis que la seconde est caractérisée par la coopération et une intention de parvenir à un accord mutuellement bénéfique. Beuret (2010) explique que les parties en conflit peuvent passer d'un type de négociation à l'autre en trouvant un « point de passage transactionnel » (p. 45), tel qu'une médiation. Allenbach (2016) précise que les processus de construction d'accords par la négociation de sens et d'alliances sont des facteurs de réussite cruciaux pour le travail collectif. Dans leur étude empirique, Germier et Marcel (2016) montrent également que la négociation est importante pour construire des accords dans un collectif d'enseignants. De ces quelques lignes, nous retenons que la négociation joue un rôle important dans la structuration du travail collectif et dans son fonctionnement.

L'état des connaissances sur le travail collectif des enseignants

Les résultats de la revue systématique de la littérature anglophone de Vangrieken et al. (2015) (82 articles empiriques analysés à partir de 2000) sont les plus récents que nous ayons trouvés. Cette revue montre que le travail collectif s'organise de différentes manières et qu'il existe une terminologie qui prête parfois à confusion. Des termes comme collaboration, coopération, coordination, travail partagé, travail en groupe, travail en équipe, etc. sont utilisés de manière interchangeable. L'intensité de la collaboration est très variable, y compris au sein d'une même modalité de travail. La revue souligne les avantages du travail collectif au niveau des élèves, des écoles mais surtout pour les enseignants (par ex., gain de temps, développement professionnel, efficacité, engagement). Certaines limites et certains désavantages sont également mis en évidence (par ex., tensions, dégradation des relations interpersonnelles, paresse sociale, perte d'autonomie). Les travaux de Hattie (2017) sur le concept d'efficacité collective des enseignants montrent que l'effet du travail collectif des enseignants sur les apprentissages des élèves ($d = 1,57$) est deux fois plus important que celui de la rétroaction ($d = 0,70$) et presque trois fois plus important que l'effet lié à la gestion de classe ($d = 0,62$). Le travail collectif des enseignants aurait donc un impact fort sur les apprentissages des élèves.

L'évaluation collective : une pratique sociale située

Par rapport aux tendances émergentes dans la littérature scientifique, nous proposons dans ce texte d'appréhender l'évaluation comme une pratique collective dans une perspective située (Lave, 1988 ; Wenger, 2005). Nous focalisons sur les enjeux du jugement professionnel des enseignants qui concernent tous les gestes et toutes les opérations constitutives de l'évaluation, quelle que soit sa fonction (Allal & Mottier Lopez, 2009 ; Wyatt-Smith et al., 2010). Pour Mottier Lopez et Allal (2008), le jugement professionnel dans une perspective située est à la fois un acte cognitif et une pratique sociale fondée sur des normes, des outils, des langages ainsi que des cultures et qui a ses contextes sociaux propres. Il permet d'appréhender le caractère complexe de l'évaluation des apprentissages des élèves. Nous conceptualisons ce jugement professionnel en évaluation dans un contexte collectif de travail entre enseignants, en exploitant les trois conditions principales développées par Wenger pour définir une communauté de pratique : un travail conjoint offrant les conditions à tous les acteurs concernés de s'y engager mutuellement et de façon active (engagement

mutuel), la poursuite et l'atteinte de buts communs face à un projet commun et partagé (entreprise commune), la construction et la négociation de sens, d'outils, d'actions, de façons de faire efficaces et responsables pour parvenir à une évaluation juste et précise (répertoire partagé).

Les questions de recherche et les objectifs spécifiques

Étant donné l'agenda des politiques éducatives internationales et les orientations pédagogiques encourageant de plus en plus les pratiques évaluatives collectives, notre recherche documentaire vise à dresser un état des lieux des connaissances scientifiques en matière de pratiques évaluatives collectives chez les enseignants. Compte tenu des potentiels enjeux aux niveaux micro, méso et macro, leur mise en œuvre et leurs effets pourraient entraîner des conséquences importantes pour tous les acteurs¹. Dans la mesure où l'évaluation des apprentissages est de moins en moins une pratique individuelle (Yerly, 2021), notre recherche vise à répondre aux quatre questions générales suivantes :

- 1) Quelles sont les finalités du travail collectif des enseignants pour l'évaluation des apprentissages ?
- 2) Quels outils et quelles méthodes utilisent-ils lors de ce travail collectif ?
- 3) Quelles sont les dimensions de l'évaluation concernées par le travail collectif ?
- 4) Quels sont les effets pour les élèves et les enseignants, pour les établissements et pour le système éducatif ?

Pour répondre aux questions, les quatre catégories d'analyse du contenu des articles consistent à :

- a) identifier les objectifs des dispositifs de travail collectif des enseignants employés pour évaluer les apprentissages des élèves ;
- b) inventorier les outils et les méthodes utilisés par les enseignants quand ils travaillent ensemble pour évaluer les apprentissages ;
- c) identifier les objets d'attention conjointe quand les enseignants évaluent les apprentissages collectivement ;

1. Les niveaux micro, méso et macro font référence à la catégorisation proposée par De Ketele (2006). L'auteur situe le niveau micro au sein des processus d'enseignement/apprentissage dans un groupe-classe ou de formation (ici, les élèves et les enseignants). Le niveau méso se situe au niveau d'une école, d'un établissement ou d'un institut de formation (ici, l'établissement scolaire). Le niveau macro est associé au plan d'un système éducatif ou de formation (ici, le système éducatif).

- d) mettre en avant les effets du travail collectif autour de l'évaluation des apprentissages pour les élèves et les enseignants, pour les établissements et pour le système scolaire.

Méthodologie

La démarche de la recherche

Pour assurer une recherche rigoureuse qui tienne compte de nos contraintes temporelles, nous avons opté pour une *rapid review* de la littérature (Grant & Booth, 2009, cités dans Dachet, 2024). Selon Ganann et al. (2010, cités dans Dachet, 2024), cette méthode conjugue un ensemble des méthodes utilisées pour accélérer et optimiser le processus de recherche systématique de la littérature dans un double objectif : gain de temps et gain en ressources humaines. Notre méthodologie utilise certaines démarches propres aux revues systématiques : questions de recherche précises, utilisation d'un protocole de recherche reproductible, critères clairement définis, objectifs explicites. Toutefois, par rapport à une recherche systématique, nous opérons certaines concessions méthodologiques : absence de processus de décision de maintien ou de rejet des articles par deux juges indépendants, réduction du nombre de bases de données, non-recours aux analyses statistiques. De ce fait, plutôt que de synthétiser les résultats en utilisant des méthodes statistiques ou quantitatives, nous effectuons une analyse et une présentation narrative des résultats.

La procédure de recherche documentaire

Des articles scientifiques issus de la littérature anglophone et francophone ont été récoltés en mai et juin 2023. Pour la littérature anglophone, nous avons d'abord interrogé la base de données ERIC qui est spécifique aux sciences de l'éducation. Pour plus d'exhaustivité, nous avons également utilisé APA PsycInfo qui a permis de croiser avec des domaines connexes (sciences sociales et humaines, psychologie). Pour la littérature francophone, nous avons utilisé les moteurs de recherche HAL Open Science, ERUDIT et CAIRN, connus pour leur sélection rigoureuse et qualitative de publications académiques. Pour commencer nos recherches dans les bases de données, nous avons mobilisé quelques écrits fondateurs correspondant à la thématique et à nos critères de recherche (Adie, 2014; Allal & Mottier Lopez, 2014; Heredia et al., 2016; Mottier Lopez & Pasquini, 2017). Ces écrits ont permis de déterminer les mots clés servant de base pour interroger les thésaurus.

Pour ERIC et PsycInfo, les équations de recherche sont construites en utilisant les *index terms* issus de ces thésaurus. Nous avons également introduit d'autres mots clés issus du vocabulaire libre (des termes non normalisés pour décrire les contenus des articles). Le tableau 1 reprend les *index terms* et le vocabulaire libre introduits dans les thésaurus. Ensuite, selon les langages, les codes et les normes propres aux bases de données, tous les termes ont été mobilisés pour effectuer des recherches dans les titres, les résumés et l'entièreté des textes. Pour les moteurs de recherche en français, étant donné l'absence de thésaurus, nous avons introduit les mots clés (traduits) issus du vocabulaire libre.

Tableau 1
Vocabulaire libre et index terms de PsycInfo et ERIC

	Concept 1 : évaluation des apprentissages	Concept 2 : travail collectif entre enseignants
Vocabulaire libre	<i>assessment, evaluation, grading, social moderation, teachers' professional development</i>	<i>community of practice, cooperation shared work, teacher collaboration</i>
<i>Index terms</i> du thésaurus de PsycInfo	<i>educational measurement, evaluation criteria, formative assessment, grading (educational)</i>	<i>collaboration, cooperation, teamwork</i>
<i>Index terms</i> du thésaurus de ERIC	<i>accountability, educational assessment, educational policy, educational practices, formative assessment, grading, interrater reliability, professional development, student evaluation, summative assessment</i>	<i>collaboration, collegiality, cooperation, cooperative plannings, peer relationship, teacher attitudes, teacher collaboration, teamwork,</i>

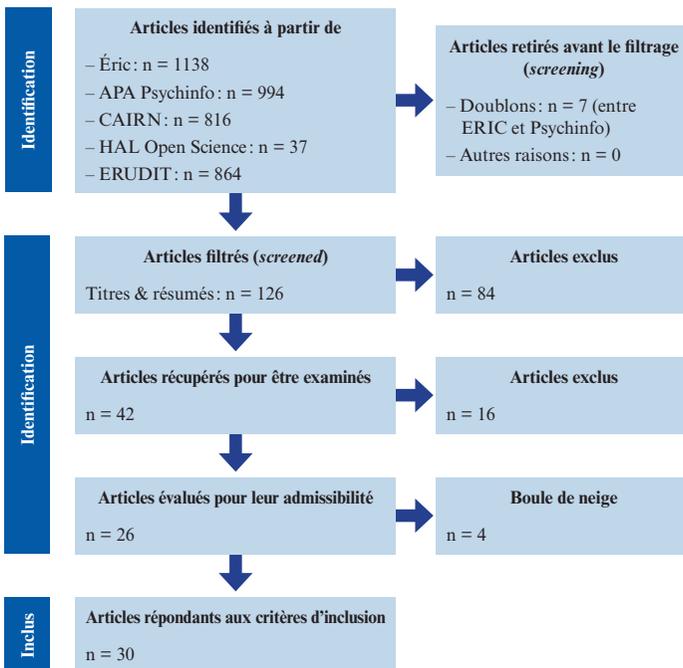
Pour assurer une recherche de qualité, des critères d'inclusion ont guidé notre recherche. Ainsi, les articles sélectionnés doivent impérativement :

- être écrits en anglais ou en français ;
- être issus de la littérature scientifique avec comité de lecture ;
- avoir été publiés à partir de 2003 ;
- être empiriques ;
- porter sur une forme de travail collectif entre enseignants, uniquement pour l'évaluation des apprentissages ;
- concerner les niveaux scolaires compris entre les grades 1 et 12.

Bien qu'intéressante, la littérature traitant de l'enseignement spécialisé a été exclue, car elle implique d'autres dynamiques et des intervenants parascolaires externes qui ne font pas l'objet de notre recherche. Nous avons activé les différents filtres (correspondant aux critères) uniformément dans toutes les plateformes de recherche. Parmi les articles proposés, nous avons fait une première sélection à partir des titres des articles et avons éliminé ceux qui n'étaient pas pertinents pour notre recherche ou qui ne correspondaient pas à nos critères de sélection. La deuxième sélection s'est opérée de la même manière sur la base des résumés. La dernière étape était la lecture du texte intégral. Enfin, grâce à la technique de la boule de neige (Vangrieken et al., 2015), l'analyse des références bibliographiques des articles sélectionnés a mené à identifier d'autres sources non apparues dans les bases de données consultées. Pour la boule de neige, nous avons utilisé les mêmes critères. Cette recherche manuelle s'est alors effectuée avec les moteurs de recherche Google Scholar et Taylor and Francis. La figure 1 ci-dessous présente le *flowchart*, inspiré du modèle PRISMA (Page et al., 2021) qui illustre le processus de sélection et le nombre d'articles sélectionnés.

Figure 1

Flowchart selon le modèle PRISMA, explicitant le processus d'extraction d'articles



Le traitement des données – une synthèse de la recherche documentaire

Afin d'analyser leur contenu, nous avons lu, catégorisé et organisé les articles sélectionnés dans une table de lecture telle que proposée par Dumez (2011, adaptée de Hart, 2009). La table contient les dates de publication, les auteurs, les titres, les résumés, les pays d'origine, les types de recherche, les méthodologies, les niveaux scolaires. Elle s'est ensuite enrichie de quatre catégories supplémentaires en lien avec nos questions de recherche et avec nos catégories d'analyse : les objectifs du travail collectif, les objets évaluatifs touchés, les outils et les méthodes utilisés et, enfin, les effets. Au fur et à mesure des lectures, nous avons élaboré un tableau de synthèse, permettant de réaliser un inventaire des caractéristiques générales des études sélectionnées. Cette façon de procéder a permis de faciliter le tri, le regroupement et une « relecture ouverte, flottante, du matériau théorique rassemblé » (Dumez, 2011, p. 24).

Les premiers constats généraux de notre recherche documentaire

Notre démarche de recherche documentaire a conduit à la sélection de 30 articles empiriques. Les dates de publication sont comprises entre 2007 et 2022. Le tableau 2 présente le corpus d'articles sélectionnés². Parmi ceux-ci, il y a 27 études qualitatives, deux études quantitatives et une étude à méthode mixte. Pour les études qualitatives, les méthodes de recherche ciblent toutes des cas spécifiques. Des entretiens individuels semi-directifs, des *focus group*, des observations d'équipes et des analyses de documents sont les outils méthodologiques privilégiés. Peu d'informations sont disponibles sur les outils d'analyse des données qualitatives. Seules deux études précisent l'utilisation du logiciel NVivo³. Pour les recherches quantitatives, les logiciels d'analyse ne sont pas nommés. Les outils statistiques les plus souvent utilisés sont les calculs des fréquences, les corrélations, les moyennes, l'écart-type et les régressions linéaires.

Il est intéressant d'observer que les pays dont sont issus les 30 articles sont des États fédéraux où la souveraineté en éducation revient aux juridictions locales (Australie, Nouvelle-Zélande, États-Unis, Royaume-Uni et Suisse) (n = 26) ou des États unitaires ayant mis en place des réformes qui

2. Une version complète de ce tableau est présentée en annexe A.

3. Smail (2020) et McFadden et al. (2022).

ont accru la décentralisation de leurs systèmes éducatifs (Norvège, Suède, Pays-Bas) (n = 4). La grande majorité des articles sont écrits en anglais (n = 27), les autres sont en français (n = 3).

Dans notre sélection d'articles, le travail collectif est davantage organisé de manière horizontale (intra-niveau – entre enseignants travaillant dans un même niveau) dans l'enseignement secondaire (n = 19) ou primaire (n = 5) que verticale (inter-niveaux – entre enseignants du primaire et du secondaire) (n = 5). Dans une configuration horizontale, la négociation est plus simple, car elle est plus homogène. La verticalité engendre en effet la rencontre de deux mondes, ce qui nécessite une démarche prolongée et complexe. Ces résultats rejoignent les conclusions de Doppenberg et al. (2012) qui expliquent que les enseignants préfèrent travailler avec des collègues du même niveau.

Dans les 19 études réalisées dans le niveau secondaire, le travail collectif est toujours organisé entre enseignants d'une même discipline (ou du même domaine). Ces résultats ne s'accordent pas avec ceux de la revue de littérature de Vangrieken et al. (2015) qui signalent que, dans le secondaire, la collaboration s'effectue principalement entre enseignants issus de disciplines différentes. Dans la plupart de nos études retenues, le profil des enseignants est hétérogène (genre, expérience, discipline). Cependant, sur le plan de l'ancienneté de service, deux recherches (Allal & Mottier Lopez, 2014; Matre & Solheim, 2016) ne concernent que des enseignants expérimentés.

Résultats

Dans la suite du texte, nous présentons les résultats de notre analyse des contenus des 30 articles. Ils sont présentés en suivant la structure des catégories d'analyse.

L'identification des objectifs du travail collectif pour l'évaluation des apprentissages

Cette partie permet de répondre à la question : Dans quel but les enseignants collaborent-ils pour évaluer les apprentissages de leurs élèves ? Le tableau 2 présente les objectifs tels que les décrivent les 30 articles. Nous ajoutons les auteurs, les dates de publication, les pays et, le cas échéant, les autorités locales responsables de l'enseignement.

Tableau 2
*Corpus complet des articles, objectifs des études, auteurs,
 années de publication et pays d'origine*

Objectifs du travail collectif	Auteurs (années de publication)	Pays (juridictions locales)
Modération sociale (n = 20)	Adie (2013)	Australie (Queensland)
	Adie (2014)	Australie (Queensland)
	Adie et al. (2012)	Australie (Queensland)
	Allal & Mottier Lopez (2014)	Suisse (Genève)
	Black et al. (2011)	Grande-Bretagne
	Colbert et al. (2012)	Australie (Queensland)
	Connolly et al. (2012)	Australie (Queensland)
	Crisp (2017)	Grande-Bretagne
	Grant (2012)	Grande-Bretagne (Écosse)
	Hipkins & Robertson (2012)	Nouvelle-Zélande
	Klenowski et al (2007)	Australie (Queensland)
	Matre & Solheim (2016)	Norvège
	Morales Villabona & Mottier Lopez (2016)	Suisse (Genève)
	Mottier Lopez & Pasquini (2017)	Suisse (Genève et Vaud)
	Mottier Lopez et al. (2012)	Suisse (Genève)
	Nilberth & Sandlund (2020)	Suède
	Reid (2007)	Grande-Bretagne (Écosse)
	Smail (2020)	Nouvelle-Zélande
	Van Der Schaaf et al. (2012)	Pays-Bas
	Wyatt Smith et al. (2010)	Australie (Queensland)
Développement des pratiques (Aft) (n = 4)	Furtak et al. (2013)	États-Unis
	Hargreaves (2013)	Grande-Bretagne
	Hargreaves et al. (2012)	Grande-Bretagne et Hong Kong
	Stewart & Houchens (2012)	États-Unis (Kentucky)
Construction d'outils (n = 5)	Buckley & Liscompe (2020)	Australie (Nouvelle-Galles du Sud)
	Heredia et al. (2016)	États-Unis
	McFadden et al. (2022)	États-Unis (Kentucky)
	Mottier Lopez & Pasquini (2017)	Suisse (Vaud)
	Morales Villabona & Mottier Lopez (2016)	Suisse (Genève)
Mise en œuvre de recommandations (n = 3)	Hermansen (2014)	Norvège
	Hermansen & Nerland (2014)	Norvège
	Yerly (2021)	Suisse (Fribourg)

L'objectif de modération sociale des résultats de l'évaluation est le plus présent (n = 20). Selon Gipps (1994), la modération sociale suggère que des enseignants se regroupent pour discuter et comparer les notes attribuées aux élèves, afin de trouver un consensus et une compréhension commune du niveau de qualité du travail. Elle engage des confrontations d'avis, d'interprétations et de jugements humains à propos de travaux d'élèves (Linn, 1993). Dans ces 20 articles, la modération sociale est parfois utilisée à des fins de reddition de comptes (*accountability*) (n = 6). Le travail collectif des enseignants est alors utilisé pour améliorer la qualité des évaluations (n = 3) ou pour vérifier la manière avec laquelle les réformes sont mises en place sur le terrain (n = 3). La modération sociale est également mobilisée à des fins de développement professionnel (*improvement*) (n = 14). Elle vise alors à améliorer la qualité du jugement professionnel (n = 6), à développer les compétences en évaluation sommative (n = 1) ou formative (n = 3), à étudier les dynamiques de négociation lors de l'attribution de notes (n = 1) ou à améliorer les apprentissages des élèves (n = 1). Elle est également un moyen de créer des communautés de pratique pour favoriser une collaboration verticale (n = 2). Dans ce cas, le travail collectif est censé faciliter les échanges entre les acteurs de l'enseignement primaire et secondaire. Dans certaines études, si on demande aux enseignants de collaborer pour évaluer les apprentissages, c'est aussi pour qu'ils construisent collectivement des connaissances en matière d'évaluation soutien d'apprentissage (*assessment for learning – Afl*) (n = 4). L'objectif du travail collectif est également de favoriser les échanges entre enseignants afin de concevoir collectivement des épreuves dans une visée formative (n = 3) ou sommative (n = 2). Enfin, si les enseignants se regroupent pour l'évaluation des apprentissages, c'est également pour mettre en œuvre les recommandations que les autorités locales (n = 2) ou fédérales (n = 1) ont édictées en matière de pratiques évaluatives.

Les outils et les méthodes soutenant le travail collectif pour l'évaluation des apprentissages

Pour atteindre les objectifs cités dans la section précédente, la littérature documente un large éventail d'outils et de méthodes. Dans cette partie, nous répondons à la question : Quels outils et méthodes utilisent-ils pour travailler ensemble sur l'évaluation des apprentissages ?

Quand il s'agit de concevoir collectivement des épreuves, les enseignants utilisent des guides de conception communs, indiquant les directives à mettre en œuvre (n = 10). Ces guides sont soit conçus par les

autorités (n = 7), soit construits collectivement par des enseignants en amont (n = 3). Parallèlement, ils s'appuient aussi sur des ressources personnelles et utilisent leurs propres évaluations en guise de modèles (n = 3). Les programmes d'études sont également mobilisés lors de la conception commune d'épreuves (n = 4). Mottier Lopez et al. (2012) constatent d'ailleurs que la première préoccupation des équipes observées dans leur étude consiste à assurer un alignement curriculaire entre l'évaluation, les pratiques d'enseignement et les objectifs d'apprentissage. Pour construire des outils d'évaluation communs, les enseignants se réfèrent également à des exemples d'épreuves externes antérieures (n = 3). Quand la collaboration est suscitée par des chercheurs (n = 3), la construction d'outils évaluatifs est aussi guidée par un scénario (n = 2), ou confrontée à un modèle théorique (n = 1). Parfois, des outils technologiques sont utilisés. Il peut s'agir de plateformes de communication qui, avec la fonction audio, permettent de mettre des personnes très éloignées géographiquement en relation (n = 3). En Australie, des modérations en ligne sont organisées. Des outils technologiques sont également utilisés pour des séances d'autoscopie en présentiel (n = 1). Les participants d'une communauté enseignante ont ainsi pu visionner ensemble la manière avec laquelle ils mettent en œuvre en classe des pratiques d'*Aff* construites collectivement auparavant. Pour comparer la qualité de l'interprétation de travaux d'élèves, les échantillons d'épreuves déjà corrigées sont les outils les plus présents dans notre corpus d'articles (n = 22). Ces épreuves sont des tests soit formatifs (n = 7), soit sommatifs avec ou sans visée certificative (n = 15). Des cadres de référence pour l'évaluation sont d'autres outils souvent utilisés dans les séances collectives (n = 17). Ces documents contiennent les normes, les critères et les guides de correction à appliquer. Ces outils servent de référence pour la construction de consensus sur les travaux des élèves.

Dans les 30 articles, les méthodes avec lesquelles les enseignants travaillent ensemble varient. Dans la plupart des études, des réunions régulières sont planifiées pour permettre aux enseignants de travailler collectivement (n = 19). Ces réunions se déroulent généralement en personne (n = 16) mais parfois aussi à distance (n = 3). Elles sont prévues dans le calendrier des enseignants et font partie intégrante de procédures à appliquer, spécifiques à l'évaluation et imposées par la direction. Le nombre et la fréquence des séances dépendent des objectifs pour lesquels ils collaborent. Par exemple, les études australiennes (n = 8) s'inscrivent toutes dans un contexte d'importantes réformes des programmes scolaires et de

l'évaluation, notamment via l'établissement et l'application de standards (Klenowski et al., 2007). Les procédures de modération sociale y sont standardisées et respectent un cahier des charges édicté par les autorités nationales. La fréquence et le nombre des réunions est fonction de l'agenda des sessions d'évaluation. Quand les procédures ne sont pas standardisées, la régularité des séances collectives dépend du fonctionnement propre aux équipes ($n = 2$). Dans certains cas, c'est par l'intermédiaire de projets gouvernementaux ($n = 4$) ou à l'initiative de fonds privés que les enseignants sont invités à collaborer autour de l'évaluation ($n = 2$). Les subventions et les investissements octroyés engendrent alors des injonctions plus ou moins douces dont les objectifs sont clairement exprimés.

Dans plusieurs recherches, les auteurs précisent que le moyen de rassembler les enseignants passe par la participation à un réseau professionnel ($n = 12$). S'appuyant sur les théories socioculturelles de l'apprentissage, plusieurs études ($n = 6$) utilisent le concept de communauté de pratique de Wenger (1998, 2005). Des communautés d'apprentissage professionnelles ($n = 1$) ou des communautés d'apprentissage enseignantes ($n = 5$) sont également utilisées. Dans ces dernières, le travail collectif est parfois organisé en ateliers ($n = 2$) ou suit la chronologie d'un mode opératoire défini ($n = 3$).

Parfois, le travail collectif pour l'évaluation des apprentissages répond à une demande différente. Il est organisé pour une expérience de recherche. Les séances sont alors organisées de manière ponctuelle ($n = 6$). Elles se déroulent sur une ou deux journées dans le cadre de séminaires de recherche ($n = 3$), d'une formation continue ($n = 2$) ou pour une recherche exploratoire ($n = 1$).

Les objets d'attention conjointe de l'évaluation collective

Dans cette partie, nous répondons à la question : Sur quelles dimensions de l'évaluation travaillent-ils collectivement ? Nous mettons en évidence les objets de l'évaluation – les dimensions – qui sont influencés par le travail collectif des enseignants. Cette section est organisée en deux parties : d'abord, lors de la conception de l'épreuve, puis, lors de la correction.

Le travail collectif pour concevoir l'épreuve

Dans 13 études, le travail collectif se concentre sur l'étape de conception de l'épreuve. Ce sont les objets du cadre de référence de l'évaluation qui sont touchés (les référents : ensemble de lignes directrices qui

définissent comment elle doit être conçue, réalisée et interprétée). Ce sont, par exemple, les objectifs à évaluer, les critères, les seuils de réussite, la pondération, les barèmes, les normes ou les activités évaluatives qui sont discutés collectivement. Dans ces discussions, le leadership est généralement partagé entre les enseignants (n = 9) mais parfois, il revient à un membre des équipes de direction (directeur, adjoint, doyen) (n = 2) ou à des intervenants extérieurs (chercheurs) (n = 2). Plusieurs auteurs expliquent que les enseignants éprouvent des difficultés pour obtenir une compréhension des référents de l'évaluation (par ex., objectifs, critères) (n = 3). Dans une étude, les consensus et la compréhension commune des attentes ont émergé grâce aux efforts visant à rendre les discours transparents. Plusieurs auteurs constatent le rôle constructif des controverses professionnelles pour concevoir collectivement des épreuves (n = 2). Dans quelques recherches, les discussions ne se concentrent que sur la compréhension commune et sur l'application des critères (n = 3). Parfois, les cadres de référence font partie de procédures à appliquer pour l'évaluation. La collaboration entraîne alors des ajustements au sein de ceux-ci. Dans Colbert et al. (2012), ces procédures sont constituées de quatre points de contrôle à remplir pour garantir l'assurance qualité de l'évaluation. Ces points de contrôle font l'objet d'une évaluation collective par les enseignants et sont ajustés en fonction des résultats. Dans l'étude norvégienne de Hermansen et Nerland (2014), les procédures concernent les modalités de mise en place d'une évaluation soutien d'apprentissage. Grâce au travail collectif, les enseignants ont retravaillé les principes historiquement établis. Ils ont négocié et recontextualisé les bases pour développer de nouvelles pratiques partagées d'*Afl*.

Le travail collectif pour corriger l'épreuve

L'analyse du corpus d'articles fait ressortir que, dans une majorité de recherches (n = 17), la collaboration prend place au moment de la correction des travaux des élèves ou juste après. La notation est alors l'objet principalement touché. Dans plusieurs articles, c'est la note finale attribuée à des évaluations internes qui est discutée et, le cas échéant, revue (n = 9). Dans d'autres articles, les discussions collectives tournent autour de la notation attribuée à des évaluations externes standardisées et certificatives (n = 3). Parfois, la collaboration se concentre sur les contenus d'activités réalisées par les élèves (n = 5). L'interprétation des preuves d'apprentissage est discutée et comparée dans des portfolios (n = 3) ou directement dans les tâches au sein d'épreuves internes (n = 2). Dans plusieurs études,

lors de séances collectives, les enseignants signalent adopter une approche compréhensive (flexible) des critères ($n = 5$). Ils justifient ce choix en évoquant qu'une utilisation instrumentale serait injuste dans certains cas. Ils exposent les circonstances de leurs arrangements évaluatifs, notamment quand un élève est à la limite du seuil de réussite, pour des élèves à besoins particuliers, par des contextes de classe particuliers (par ex., à la suite d'une absence de l'enseignant). Les discussions à propos de l'utilisation des critères ont suscité de vifs débats. Certains enseignants, souvent les plus expérimentés, se montrant plus rigides ($n = 2$).

Les notes qui apparaissent dans le bulletin peuvent également être discutées collectivement ($n = 1$). Les enseignants ayant participé à cette étude déclarent que, juste avant d'y inscrire les notes, les discussions avec leurs collègues ont été très utiles. Ils expliquent que les informations échangées leur ont permis de porter un regard différent sur les élèves et d'interagir davantage avec eux. Cela leur a permis d'améliorer la cohérence de la communication dans les rubriques qualitatives du bulletin.

Les effets du travail collectif aux niveaux micro, méso et macro

Dans cette partie, nous répondons à la question : Quelles conséquences ce travail collectif provoque-t-il au niveau des élèves et des enseignants, des établissements et du système ? Pour proposer un panorama clair, nous présentons les résultats de l'analyse des 30 articles sous forme de deux tableaux à trois niveaux : le niveau micro (où nous distinguons deux sous-niveaux : les élèves et les enseignants), le niveau méso (les établissements) et le niveau macro (le système éducatif). Comme l'ont fait Vangrieken et al. (2015) dans leur revue systématique, nous choisissons de distinguer les effets positifs des effets négatifs, bien que nous soyons conscients du fait que cette dichotomie n'est pas toujours objective. En effet, une telle classification ne reflète pas totalement la complexité des effets. Malgré tout, ce choix vise à faciliter la lisibilité et la compréhension des différents effets, en les regroupant de manière claire et intuitive pour le lecteur.

Les effets positifs

Comme dit plus haut, la littérature montre de nombreux effets positifs du travail collectif des enseignants de manière globale. S'appuyant sur les 30 articles, le tableau 3 présente ces effets dans le contexte spécifique de l'évaluation des apprentissages.

Tableau 3

Effets positifs du travail collectif pour l'évaluation des apprentissages selon les niveaux élèves et enseignants, établissements et système

	Niveau micro	Niveau méso	Niveau macro
Niveau élèves	Niveau enseignants	Niveau établissements	Niveau système
– Implication des élèves dans le processus évaluatif (Buckley & Liscombe, 2020 ; Hermansen, 2014; Mottier Lopez & Pasquini, 2017; Smail, 2020)	– Transformation et amélioration du jugement professionnel (Adie et al., 2012; Allal & Mottier Lopez, 2014; Connolly et al., 2012; Furtak et al., 2016; Klenowski et al., 2007; Mottier Lopez et al., 2012; Nilsberth & Sandlund, 2020; Van Der Schaaf et al., 2012; Wyatt Smith et al., 2010) – Apprentissage mutuel (Reid, 2007; Stewart & Houchens, 2014) – Harmonisation des pratiques (Allal & Mottier Lopez, 2014; Yerly, 2021)	– Culture partagée de l'évaluation (Hargreaves, 2013; Mottier Lopez & Pasquini, 2017; Smail, 2020) – Meilleures relations interinstitutionnelles (Adie, 2014; Crisp, 2017; Hargreaves, 2013; Morales Villabona & Mottier Lopez, 2016; Reid, 2007) – Production de ressources pédagogiques (Black et al., 2011; Hipkins & Robertson, 2012; Matre & Solheim, 2016; Smail, 2020; Stewart & Houchens, 2014) – Mise en lien avec des opérateurs parascolaires (Allal & Mottier Lopez, 2014)	– Amélioration de la qualité de l'évaluation (validité, cohérence, fidélité) (Allal & Mottier Lopez, 2014; Black et al., 2011; Buckley & Liscombe, 2020; Connolly et al., 2012; Furtak et al., 2016; Hermansen, 2014; Hipkins & Robertson, 2012; Mottier Lopez & Pasquini, 2017; Smail, 2020; Van Der Schaaf et al., 2012; Wyatt Smith et al., 2010 Yerly, 2021)
– Soutien des progrès des élèves (Hargreaves, 2013)	– Amélioration de l'usage pédagogique de l'évaluation (Buckley & Liscombe, 2020; Hermansen, 2014; Hermansen & Nerland, 2014; McFadden et al., 2022; Mottier Lopez et Pasquini, 2017)		
– Communication des résultats (Mottier Lopez & Pasquini, 2017)	– Changement/transformation des pratiques évaluatives (Adie, 2014; Allal & Mottier Lopez, 2014; Black et al., 2011; Buckley & Liscombe, 2020; Grant, 2012; Hermansen, 2014; McFadden et al., 2022; Morales Villabona & Mottier Lopez, 2016; Mottier Lopez & Pasquini, 2017; Smail, 2020; Stewart & Houchens, 2014; Yerly, 2021)		
– Meilleurs résultats scolaires (Furtak et al., 2016; Stewart & Houchens, 2014)	– Émergence d'un débat professionnel (Mottier Lopez et al., 2012; Mottier Lopez & Pasquini, 2017; Reid, 2007; Smail, 2020)		

Tableau 3 (suite)

Effets positifs du travail collectif pour l'évaluation des apprentissages selon les niveaux élèves et enseignants, établissements et système

	Niveau micro	Niveau méso	Niveau macro
Niveau élèves	Niveau enseignants	Niveau établissements	Niveau système
	<ul style="list-style-type: none"> – Construction et soutien de nouvelles connaissances professionnelles (Crisp, 2017; Hermansen & Nerland, 2014; Mottier Lopez et al., 2012; Mottier Lopez & Pasquini, 2017; Reid, 2007; Stewart & Houchens, 2014) – Réflexion sur les pratiques d'évaluation, métacognition (Hargreaves, 2013; Heredia et al., 2016; McFadden et al., 2022) – Renforcement des liens (Adie, 2013; Hargreaves, 2013) – Construction d'une compréhension commune des standards (Adie, 2013, 2014; Buckley & Liscompe, 2020; Connolly et al., 2012; Crisp, 2017; Hermansen, 2014; Hipkins & Robertson, 2012; Klenowski et al., 2007; Matre & Solheim, 2016; Smail, 2020; Wyatt Smith et al., 2010) – Utilisation plus flexible des standards (Connolly et al., 2012; Klenowski et al., 2007; Matre & Solheim, 2016; Mottier Lopez et al., 2012; Wyatt Smith et al., 2010) – Responsabilisation accrue (Adie, 2014; Grant, 2012; Smail, 2020) – Construction d'un répertoire partagé (ressources, stratégies, connaissances) (Adie, 2014; Allal & Mottier Lopez, 2014; Hermansen & Nerland, 2014; Klenowski et al., 2007; Matre & Solheim, 2016; Mottier Lopez et al., 2012; Smail, 2020; Stewart & Houchens, 2014) 	<ul style="list-style-type: none"> – Leadership partagé entre enseignants et direction (Hipkins & Robertson, 2012) – Harmonisation des processus (Allal & Mottier Lopez, 2014; Buckley & Liscompe, 2020; Matre & Solheim, 2016; Stewart & Houchens, 2014; Yerly, 2021) – Facilitation de la poursuite du travail collectif (Grant, 2012) 	

Tableau 3 (suite)

Effets positifs du travail collectif pour l'évaluation des apprentissages selon les niveaux élèves et enseignants, établissements et système

	Niveau micro	Niveau méso	Niveau macro
Niveau élèves	Niveau enseignants	Niveau établissements	Niveau système
	<ul style="list-style-type: none"> – Planification des prochaines étapes du processus d'enseignement (Smail, 2020) – Construction d'une identité professionnelle partagée (Adie, 2013, 2014; Klenowski et al., 2007) – Réduction des écarts (Allal & Mottier Lopez, 2014) – Ajustements évaluatifs (Mottier Lopez et al., 2012; Nilsberth & Sandlund, 2020) – Amélioration des tâches évaluatives (Black et al., 2011; Mottier Lopez & Pasquini, 2017) – Construction commune de consensus (Buckley & Liscompe, 2020; McFadden et al., 2020; Morales Villabona & Mottier Lopez, 2016; Mottier Lopez et al., 2012; Wyatt Smith et al., 2010; Yerly, 2021) – Augmentation de l'engagement (Connolly et al., 2012; Hargreaves et al., 2013) – Renforcement du leadership (Stewart & Houchens, 2014) – Création de sens, de liens entre théorie et pratique (Hargreaves, 2012; McFadden et al., 2020) – Apaisement des tensions (Morales Villabona et Mottier Lopez, 2016) 		

Bien que tous les niveaux bénéficient de la collaboration, la plupart des effets positifs se situent au niveau micro, plus particulièrement au niveau des enseignants. Les conséquences positives les plus citées concernent le développement des pratiques évaluatives (n = 12), l'amélioration du jugement professionnel (n = 9), la construction de nouvelles connaissances professionnelles (n = 6), la construction d'une compréhension commune des critères et des normes (n = 11), le partage d'un répertoire commun (n = 8) et l'obtention de consensus (n = 6). Comme les autres effets repris dans cette partie du tableau, ils participent tous à l'amélioration du développement professionnel (*improvement*).

Au niveau des établissements, la collaboration entre différents niveaux (primaire et secondaire) permet aux enseignants d'appréhender les réalités de leurs collègues et de comprendre les politiques d'évaluation instaurées dans leurs institutions respectives (n = 5). La collaboration permet également d'harmoniser les processus soit en ce qui concerne l'évaluation proprement dite (n = 2), soit à propos d'éléments connexes comme l'organisation des rencontres avec les parents ou la manière de formuler les *feedbacks* directement dans les épreuves (n = 2) ou dans les bulletins (n = 1). La construction d'une culture partagée de l'évaluation est observée dans plusieurs études (n = 3). Selon Mottier Lopez (2013), une culture commune de l'évaluation est importante car elle fournit les bases pour rendre l'évaluation des apprentissages plus équitable dans une école.

Un autre constat qui ressort des résultats concerne le niveau macro. En effet, 12 études mettent en évidence la fonction liée à la reddition de comptes (*for accountability*). Les différents auteurs évoquent une amélioration de la qualité de l'évaluation, que ce soit sur le plan de la pertinence, de la validité et/ou de la fiabilité⁴. Avec le développement des pratiques évaluatives (n = 12), cet effet est celui qui apparaît le plus fréquemment dans tous les articles.

4. Selon De Ketele et Gérard (2005), «la pertinence est le caractère plus ou moins approprié de l'épreuve, selon qu'elle s'inscrit dans la ligne des objectifs visés [...], la validité est le degré d'adéquation entre ce que l'on déclare faire (évaluer telle ou telle dimension) et ce que l'on fait réellement, entre ce que l'outil mesure et ce qu'il prétend mesurer (...) et la fiabilité est le degré de confiance que l'on peut accorder aux résultats observés [...]» (p. 3).

Les effets négatifs

Comme pour la partie précédente, le tableau 4 présente la liste des désavantages du travail collectif. Les résultats montrent que le niveau macro n'est pas du tout impacté. À nouveau, c'est au niveau micro que les effets sont les plus nombreux et, plus particulièrement, au niveau des enseignants. À la différence du tableau précédent, nous constatons une prévalence accrue d'effets de nature socioémotionnelle. Le travail collectif portant sur l'évaluation des apprentissages crée des tensions interpersonnelles ($n = 2$), suscite de la timidité ($n = 3$), des réactions émotionnelles vives (non précisées) ($n = 1$) ou la crainte d'exposer ses pratiques évaluatives ($n = 2$). Il arrive que les discussions autour de l'évaluation engendrent des réticences plus ou moins fortes. Les enseignants refusent par exemple d'adopter de nouvelles pratiques évaluatives ($n = 1$) ou s'engagent moins dans les échanges ($n = 3$). Dans plusieurs études, nous constatons que la collaboration renforce la tendance de certains enseignants à adopter une approche instrumentale des critères ($n = 2$). Ils s'y réfèrent alors strictement et justifient leur choix à l'aide d'arguments prônant l'égalité de traitement.

Au niveau des établissements, nous observons peu de conséquences négatives dans les 30 articles. Dans leur étude, Black et al. (2011) expliquent que les pratiques évaluatives collectives sont trop influencées par les évaluations externes lors de la conception commune d'épreuves. Les auteurs ont observé que les enseignants avaient tendance à en reproduire la forme sans réellement en comprendre le fond et que cela ne participait pas vraiment à leur développement professionnel. Ils constatent aussi que les mésententes et les controverses ont provoqué un désalignement entre l'évaluation et l'enseignement. L'évaluation proposée peut alors être entachée de biais importants. Dans une autre étude où le travail collectif a été provoqué entre enseignants du primaire et du secondaire, les échanges ont contribué à raviver des tensions existantes entre les écoles.

Tableau 4

Effets négatifs du travail collectif pour l'évaluation des apprentissages selon les niveaux élèves et enseignants, établissements et système

	Niveau micro	Niveau méso	Niveau macro
Niveau élèves	Niveau enseignants	Niveau établissements	Niveau système
– Construction de critères désavantageux (McFadden et al., 2020)	<ul style="list-style-type: none"> – Jugement professionnel davantage impacté par les opinions personnelles (Van Der Schaaf et al., 2012) – Fixation mécanique des seuils (Mottier Lopez & Pasquini, 2017) – Réactions émotionnelles (Crisp, 2017) – Incompréhension des critères (Reid, 2007) – Crainte d'exposer ses pratiques (Adie, 2013; Nilsberth & Sandlund, 2020) – Réduction de l'implication, réticences (Adie, 2013; Connolly et al., 2012; Wyatt Smith et al., 2010) – Timidité, gêne (Adie, 2013; Nilsberth & Sandlund, 2020; Wyatt Smith et al., 2010) – Pertes de temps, surcharge de travail (Klenowski et al., 2007) – Luttes d'influence (Morales Villabona & Mottier Lopez, 2016; Nilserth & Sandlund, 2020) – Approche instrumentale (inflexible) des standards (Hipkins & Robertson, 2012; Nilserth & Sandlund, 2020) – Tensions interpersonnelles (Black et al., 2011; Yerly, 2021) – Arrangements évaluatifs (Allal & Mottier Lopez, 2014; Colbert et al., 2012; Hipkins & Robertson, 2012; Wyatt Smith et al., 2010) – Notation analytique (Klenowski et al., 2007) – Difficulté d'obtenir des consensus (Mottier Lopez et al., 2012) – Conservation des pratiques ancestrales (Heredia et al., 2016) 	<ul style="list-style-type: none"> – Mésusage des tests externes (Black et al., 2011) – Désalignement entre l'évaluation et l'enseignement (Black et al., 2011) – Relations interinstitutionnelles (Morales Villabona & Mottier Lopez, 2016) 	

Discussion

Les questionnements soulevés dans les différentes parties de cette discussion apparaissent comme des enjeux contemporains intrinsèquement liés à une tendance émergente en éducation : le travail collectif des enseignants pour l'évaluation des apprentissages. Pour ce faire, nous discutons les résultats en questionnant plusieurs tensions qui émergent à la suite de notre analyse documentaire.

La distinction entre contrôle et évaluation permet d'éclairer les objectifs et les implications de la modération sociale. Elle met en évidence les différentes dimensions du processus et incite à réfléchir sur la façon dont elle est pratiquée et interprétée. Nous discutons ensuite des défis du travail formel et informel en évaluation. Nous observons comment le passage de l'informel au formel peut favoriser un jugement professionnel situé. La discussion porte également sur le défi consistant à équilibrer les besoins en matière de formation des enseignants à l'évaluation et les bénéfices du développement professionnel engendrés par le travail collectif. Enfin, le concept du jugement professionnel situé est confronté aux résultats de l'analyse documentaire.

En ce qui concerne les objectifs du travail collectif pour l'évaluation, nos résultats montrent que la modération sociale est la finalité la plus répandue. Au début des années 2000, pour Maxwell (2001), la fonction de développement professionnel (*for improvement*) de la modération sociale est secondaire par rapport à la fonction de reddition des comptes (*for accountability*). Notre analyse montre une tendance inverse 20 ans plus tard : la finalité de développement professionnel ($n = 13$) est plus présente que celle de reddition de comptes ($n = 3$). Dans ce contexte, comme l'ont fait Ardoino et Berger (1990) en appréhendant l'évaluation en éducation sous des angles fondateurs, nous différencions ici le contrôle de l'évaluation. Les auteurs distinguent contrôle et évaluation par une approche étymologique qui fait jouer le dialogisme plus-valeur/plus-value. Ils définissent le contrôle comme un ensemble de procédures visant à vérifier la mise en œuvre des plans ou des programmes, en lien avec la reddition des comptes et le pilotage. En revanche, ils conçoivent l'évaluation comme un processus qui vise à extraire la valeur d'une tâche, et qui se concentre sur l'amélioration et le développement. Cette distinction nous permet d'éclairer les différentes finalités du travail collectif en évaluation, en montrant que l'accent est de plus en plus mis sur l'amélioration des pratiques plutôt que sur le simple contrôle des résultats.

Nos résultats montrent que, lorsque les enseignants utilisent des critères communs, des pratiques informelles (individuelles) se confrontent à d'autres qui sont plus formelles. Dès lors, une question émerge : comment ces deux aspects interagissent-ils pour soutenir un jugement professionnel situé ? Les résultats montrent que le travail formel peut permettre de visibiliser des pratiques informelles, parfois clandestines, comme effectuer des arrangements évaluatifs. Or, ce genre de pratique pourrait être une source de tensions quant à l'application de procédures, de recommandations ou de normes en matière de pratiques évaluatives. En ce qui concerne l'évaluation informelle, les observations en classe sur lesquelles les enseignants se basent ne sont pas nécessairement planifiées. Les pratiques informelles sont influencées par des considérations contextuelles. En ce sens, le travail informel autour de l'évaluation est situé. Les pratiques informelles peuvent donc servir de point de départ à des discussions formelles, pouvant générer un dialogue professionnel, par exemple sur les approches d'évaluation et sur les jugements professionnels. En progressant de l'informel vers le formel, les enseignants peuvent réfléchir sur leurs pratiques évaluatives. Le travail formel peut alors être l'occasion de clarifier et de formaliser les normes professionnelles en matière d'évaluation. Il peut aussi contribuer à la construction d'un répertoire partagé d'outils, de pratiques et de méthodes, conformément aux attentes institutionnelles. Les discussions formelles sont aussi l'occasion d'approfondir la compréhension des considérations contextuelles, fournissant des occasions de les intégrer dans le processus d'évaluation des apprentissages, favorisant ainsi un jugement professionnel situé.

En ce qui a trait aux conséquences du travail collectif, nos résultats s'accordent avec ceux de Vangrieken et al. (2015). Bien que tous les niveaux soient touchés, les effets se concentrent sur les enseignants. Il semblerait que les avantages dépassent les inconvénients. Alors que les effets positifs contribuent principalement au développement professionnel, les effets négatifs sont majoritairement liés à des aspects socioémotionnels (par ex., timidité, gêne). Il semblerait que les enseignants aient du mal à exposer leurs pratiques évaluatives publiquement. Pour Mottier Lopez et Allal (2014), ces pratiques restent en effet souvent une affaire privée. Ces constats peuvent être mis en lien avec les recherches qui s'accordent à la fois pour dire que les enseignants se sentent mal formés (Mertler, 2004) et pour pointer des lacunes dans leurs pratiques évaluatives (De Luca et al., 2016). Ainsi, se dessine une tension entre la nécessité de mieux former les enseignants à l'évaluation et les bénéfices du développement professionnel via le travail collectif.

Cette tension souligne le défi qui consiste à équilibrer ces deux aspects. Dès lors, devrait-on miser davantage sur la formation (initiale et continue) ou sur la mise en place plus systématique d'une évaluation collective ?

L'un des effets les plus présents au niveau des enseignants concerne le développement de leur jugement professionnel (n = 9). Dans plusieurs recherches (n = 4), les participants modifient leur jugement professionnel en tenant compte du contexte socioculturel. Dans une perspective située de l'évaluation (Mottier Lopez, 2013 ; 2021), le contexte de la classe et de l'école ainsi que le rôle du langage, de la parole et des outils culturels sont importants. Une analyse plus fine des résultats montre que sur les neuf articles qui évoquent le développement du jugement professionnel, sept concernent des dispositifs organisés pour modérer les résultats. Il semblerait que le jugement professionnel n'y soit conceptualisé qu'à la fin du processus, n'intervenant que lors de l'attribution de la note. Or, les travaux de Allal et Mottier Lopez (2008) invitent à considérer que ce jugement professionnel intervient tout au long du processus. Ces travaux permettent de se rendre compte que le jugement professionnel est un acte cognitif et une pratique sociale située qui s'exerce tout au long du processus évaluatif. Cette conception du jugement professionnel en continu se confronte donc aux contextes ne l'instituant que lors des étapes de correction et de notation (par ex., lors des séances de modération en Australie), mais s'aligne avec les études menées en milieu naturel par Yerly (2021) dans une école secondaire en Suisse romande et par Holmeier et al. (2017)⁵ dans plusieurs écoles secondaires en Suisse alémanique. Dès lors, il est légitime de se demander à quelles étapes de l'évaluation la collaboration doit nécessairement jouer un rôle et si certaines de ces étapes sont prioritaires pour atteindre les objectifs.

Conclusion

L'objectif principal de cette recherche est de donner un aperçu de la littérature scientifique anglophone et francophone sur le travail collectif des enseignants portant sur l'évaluation des apprentissages, au niveau de son organisation et de ses conséquences. L'examen s'est concentré sur quatre axes, à savoir 1) les finalités, 2) les outils et les méthodes, 3) les dimensions de l'évaluation et 4) les effets pour tous les acteurs de l'enseignement.

5. Cet article ne fait pas partie de notre sélection, car il est écrit en allemand.

Il apparaît que le concept d'évaluation collective est émergent et plus présent dans la littérature anglophone que francophone. En francophonie, l'organisation des systèmes éducatifs est moins homogène. Par exemple, le système français se démarque par son centralisme national au regard du monde anglo-saxon, plus fédératif. Cela engendre des différenciations culturelles, notamment en matière de politiques évaluatives. Le travail collectif pour l'évaluation suit plusieurs objectifs : modérer les résultats, concevoir des outils, développer de nouvelles pratiques et mettre des directives en œuvre. Quand les enseignants travaillent collectivement, ils utilisent des outils personnels construits collectivement ou mis à leur disposition. Lorsqu'ils se regroupent, ils travaillent en équipe ou ont recours à des communautés. La notation est la dimension de l'évaluation la plus touchée par le travail collectif. Les référents de l'évaluation (objectifs, tâches, critères, seuils, pondérations) sont également des dimensions fortement discutées lors de la conception commune d'épreuves. Tous les niveaux bénéficient du travail collectif. Les effets positifs se situent majoritairement au niveau des enseignants (par ex., développement des pratiques et des connaissances, amélioration du jugement professionnel). Notre aperçu montre également que le travail collectif permet au système d'améliorer la qualité des évaluations qu'il propose. Cependant, certains effets négatifs sont également observés, surtout pour les enseignants (par ex., timidité, craintes, tensions interpersonnelles).

Même si elle se veut la plus rigoureuse possible, cette recherche comporte certaines limites. Tout d'abord, nous pointons un biais méthodologique lié à la méthode de recherche. Si la *rapid review* rationalise et accélère le processus de recherche systématique, elle augmente également le risque de biais, par exemple en matière de sélection d'études ou d'inclusion d'études faibles. Le recours à d'autres bases de données pourrait par exemple mener à une sélection plus riche. Il est donc possible que nous soyons passés à côté d'études pertinentes. Une approche strictement systématique aurait sans doute permis de conduire à une sélection plus exhaustive et qualitative des sources et aurait balisé plus efficacement la prise de décision. Un biais linguistique est également présent. En effet, nos critères d'inclusion restreignent nos recherches aux seuls articles en anglais et en français, excluant *de facto* des études publiées dans d'autres langues. Un biais quant à l'interprétation des articles est également à prendre en considération.

Notre analyse documentaire débouche sur des enjeux contemporains et émergents en matière d'évaluation des apprentissages. Cependant, elle en fait apparaître d'autres que nous n'avons pas l'occasion de traiter. Par exemple, dans nos résultats, nous analysons indifféremment le travail collectif pour les évaluations à faibles et à hauts enjeux. La réalisation de futures recherches visant à déterminer si les dynamiques collectives, telles que les attitudes, les outils, les méthodes et les effets varient en fonction du contexte constitue une question intéressante à explorer. La littérature montre que le travail collectif est plus constructif lorsque les enseignants ne sont pas (trop) contraints (Hargreaves, 2012, 2021 ; Vangrieken et al., 2015). Étudier cet aspect dans un contexte spécifique à l'évaluation des apprentissages en est également un autre.

Révision linguistique : Marie-Claire Legaré

Mise en page : Emmanuel Gagnon

Résumé en portugais : Eusébio André Machado

Réception : 30 septembre 2023

Version finale : 18 avril 2024

Acceptation : 12 septembre 2024

LISTE DE RÉFÉRENCES

- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. Dans V. Klenowski (dir.), *International Teacher Judgement Practices* (p. 101-116). Routledge. <https://doi.org/10.1080/0969594X.2011.650150>
- Adie, L. (2014). The development of shared understandings of assessment policy: Travelling between global and local contexts. *Journal of Education Policy*, 24(4), 532-545. <https://doi.org/10.1080/02680939.2013.853101>
- Adie, L. E., Klenowski, V. & Wyatt-Smith, C. (2012). Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*, 64(2), 223-240. <https://doi.org/10.1080/00131911.2011.598919>
- Allal, L. & Mottier Lopez, L. (2008). Mieux comprendre le jugement professionnel en évaluation : apports et implications de l'étude genevoise. Dans L. Lafortune & L. Allal (dir.), *Jugement professionnel en évaluation : pratiques enseignantes au Québec et à Genève* (p. 223- 239). Presses de l'Université du Québec.

- Allal, L. & Mottier Lopez, L. (2009). Au cœur du jugement professionnel en évaluation : des démarches de triangulation. *Les dossiers des sciences de l'éducation*, 22(1), 25-40. <https://doi.org/10.3406/dsedu.2009.1164>
- Allal, L. & Mottier Lopez, L. (2014). Teachers' professional judgment in the context of collaborative assessment practice. Dans C. Wyatt-Smith, V. Klenowski & P. Colbert (dir.), *Designing assessment for quality learning* (p. 151-165). https://doi.org/10.1007/978-94-007-5902-2_10
- Allenbach, M. (2016). Négociation de pratiques collaboratives et construction du rôle des intervenant-es à l'école [Thèse de doctorat, Université de Fribourg].
- Ardoino, J. & Berger, G. (1990). Étude critique - Vers une nouvelle conception de l'évaluation? *Revue française de pédagogie*, (90), 105-112. https://www.persee.fr/doc/rfp_0556-7807_1990_num_90_1_2466_t1_0105_0000_1
- Barrère, A., Saujat, F. & Lantheaume, F. (2008). Rendre visible le travail enseignant. Questions de méthodes : Entretien de Anne Barrère et Frédéric Saujat par Françoise Lantheaume. *Recherche et formation*, 57(1). <https://doi.org/10.4000/rechercheformation.861>
- Beuret, J. E. (2010). De la négociation conflictuelle à la négociation concertative : un «Point de Passage Transactionnel». *Négociations*, (1), 43-60. <https://www.cairn.info/revue-negociations-2010-1-page-43.htm>
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451-469. <https://doi.org/10.1080/0969594X.2011.557020>
- Borgès, C. & Lessard, C. (2007). Qu'arrive-t-il quand la collaboration enseignante devient une norme? Dans J.-F. Marcel (dir), *Coordonner, collaborer, coopérer : de nouvelles pratiques enseignantes* (p. 61-74). De Boeck Supérieur. <https://doi.org/10.3917/dbu.marce.2007.01.0061>
- Buckley, K. & Lipscombe, K. (2022). Validity and the design of classroom assessment in 77478/teacher teams. *The Australian Educational Researcher*, 49(2), 425-444. <http://dx.doi.org/10.1007/s13384-021-00437-9>
- Cohen, S. G. & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23, 239-290. <https://doi.org/10.1177/01492063970230>
- Colbert, P., Wyatt-Smith, C. & Klenowski, V. (2012). A systems-level approach to building sustainable assessment cultures: Moderation, quality task design and dependability of judgement. *Policy Futures in Education*, 10(4), 386-401. <https://doi.org/10.2304/pfe.2012.10.4.386>
- Connolly, S., Klenowski, V. & Wyatt-Smith, C. M. (2012). Moderation and consistency of teacher judgement: teachers' views. *British Educational Research Journal*, 38(4), 593-614. <https://doi.org/10.1080/01411926.2011.569006>
- Crisp, V. (2017). The judgement processes involved in the moderation of teacher-assessed projects. *Oxford Review of Education*, 43(1), 19-37. <https://doi.org/10.1080/03054985.2016.1232245>
- Dachet, D. (2024). Umbrella Review : de sa méthodologie aux considérations méthodologiques relatives à son application dans le champ des sciences de l'Éducation [Thèse de doctorat, Université de Liège].

- Decuyper, S., Dochy, F. & Van den Bossche, P. (2010). Grasping the dynamic complexity of team learning: An integrative model for effective team learning in organizations. *Educational Research Review*, 5, 111–133. <https://doi.org/10.1016/j.edurev.2010.02.002>
- De Ketele, J.-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation*, 29(1), 99-118. <https://doi.org/10.7202/1086970ar>
- De Ketele, J.-M. & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et Évaluation en Éducation*, 28(3), 1-26. <https://doi.org/10.7202/1087028ar>
- 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. <https://doi.org/10.1007/s11092-015-9233-6>
- Doppenberg, J. J., Den Brok, P. J. & Bakx, A. W. E. A. (2012). Collaborative teacher learning across foci of collaboration: Perceived activities and outcomes. *Teachers and Teacher Education*, 28, 899–910. <https://doi.org/10.1016/j.tate.2012.04.007>
- Dumez, H. (2011). Faire une revue de littérature : pourquoi et comment? *Le libellio d'aegis*, 7(2), 15-27. <https://hal.archives-ouvertes.fr/hal-00657381>
- Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D. & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science*, 44, 267-291. <https://doi.org/10.1007/s11251-016-9371-3>
- Germier, C. & Marcel, J. F. (2016). La négociation des accords. Approche ethnographique des pratiques de deux collectifs d'enseignants. *Les dossiers des sciences de l'éducation*, (35), 73-96. <https://doi.org/10.4000/dse.1271>
- Gilbert, A.-F. (2018). *Le travail collectif enseignant, entre informel et institué*. Dossier de veille de l'IFÉ, n° 124.
- Gipps, C. (1994, 4-8 avril). *Quality assurance in teachers' assessment* [Présentation]. Annual meeting of the American Educational Research Association, New-Orleans, USA.
- Grangeat, M. (2011). Le travail collectif enseignant : éléments de modélisation du développement professionnel. Dans M. Grangeat (dir.), *Les démarches d'investigation dans l'enseignement scientifique. Pratiques de classe, travail collectif enseignant, acquisitions des élèves* (p. 79-106). *École Normale Supérieure*. <https://shs.hal.science/halshs-01205850>
- Grant, L. (2012). Cross-sector moderation as a means of engaging staff with assessment and teaching pedagogy. *Policy Futures in Education*, 10(4), 434-446. <https://doi.org/10.2304/pfie.2012.10.4.434>
- Hargreaves, A. (1994). *Changing teachers, changing times: Teachers' work and culture in the postmodern age*. OISE Press.
- Hargreaves, A. (2021). Teacher collaboration: 30 years of research on its nature, forms, limitations and effects. *Teachers and Teaching*, 25(5), 603-621. <https://doi.org/10.1080/13540602.2019.1639499>
- Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education*, 14(2), 185-199. <https://doi.org/10.1080/09695940701478594>
- Hargreaves, E. (2012). Assessment for learning and teacher learning communities: UK teachers' experiences. *Teaching Education*, 24(3), 327-344. <https://doi.org/10.1080/10476210.2012.713931>

- Hargreaves, E., Berry, R., Lai, Y. C., Leung, P., Scott, D. & Stobart, G. (2013). Teachers' experiences of autonomy in continuing professional development: Teacher learning communities in London and Hong Kong. *Teacher development*, 17(1), 19-34. <https://doi.org/10.1080/13664530.2012.748686>
- Hattie, J. (2017). *L'apprentissage visible pour les enseignants. Connaître son impact pour maximiser le rendement des élèves*. PUQ.
- Heredia, S. C., Furtak, E. M., Morrison, D. & Renga, I. P. (2016). Science teachers' representations of classroom practice in the process of formative assessment design. *Journal of Science Teacher Education*, 27(7), 697-716. <https://doi.org/10.1007/s10972-016-9482-3>
- Hermansen, H. (2014). Recontextualising assessment resources for use in local settings: Opening up the black box of teachers' knowledge work. *Curriculum Journal*, 25(4), 470-494. <https://doi.org/10.1080/09585176.2014.956771>
- Hermansen, H. & Nerland, M. (2014). Reworking practice through an AfL project: An analysis of teachers' collaborative engagement with new assessment guidelines. *British educational research journal*, 40(1), 187-206. <https://doi.org/10.1002/berj.3037>
- Hipkins, R. & Robertson, S. (2012). The complexities of moderating student writing in a community of practice. *Assessment Matters*, 4, 30-52. <https://doi.org/10.18296/am.0099>
- Holmeier, M., Maag Merki, K. & Hirt, C. (2017). *Gemeinsames Prüfen. Eine Fallanalyse in Gymnasien in der Schweiz*. Springer. <https://doi.org/10.1007/978-3-658-17060-8>
- Klenowski, V., Adie, L., Gunn, S., Looney, A., Elwood, J., Wyatt-Smith, C. & Colbert, P. (2007, 25-29 novembre). *Moderation as judgement practice: Reconciling system level accountability and local level practice*. Dans P.-L. Jeffery (chair), *Conference: Research impacts: Proving or improving?* [Conférence]. Australian Association for Research in Education, Fremantle, Australia.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press.
- Laveault, D. & Yerly, G. (2017). Modération statistique et modération sociale des résultats scolaires : approches opposées ou complémentaires? *Mesure et Évaluation en Éducation*, 40(2), 91-123. <https://doi.org/10.7202/1043569ar>
- Letor, C. (2010). Moments de coopération entre enseignants : entre logiques institutionnelles, organisationnelles et professionnelles. *Travail et formation en éducation*, 7, 107-120. <http://journals.openedition.org/tfe/1458>
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83-102. https://doi.org/10.1207/s15324818ame0601_5
- Marcel, J.-F., Dupriez, V., Perisset Bagnoud, D. & Tardif, M. (2007). *Coordonner, collaborer, coopérer : de nouvelles pratiques enseignantes*. De Boeck.
- Maroy, C. (2005). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? *Cahiers de la recherche en Éducation et Formation*, (49). <http://hdl.handle.net/2078.1/81045>
- Maroy, C. & Voisin, A. (2014). Une typologie des politiques d'accountability en éducation : l'incidence de l'instrumentation et des théories de la régulation. *Éducation comparée. Revue de recherche internationale et comparative en éducation*, 11, 31-58. <https://doi.org/1866/14074>

- Matre, S. & Solheim, R. (2016). Opening dialogic spaces: Teachers' metatalk on writing assessment. *International Journal of Educational Research*, 80, 188-203. <https://doi.org/10.1016/j.ijer.2016.07.001>
- Maxwell, G. S. (2001). *Moderation of assessments in vocational education and training*. Queensland Government Departement of Employment and Training.
- McFadden, J., Jung, K., Robinson, B. & Tretter, T. R. (2022). Teacher-developed Multi-Dimensional Science Assessments Supporting Elementary Teacher Learning about the Next Generation Science Standards. *Journal of Science Teacher Education*, 33(1), 55-82. <https://doi.org/10.1080/1046560X.2021.1905331>
- McLaughlin, M. W. & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. University of Chicago Press.
- Meirieu, P. (2023). Du bon usage de l'autonomie des établissements scolaires. *Administration et Éducation*, 1, 127-141. <https://doi.org/10.3917/admed.177.0127>
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American secondary education*, 33(1) 49-64. <https://www.jstor.org/stable/41064623>
- Morales Villabona, F. & Mottier Lopez, L. (2016). Quelle évaluation collaborative dans la modération sociale entre enseignants? Dans. Mottier Lopez & W. Tessaro (dir.), *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages* (p. 289-312). Peter Lang.
- Mottier Lopez, L. (2008). *Apprentissage situé: la microculture de classe en mathématiques*. Peter Lang.
- Mottier Lopez, L. (2013). De la mesure à l'évaluation collaborative en éducation. *Revue française d'administration publique*, 148(4), 939-952. <https://doi.org/10.3917/rfap.148.0939>
- Mottier Lopez, L. (2021). Une évaluation continue pour apprendre durablement, une évaluation à visée inclusive. *Revue suisse de pédagogie spécialisée*, 11(4), 9-16. <https://ojs.szh.ch/revue/article/view/168>
- Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation: un acte cognitif et une pratique sociale située. *Schweizerische Zeitschrift für Bildungswissenschaften*, 30(3), 465-482. <https://doi.org/10.25656/01:4229>
- Mottier Lopez, L. & Pasquini, R. (2017). Professional controversies between teachers about their summative assessment practices: A tool for building assessment capacity. *Assessment in Education: Principles, Policy & Practice*, 24(2), 228-249. <https://doi.org/10.1080/0969594X.2017.1293001>
- Mottier Lopez, L., Tessaro, W., Dechamboux, L. & Villabona, F.- M. (2012). La modération sociale: un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves. *Questions vives: Recherches en éducation*, 6(18), 159-175. <https://doi.org/10.4000/questionsvives.1235>
- Nilsberth, M. & Sandlund, E. (2021). On the interactional challenges of revealing summative assessments: Collaborative scoring talk among teachers and students in Swedish national tests. *Linguistics and education*, 61(2), 1-16. <https://doi.org/10.1016/j.linged.2020.100899>
- Page M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffman, T. C. & Mulrow, C. D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic review. *British Medical Journal*, 372(1). <https://doi.org/10.1136/bmj.n71>

- Reid, L. (2007). Teachers talking about writing assessment: valuable professional learning? *Improving Schools*, 10(2), 132-149. <https://doi.org/10.1177/1365480207077843>
- Salas, E., Burke, S. C. & Cannon-Bowers, J. A. (2000). Teamwork: Emerging principles. *International Journal of Management Reviews*, 2, 339-356. <https://doi.org/10.1111/1468-2370.00046>
- Smail, E. (2020). Using involvement in moderation to strengthen teachers' assessment for learning capability. *Assessment in Education: Principles, Policy & Practice*, 27(5), 522-543. <https://doi.org/10.1080/0969594X.2020.1777087>
- Spillane, J. P., Shirrell, M. & Hopkins, M. (2016). Designing and deploying a professional learning community (PLC) organizational routine: Bureaucratic and collegial arrangements in tandem. *Les dossiers des sciences de l'éducation*, 35, 97-122. <https://doi.org/10.4000/dse.1283>
- Stewart, T. A. & Houchens, G. W. (2014). Deep impact: How a job-embedded formative assessment professional development model affected teacher practice. *Qualitative Research in Education*, 3(1), 51-82. <https://doi.org/10.4471/qre.2014.36>
- Strauss, A. (1992). *La trame de la négociation*. L'Harmattan.
- Van der Schaaf, M., Baartman, L. & Prins, F. (2012). Exploring the role of assessment criteria during teachers' collaborative judgement processes of students' portfolios. *Assessment & Evaluation in Higher Education*, 37(7), 847-860. <https://doi.org/10.1080/002602938.2011.576312>
- Vangrieken, K., Dochy, F., Raes, E. & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational research review*, 15, 17-40. <http://dx.doi.org/10.1016/j.edurev.2015.04.002>
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.
- Wenger, E. (2005). *La théorie des communautés de pratique*. Presses de l'Université Laval.
- Wyatt-Smith, C., Klenowski, V. & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice*, 17(1), 59-75. <https://doi.org/10.1080/09695940903565610>
- Yerly, G. (2021). La collaboration des enseignants autour de l'évaluation des apprentissages comme moyen de piloter le système? Le cas de la politique «évaluation commune» dans le secondaire supérieur en Suisse. *Les dossiers des sciences de l'éducation*, 44, 77-97. <https://doi.org/10.4000/dse.5322>

Annexe A - Corpus complet d'articles : auteurs, titres, années de publication, pays d'origine, niveaux et méthodes de recherche

Auteurs	Titres	Années	Pays	Niveaux d'enseignement	Types de recherche
Adie	<i>The development of teacher assessment identity through participation in online moderation</i>	2013	Australie (Queensland)	Primaire et secondaire	Qualitative
Adie	<i>The development of shared understandings of assessment policy: Travelling between global and local contexts</i>	2014	Australie (Queensland)	Secondaire	Qualitative
Adie et al.	<i>Towards an understanding of teacher judgement in the context of social moderation</i>	2012	Australie (Queensland)	Secondaire	Qualitative
Allal & Mottier Lopez	<i>Mieux comprendre le jugement professionnel en évaluation: apports et implications de l'étude genevoise</i>	2014	Suisse (Genève)	Secondaire	Qualitative
Black et al.	<i>Can teachers' summative assessments produce dependable results and also enhance classroom learning?</i>	2011	Grande-Bretagne	Secondaire	Qualitative
Buckley & Liscompe	<i>Validity and the design of classroom assessment in teacher teams</i>	2022	Australie (Nouvelle Galles du Sud)	Primaire	Qualitative
Colbert et al.	<i>A systems-level approach to building sustainable assessment cultures: Moderation, quality task design and dependability of judgement</i>	2012	Australie (Queensland)	Secondaire	Qualitative
Connolly et al.	<i>Moderation and consistency of teacher judgement: teachers' views</i>	2012	Australie (Queensland)	Secondaire	Qualitative
Crisp	<i>The judgement processes involved in the moderation of teacher-assessed projects</i>	2017	Grande-Bretagne	Secondaire	Qualitative
Furtak et al.	<i>Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study</i>	2016	États-Unis	Secondaire	Quantitative

Annexe A - Corpus complet d'articles: auteurs, titres, années de publication, pays d'origine, niveaux et méthodes de recherche (suite)

Auteurs	Titres	Années	Pays	Niveaux d'enseignement	Types de recherche
Grant	<i>Cross-sector moderation as a means of engaging staff with assessment and teaching pedagogy</i>	2012	Grande-Bretagne (Écosse)	Primaire et secondaire	Qualitative
Hargreaves	<i>Assessment for learning and teacher learning communities: UK teachers' experiences</i>	2012	Grande-Bretagne	Secondaire	Qualitative
Hargreaves et al.	<i>Teachers' experiences of autonomy in continuing professional development</i>	2013	Grande-Bretagne et Hong Kong	Secondaire	Qualitative
Heredia et al.	<i>Science teachers' representations of classroom practice in the process of formative assessment design</i>	2016	États-Unis	Secondaire	Qualitative
Hermansen	<i>Recontextualising assessment resources for use in local settings: Opening up the black box of teachers' knowledge work</i>	2014	Norvège	Secondaire	Qualitative
Hermansen & Nerland	<i>Reworking practice through an AfL project: An analysis of teachers' collaborative engagement with new assessment guidelines</i>	2014	Norvège	Secondaire	Qualitative
Hipkins & Robertson	<i>The complexities of moderating student writing in a community of practice</i>	2012	Nouvelle-Zélande	Secondaire	Qualitative
Klenowski et al.	<i>Moderation as judgement practice: Reconciling system level accountability and local level practice</i>	2007	Australie (Queensland)	Secondaire	Qualitative
Matre & Solheim	<i>Opening dialogic spaces: Teachers' metatalk on writing assessment</i>	2016	Norvège	Secondaire	Qualitative

Annexe A - Corpus complet d'articles : auteurs, titres, années de publication, pays d'origine, niveaux et méthodes de recherche (suite)

Auteurs	Titres	Années	Pays	Niveaux d'enseignement	Types de recherche
McFadden et al.	<i>Teacher-developed Multi-Dimensional Science Assessments Supporting Elementary Teacher Learning about the Next Generation Science Standards</i>	2022	États-Unis (Kentucky)	Primaire	Qualitative
Morales Villabona & Mottier Lopez	<i>Quelle évaluation collaborative dans la modération sociale entre enseignants?</i>	2016	Suisse (Genève)	Primaire et secondaire	Qualitative
Mottier Lopez & Pasquini	<i>Professional controversies between teachers about their summative assessment practices: A tool for building assessment capacity</i>	2017	Suisse (Genève et Vaud)	Primaire ou secondaire (deux recherches distinctes)	Qualitative
Mottier Lopez et al.	<i>La modération sociale : un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves</i>	2012	Suisse (Genève)	Primaire	Qualitative
Nilsberth & Sandlund	<i>On the interactional challenges of revealing summative assessments: Collaborative scoring talk among teachers and students in Swedish national tests</i>	2021	Suède	Secondaire	Qualitative
Reid	<i>Teachers talking about writing assessment: valuable professional learning?</i>	2007	Grande-Bretagne (Écosse)	Primaire et secondaire	Qualitative
Smail	<i>Using involvement in moderation to strengthen teachers' assessment for learning capability</i>	2020	Nouvelle-Zélande	Primaire	Qualitative
Stewart & Houchens	<i>Deep impact: How a job-embedded formative assessment professional development model affected teacher practice</i>	2014	États-Unis (Kentucky)	Primaire	Qualitative
Van Der Schaaf et al.	<i>Exploring the role of assessment criteria during teachers' collaborative judgement processes of students' portfolios</i>	2012	Pays-Bas	Secondaire	Quantitative

Annexe A - Corpus complet d'articles : auteurs, titres, années de publication, pays d'origine, niveaux et méthodes de recherche (suite)

Auteurs	Titres	Années	Pays	Niveaux d'enseignement	Types de recherche
Wyatt Smith et al.	<i>The centrality of teachers' judgement practice in assessment: A study of standards in moderation</i>	2010	Australie (Queensland)	Primaire et secondaire	Mixte
Yerly	<i>La collaboration des enseignants autour de l'évaluation des apprentissages comme moyen de piloter le système? Le cas de la politique « évaluation commune » dans le secondaire supérieur en Suisse.</i>	2021	Suisse (Fribourg)	Secondaire	Qualitative