Document généré le 12 août 2025 17:58

Canadian Journal of Education
Revue canadienne de l'éducation

Comment appréhender une littératie en évaluation pour décrire et comprendre le développement des compétences des enseignants ?

Raphaël Pasquini et Fernando Morales Villabona

Volume 45, numéro 2, été 2022

URI : https://id.erudit.org/iderudit/1090614ar
DOI : https://doi.org/10.53967/cje-rce.v45i2.5171

Aller au sommaire du numéro

Éditeur(s)

Canadian Society for the Study of Education

ISSN

0380-2361 (imprimé)
1918-5979 (numérique)

Découvrir la revue

Citer cet article

Pasquini, R. & Morales Villabona, F. (2022). Comment appréhender une littératie en évaluation pour décrire et comprendre le développement des compétences des enseignants ? *Canadian Journal of Education / Revue canadienne de l'éducation*, *45*(2), 512–554. https://doi.org/10.53967/cje-rce.v45i2.5171

Résumé de l'article

Notre recherche interroge dans quelle mesure la littératie en évaluation (LE), peu développée dans la communauté de recherche francophone, apparaît porteuse pour décrire et comprendre le développement de compétences chez six enseignants de mathématiques et de français du secondaire, en Suisse romande. L'analyse des données issues d'un dispositif de formation favorisant les échanges et la création d'épreuves sommatives en exploitant des apports théoriques spécifiques permet d'explorer le développement dynamique des compétences des enseignants, et par le fait même, celui de leur LE, à l'aide d'une méthodologie exploitant les dimensions recensées dans la littérature contemporaine. Plus précisément, nos résultats montrent que les pratiques d'évaluation gagnent en cohérence, mais que les pratiques de notation rencontrent toujours des obstacles. La LE s'inscrit ici dans une épistémologie située et non technique. Elle montre toutefois la nécessité d'élargir encore l'analyse des négociations menées par les enseignants dans leurs pratiques évaluatives pour mieux en comprendre les caractéristiques.

© Canadian Society for the Study of Education, 2022

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Comment appréhender une littératie en évaluation pour décrire et comprendre le développement des compétences des enseignants ?

Raphaël Pasquini
Haute École Pédagogique du canton de Vaud

Fernando Morales Villabona
Haute École Pédagogique du canton de Vaud

Résumé

Notre recherche interroge dans quelle mesure la littératie en évaluation (LE), peu développée dans la communauté de recherche francophone, apparaît porteuse pour décrire et comprendre le développement de compétences chez six enseignants de mathématiques et de français du secondaire, en Suisse romande. L'analyse des données issues d'un dispositif de formation favorisant les échanges et la création d'épreuves sommatives en exploitant des apports théoriques spécifiques permet d'explorer le développement dynamique des compétences des enseignants, et par le fait même, celui de leur LE, à l'aide d'une méthodologie exploitant les dimensions recensées dans la littérature contemporaine.

Canadian Journal of Education / Revue canadienne de l'éducation 45:2 (2022)
**©2022 Canadian Society for the Study of Education/
Société canadienne pour l'étude de l'éducation**
www.cje-rce.ca

Plus précisément, nos résultats montrent que les pratiques d'évaluation gagnent en cohérence, mais que les pratiques de notation rencontrent toujours des obstacles. La LE s'inscrit ici dans une épistémologie située et non technique. Elle montre toutefois la nécessité d'élargir encore l'analyse des négociations menées par les enseignants dans leurs pratiques évaluatives pour mieux en comprendre les caractéristiques.

Mots clés : littératie en évaluation, compétences en évaluation, évaluation sommative, alignement curriculaire élargi

Abstract

Our research examines the extent to which assessment literacy (AL), which is little developed in the French-speaking research community, appears to be a useful tool for documenting and understanding the development of assessment skills among six secondary school mathematics and French teachers in French-speaking Switzerland. Resulting from a training process promoting exchanges and the creation of summative tests by exploiting specific theoretical contributions, the analysis of our data shows the dynamic development of teachers' skills, and through that of their AL, using a methodology exploiting the dimensions identified in contemporary literature. More specifically, our results highlight that assessment practices are becoming more consistent, but that grading practices still face obstacles. AL conceptualized in this way is part of a situated, non-technical epistemology. However, it shows the need to further expand the analysis of the negotiations carried out by teachers in their assessment practices to better understand their characteristics.

Keywords: assessment literacy, assessment skills, summative assessment, expanded curricular alignment

Problématique

À l'heure où le lecteur s'empare de ce texte, la majorité des systèmes éducatifs de l'Organisation de coopération et de développement économiques (OCDE) travaillent à la mise en œuvre d'une évaluation au service des apprentissages de tous les élèves, de manière à lutter contre l'échec scolaire et à favoriser les progrès de chacun. Cette acception de l'évaluation se concrétise dans le concept d'*Assessment for Learning* (AfL) (Assessment Reform Group, 2002), ou Évaluation-soutien d'Apprentissage (EsA) (Allal et Laveault, 2009), défini comme une démarche d'évaluation faisant « partie des pratiques quotidiennes des élèves et des enseignants qui, individuellement et en interaction, recherchent, réfléchissent sur et réagissent à l'information provenant d'échanges, démonstrations et observations afin de favoriser les apprentissages en cours » (Allal et Laveault, 2009, p. 102).

Pensée depuis les travaux fondateurs de Gipps (1994) et développée dans des réseaux internationaux et nationaux rassemblant chercheurs et acteurs du politique[1], la mise en œuvre de cette évaluation inclusive (Klenowski et Wyatt-Smith, 2014) s'intensifie depuis peu. Elle a de ce fait non seulement un impact sur les institutions de formation, qui doivent élaborer leurs contenus à l'aune de nouvelles théorisations de l'évaluation, mais elle conditionne aussi les attentes que les systèmes éducatifs développent envers le corps enseignant. La question se pose alors de savoir quelles connaissances et compétences les enseignants devraient idéalement posséder ou développer pour être des évaluateurs centrés sur les apprentissages et les progrès des élèves.

Cette question n'est toutefois pas nouvelle et a été abordée par les recherches menées dans le domaine de l'évaluation (p. ex., Brookhart, 2011; Popham, 2009). Mais nous pensons également aux travaux ayant étudié l'impact des méthodes de recherche collaborative (p. ex., Bednarz, 2013, 2015), ceux qui ont examiné les forces et limites des communautés d'apprentissage professionnel (p. ex., Dionne et al., 2010; Dogan et al., 2016), ou aux études sur le développement du jugement professionnel en évaluation (p. ex., Mottier Lopez et Tessaro, 2016; Mottier Lopez et Morales Villabona, 2018).

Malgré ces apports, il reste toujours délicat de délimiter le « portrait type » d'un enseignant-évaluateur aux pratiques justes (Nisbet et Shaw, 2020), cohérentes (Biggs, 2003; Pasquini, 2019, 2021) ou valides (Bonner, 2013). En effet, un nombre

1 Nous pensons, par exemple, à l'International Educational Assessment Network (IEAN).

conséquent de travaux ont montré que les pratiques évaluatives sont le fruit de négociations complexes que mène l'enseignant entre de multiples facteurs comme ses valeurs, ses croyances, ses connaissances et ses expériences, mais aussi son contexte, son public d'élèves, les normes légales qui cadrent ses pratiques, ou encore les cultures d'établissement ou les modes de leadership de ses équipes et de son école (Fulmer et al., 2015; Timperley et al., 2007; Xu et Brown, 2016). Ces négociations permanentes sont éminemment contextuelles et donc situées (DeLuca et al., 2019; Pasquini et DeLuca, 2021; Willis et al., 2013). Relativement aux pratiques évaluatives sommatives dont il est question dans cet article, d'autres études ont pointé le fait qu'elles sont le plus souvent le résultat d'essais-erreurs que de pratiques modélisées (Braxmeyer et al., 2004; Moss, 2013), et que les pratiques de notation qui y sont articulées ressemblent davantage à un fatras qu'à l'expression d'une forme de cohérence (Cross et Frary, 1999; Pasquini, 2021; Tierney et al., 2011).

En lien avec cette épistémologie située, nous postulons que la littératie en évaluation (LE)[2] est pertinente pour comprendre et décrire l'appropriation de connaissances et leur complexe mise en œuvre sous forme de compétences. Les travaux en LE se développent intensément dans la communauté de recherche anglophone en évaluation depuis les années 1990. En 2017, DeLuca et Johnson affirment qu'« une littératie en évaluation est une condition *sine qua non* pour un enseignant compétent aujourd'hui » (p. 121, trad. libre). Ainsi, ils soulignent la nécessité d'approfondir les concepts de compétences et de connaissances en évaluation, et de réfléchir aux conditions nécessaires pour que la LE permette d'accompagner le développement des pratiques évaluatives dans une épistémologie située.

Cet article s'empare de cette intention en posant deux hypothèses :

1. Le développement de compétences en évaluation est notamment rendu possible par l'appropriation de connaissances théoriques, qui sont incontournables au développement d'une littératie en évaluation.
2. Cette littératie en évaluation ne se réduit pas à une perspective technique visant à rendre compte d'acquis, mais constitue le fruit de négociations situées par l'enseignant en fonction de sa discipline et de son contexte.

2 Volante et Fazio (2007), au Canada, ont traduit *Assessment Literacy* en français par « capacité d'évaluation ». Nous préférons ici en adopter une traduction littérale, pour éviter des malentendus, le terme « capacité » étant polysémique dans la recherche.

Au regard de ces éléments de problématisation, notre question de recherche est la suivante : « Dans quelle mesure le développement de compétences en évaluation — rendu possible par l'appropriation de concepts théoriques — permet-il aux enseignants de résoudre des problèmes qu'ils identifient dans leurs pratiques évaluatives et d'ainsi rendre compte d'un développement d'une littératie en évaluation ? »

Pour la traiter, nous commençons par dresser notre cadre conceptuel en y définissant les éléments théoriques clés de notre étude. La méthodologie présente ensuite le corpus d'enseignants avec qui nous avons travaillé, les particularités de notre dispositif, ainsi que nos données et outils d'analyse. Les résultats sont organisés de manière à mettre en évidence l'évolution des compétences des enseignants et le rapport avec le développement de leur LE dans une perspective située (DeLuca et al., 2019; Willis, 2013). La discussion relève dans la foulée quelques questions et tensions en vue de futures problématisations. Enfin, nous terminons avec une conclusion mettant notamment en évidence les atouts et les limites du présent travail, ainsi que des pistes pour de futures recherches.

Cadre conceptuel

L'évaluation sommative

Comme Brookhart (2017), nous considérons que « [l']évaluation sommative est effectuée après l'apprentissage pour certifier ce qui a été appris » (p. 8, trad. libre). Une telle démarche relève d'un bilan devant porter sur les apprentissages essentiels qui s'accompagne d'une note chiffrée (sur 6, 10, 100, ou en pourcentage), d'une cote lettrée (A, B, C, etc.) ou de tout autre type d'appréciation. L'évaluation sommative reste, dans la majorité des systèmes éducatifs, la démarche d'évaluation que le corps enseignant privilégie et que les établissements d'enseignement imposent (Klenowski et Wyatt-Smith, 2014; OCDE, 2012). Des travaux s'accordent également sur le fait que la note représente l'information la plus concrète que les élèves puissent obtenir sur les bilans de leurs apprentissages (Guskey et Link, 2019). Il y a donc une nécessité d'étudier spécifiquement ces pratiques. En effet, bien que potentiellement statique, la dimension formelle de l'évaluation sommative notée ne devrait pas exclure la possibilité de faire des rétroactions

aux élèves (Taras, 2005), puisque c'est l'utilisation faite des informations issues de la démarche d'évaluation qui en détermine la fonction, et non l'inverse (Good, 2011; Wiliam, 2000). Dans ce sens, nous postulons que toute démarche d'évaluation sommative est un processus complexe centré sur une prise d'informations pouvant également fournir des rétroactions aux élèves, et ainsi poursuivre conjointement une fonction de soutien à l'apprentissage (Harlen, 2005; Laveault et Allal, 2016), dans une logique de progression.

L'alignement curriculaire élargi

Pour favoriser le développement des compétences en évaluation des enseignants participant à notre étude, et par la même occasion celui de leur LE, nous avons exploité avec eux le modèle théorique de l'alignement curriculaire élargi (Pasquini, 2019, 2021) qui reprend les caractéristiques principales de l'alignement curriculaire (Anderson, 2002; Gauthier et al., 2005), reconnu comme un indice de pratiques évaluatives cohérentes (Biggs, 2003; Martone et Sireci, 2009) participant significativement au développement de la LE chez les enseignants (Klenowski et Wyatt-Smith, 2014). Le modèle est ici élargi, car il intègre deux composantes spécifiques aux pratiques évaluatives lors d'épreuves sommatives : la pondération (le fait d'attribuer des points sur la base de critères) et la notation (le fait de définir des seuils pour la construction des notes). Aussi, l'alignement curriculaire élargi désigne, dans une épreuve d'évaluation sommative, « le fort lien de cohérence systémique existant entre les objectifs évalués, les tâches évaluatives, les points et/ou les critères attribués et l'échelle de notation » (Pasquini, 2019, p. 83–84). Les relations de cohérence sont ici lues à l'aide de la taxonomie d'Anderson et Krathwohl (2001). Par exemple, en lien avec les processus d'enseignement et d'apprentissage prévus ou effectifs, nous avons travaillé à comprendre avec les enseignants dans quelle mesure l'objectif d'apprentissage en mathématiques « résoudre des problèmes géométriques » se retrouvait dans les tâches de l'épreuve sommative, mais était également critérié pour établir les différents seuils des notes (de 1 à 6). Le questionnement permis par notre modèle nous a amenés à parler avec les enseignants de cohérence curriculaire élargie de l'épreuve, en cas de cohérence globale, et de désalignements curriculaires (Airasian et Miranda, 2002; Biggs, 2003) quand cette cohérence se perdait, par exemple, lorsque le seuil des notes était déterminé à l'aide de barèmes standardisés sans référence aux apprentissages évalués.

Les compétences retenues dans le cadre de notre recherche

Les connaissances sur lesquelles nous nous sommes concentrés étaient issues des composantes du modèle de l'alignement curriculaire élargi exploité avec les enseignants (Pasquini, 2019, 2020) : rapport au plan d'études, nature du contenu, design des tâches évaluatives, pratiques de pondération, pratiques de notation. Nous avons postulé que le développement des compétences relevait ensuite d'une mobilisation de ces connaissances en contexte, en vue de résoudre des situations complexes et de répondre aux problèmes identifiés par les enseignants (Paquay et al., 2010).

Conformément à la littérature, nous avons également postulé que le développement des compétences évaluatives des enseignants et leur capacité à justifier leurs décisions en fonction de leur discipline, des caractéristiques des classes et du contexte large, seraient des signes de développement de leur LE. Dès lors, en référence aux différentes composantes de l'alignement curriculaire élargi, et en articulant les apports de la recherche sur les compétences clés en évaluation (Brookhart, 2011; Popham, 2009) avec les spécificités contextuelles des enseignants, nous avons formulé six compétences (Pasquini, 2018, 2020) :

1. Analyser le plan d'études du point de vue des objectifs et des contenus disciplinaires qui y circulent à des fins d'évaluation sommative, en mettant en relation de manière cohérente les différents niveaux de formalisation des objectifs, et en analysant les contenus pour en déterminer les caractéristiques essentielles à des fins d'évaluation sommative.
2. Élaborer des tâches évaluatives complexes en rapport avec les objectifs des plans d'études qui donnent des preuves d'apprentissages maîtrisés ou en voie de réalisation, en privilégiant des tâches qui donnent à voir des unités d'apprentissage significatives (approche qualitative).
3. Adopter des systèmes de pondération permettant de mettre en évidence le poids donné à l'apprentissage se référant aux objectifs et aux contenus, en privilégiant, autant que possible, une pondération critériée, notamment pour permettre de la rétroaction aux élèves dans une visée de régulation.
4. Construire des notes de manière à ce qu'elles rendent compte des acquis et lacunes des apprentissages des élèves, en élaborant ses propres barèmes

en référence au système de pondération, à la complexité des tâches et aux objectifs évalués.

5. Penser ses pratiques évaluatives sommatives en matière de validité et de cohérence d'alignement curriculaire élargi, en les référant aux processus d'enseignement et d'apprentissage, et en analysant leur cohérence à l'aune d'éléments taxonomiques.
6. Justifier ses choix en matière de pratiques évaluatives sommatives au regard de son contexte institutionnel, cantonal, local (établissement), et de la culture des équipes dans lesquelles elles s'inscrivent, en les adaptant aux prescriptions institutionnelles tout en étant attentif à ce qu'elles soutiennent les apprentissages des élèves.

La littératie en évaluation

On traduit généralement *literacy* par « littératie ». L'OCDE la définit comme « l'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités » (2000, p. x [Introduction]). Cette définition générique souligne que tout individu lettré devrait maîtriser des connaissances (en langue, dans la définition de départ), les utiliser en situation, exploiter ses acquis pour atteindre certains buts, en fonction des contextes dans lesquels ils sont poursuivis, et mettre à profit ses acquis pour le développement de connaissances et de compétences futures. Exploité dans divers champs sociaux et professionnels — en littératie numérique (pour comprendre, utiliser et créer des outils et médias numériques), en littératie en santé (pour comprendre et utiliser des informations médicales afin de prendre des décisions de traitements) ou encore, et de manière non exhaustive, en littératie visuelle (pour comprendre et utiliser tous les symboles autres que l'écriture) —, ce concept prend aujourd'hui de l'ampleur dans le domaine de l'évaluation.

Stiggins (1995) définissait la LE comme « la capacité à faire la différence entre une évaluation fondée et une évaluation non fondée » (p. 240, trad. libre). Inscrite dans une épistémologie psychométrique et axée sur la maîtrise de connaissances, cette définition a inspiré plusieurs chercheurs qui ont tenté de décrire les connaissances et les compétences que tout enseignant devrait posséder pour être un « bon » évaluateur

(p. ex., Mertler, 2004; Popham, 2004; Volante et Fazio, 2007). Cette vision relativement techniciste et statique a ensuite été étendue en s'appuyant sur les avancées de la recherche sur les pratiques évaluatives et les théories de l'apprentissage, en donnant une plus grande attention aux aspects sociaux, contextuels et théoriques en jeu dans les pratiques d'évaluation.

Ainsi, les conceptions contemporaines de la LE la considèrent comme un espace professionnel dans lequel les enseignants font interagir leurs connaissances de l'évaluation avec leurs connaissances de la pédagogie, du contenu et du contexte d'apprentissage. Cette conception de la LE se définit alors comme

> une pratique sociale dynamique dépendant du contexte qui implique que les enseignants articulent et négocient les connaissances scolaires et culturelles entre eux et avec les élèves, dans l'initiation, le développement et la pratique de l'évaluation pour atteindre les objectifs d'apprentissage. (Willis et al., 2013, p. 242, trad. libre)

L'épistémologie située de la LE se retrouve plus précisément à deux niveaux qui sont au cœur de notre étude :

1. Les pratiques évaluatives s'appréhendent dans des contextes spécifiques de niveaux micro (classes, élèves), méso (écoles, équipes pédagogiques) et macro (directives légales, relations aux parents) qui interagissent entre eux (Fulmer et al., 2015).
2. Les démarches d'évaluation s'inscrivent dans les processus d'enseignement et d'apprentissage, eux aussi toujours contextualisés, notamment du point de vue des disciplines (Anderson, 2002; Pasquini, 2019).

Au cœur de cette épistémologie située, Xu et Brown (2016), en s'appuyant sur de multiples sources, identifient plusieurs connaissances indispensables à un développement de la LE qui ont du sens dans le cadre de notre recherche portant sur les pratiques d'évaluation sommative. Ce sont notamment des connaissances :

- disciplinaires sur les contenus des curricula;
- des buts, méthodes et contenus des évaluations, notamment en ce qui concerne les tâches évaluatives;
- en pondération (exploitation de divers systèmes critériés) et en notation;

- en rétroaction;
- au niveau de l'interprétation et de la communication des résultats.

Ces connaissances participent au développement de compétences évaluatives, car elles vont en être les descripteurs. Mais elles nécessitent d'être contextualisées et perpétuellement négociées à la lumière de dimensions plus larges. Dans cette perspective, et comme nous le détaillerons plus loin, la formation à la LE se doit de valoriser les débats entre pairs au travers de confrontations de pratiques évaluatives articulant, par exemple, la création et l'analyse de documents pour la classe (dans notre cas, des épreuves sommatives), afin de « donner aux participants la possibilité de problématiser de manière critique et de recontextualiser les normes des communautés d'évaluation, de comprendre les besoins en matière d'évaluation » (Willis et al., 2013, p. 252, trad. libre) pour, *in fine*, soutenir les apprentissages des élèves.

Méthodologie

Le corpus d'enseignants

Nous avons travaillé avec trois enseignants de mathématiques et trois enseignantes de français volontaires, du secondaire (élèves de 13 à 16 ans)[3], en privilégiant une perspective collaborative (Pasquini, 2020). Les enseignantes de français travaillaient dans le même établissement. Deux enseignants de mathématiques exerçaient dans un établissement différent, et le troisième provenait d'une autre école présentant des caractéristiques socio-économiques similaires à celle de ses collègues de même discipline. Leur expérience variait de trois à quinze ans, mais personne n'avait suivi de formation continue récente portant sur l'évaluation sommative. Du point de vue prescriptif, ils étaient légalement tenus d'évaluer formellement leurs élèves au travers d'épreuves sommatives complexes et de les certifier à l'aide de notes allant de 1 à 6, avec demi-notes, le 4 traduisant le seuil de suffisance.

3 Par souci d'anonymat, des pseudonymes ont été attribués.

Le dispositif

Lorsqu'il s'agit de caractériser les dispositifs qui participent à développer une LE, la littérature s'accorde sur plusieurs points : les enseignants se forment avec leurs collègues d'école (Willis et al., 2013), afin de partager un langage commun et de négocier certains choix évaluatifs (Fleer, 2015; Lukin et al., 2004); les actions de formation portent sur la création de critères (Harlen, 2005), le *design* de tâches évaluatives (Adie, 2013; Laveault et Allal, 2016) ou encore la création d'outils de notation (Brookhart, 2017; Pasquini, 2021); un temps conséquent est alloué à échanger autour des processus de création de démarches, afin que le développement des compétences puisse se réaliser de manière progressive et s'ancrer dans des expériences de pratiques en classe (Guskey et Yoon, 2009). Autant de caractéristiques que nous avons respectées dans notre dispositif décrit dans le Tableau 1.

Plus précisément, les temps 1 et 8 ont donné lieu à des entretiens compréhensifs (Kaufmann, 2011) individuels permettant de colliger les descriptions des participants concernant leurs pratiques évaluatives sommatives en contexte, avant et après la recherche. Les temps 2, 4 et 7 ont concrétisé trois actes d'écriture au travers de la production d'un récit exemplaire de pratique (Desgagné, 2005; Pasquini, 2013, 2016) : au début de la recherche (pour faire état de leur questionnement), pendant la recherche (pour expliciter l'évolution de leur questionnement et de leurs acquisitions) et à la fin de celle-ci (pour dresser un bilan des changements et des questions toujours en suspens). Lors d'une première rencontre commune (temps 3), les enseignants ont analysé une épreuve sommative inconnue à l'aide du modèle théorique de l'alignement curriculaire élargi (Pasquini, 2018, 2019), puis ils ont commencé à élaborer une épreuve pour leurs élèves, conformément à leur contexte, individuellement ou avec leurs collègues directs. La deuxième rencontre (temps 5) a eu lieu six semaines plus tard, une fois entre spécialistes de français, puis entre spécialistes de mathématiques, afin d'approfondir les questions liées aux contenus des deux disciplines. Ce laps de temps a permis aux enseignants de mettre en œuvre leur démarche en classe. Durant cette deuxième séance, le chercheur et les enseignants ont discuté de la première épreuve, puis ces derniers ont commencé à construire une seconde épreuve. Deux mois plus tard, au temps 6, une dernière rencontre a eu lieu en grand groupe, durant laquelle les participants ont réitéré le processus de discussion de chaque épreuve. L'entièreté du dispositif a duré sept mois. Quatre types de données ont permis d'analyser le développement des compétences de chacun : les épreuves sommatives; les enregistrements des échanges, pendant les rencontres, lors des discussions concernant ces épreuves; les entretiens compréhensifs; et les écrits exemplaires de pratique.

Tableau 1

Dispositif de récolte de données

Temps	1	2	3	4	5	6	7	8
Séances	Entretiens individuels 1	Récit de pratique 1	Rencontre 1 commune	Récit de pratique 2	Rencontre 2 différenciée mathématiques et français	Rencontre 3 commune	Récit de pratique 3	Entretiens individuels 2
Données	Transcription des entretiens	1er écrit sur les pratiques	1re épreuve sommative créée par les enseignants Enregistrement et transcription des séances	2e écrit sur les pratiques	Enregistrement et transcription des discussions des épreuves sommatives réalisées 2e épreuve sommative créée par les enseignants	Enregistrement et transcription des discussions des épreuves sommatives réalisées	3e écrit sur les pratiques	Transcription des entretiens

Les données pour décrire le développement de la LE

Dans un précédent article (Pasquini, 2020), nous avons explicité le processus de développement des compétences de chaque enseignant en fonction des disciplines, des contextes et de leurs questions de départ. Nous avons décrit la dimension dynamique de ce développement, soulignant notamment certaines tensions entre les pratiques d'évaluation et les pratiques de notation, sur lesquelles nous reviendrons. C'est donc à partir de ces résultats que nous avons examiné un second développement de la LE pour chaque enseignant.

La méthode d'analyse

C'est dans un premier temps en nous inspirant des travaux de Klinger et ses collègues (2015) que nous avons pensé un outil d'analyse répondant à une épistémologie située de la LE. En articulant des apports issus de la recherche avec des dimensions contextuelles, ces chercheurs ont catégorisé 19 facteurs influençant les pratiques évaluatives en trois processus évaluatifs clés permettant aux enseignants d'exercer un jugement professionnel pour une évaluation sommative cohérente au service des apprentissages de tous les élèves. Ces processus sont les suivants : les *fondements* relèvent la capacité chez l'enseignant à identifier les finalités et les fonctions de l'évaluation; la *mise en pratique* se lit au travers des liens que fait l'enseignant entre les objectifs d'apprentissage évalués, la conception des épreuves, l'analyse des travaux des élèves, la formulation de rétroactions et l'établissement des notes; et la *qualité* correspond au souci qu'a l'enseignant de prendre des décisions jugées justes, en contrant les éventuels biais et en tenant compte des diversités des élèves.

Nous avons ensuite suivi les apports de DeLuca et ses collègues (2019) qui croisent ces trois processus comme autant de nouvelles catégories avec les travaux en LE de Willis et ses collègues (2013) et ceux de Tierney (2006), afin de souligner l'importance des connaissances théoriques que l'enseignant met au service de ses pratiques. Ainsi, DeLuca et ses collègues (2019) modélisent quatre dimensions inhérentes à une LE située : a) le but de la démarche d'évaluation est clair (l'enseignant explique en quoi elle peut soutenir les apprentissages); b) l'exploitation de la théorie dans les pratiques est au service de leur cohérence (l'enseignant peut expliciter cette dernière); c) la justesse des pratiques est justifiée (l'enseignant l'exprime au regard des attentes

relatives aux apprentissages); d) les décisions prises sont assumées (l'enseignant accepte les compromis ou dilemmes éventuels jusqu'aux processus de notation). Ce sont ces quatre dimensions, en cohérence avec les définitions contemporaines d'une LE située, qui nous ont servi à rendre compte du développement des compétences en évaluation des six enseignants et de leur LE.

Résultats

Le développement de la LE prenant appui sur celui des compétences, nous présentons les résultats en deux temps. Tout d'abord, pour chaque enseignant, un tableau rend compte de l'évolution de ses compétences précédemment énoncées (Pasquini, 2020) (Tableaux 2, 4, 6, 8, 10 et 12 ci-après). Des extraits de verbatim, signalés entre guillemets dans les tableaux, exemplifient les éléments significatifs de ce développement. Les codes E1 et E2 symbolisent les contenus des épreuves qu'ils ont créées. Puisque les compétences E et F portaient sur la mobilisation des apports proposés, les indices de leur développement n'apparaissent que dans la seconde partie du tableau. Une brève synthèse suit chaque tableau dans une perspective interprétative. Dans un second temps, nous présentons en quoi un développement d'une LE a eu lieu à la lumière de nos quatre dimensions issues de la littérature, également à l'aide d'un tableau par enseignant, en nous référant aux données récoltées en fin de recherche, notamment au travers d'extraits de verbatim, également cités entre guillemets.

Les enseignants de mathématiques

Fernand

À la fin de la formation, Fernand exploite davantage le plan d'études, ce qui lui permet de penser la cohérence de ses épreuves jusqu'aux critères. À ce titre, il peut justifier la cohérence de ses épreuves à l'aide d'apports théoriques, même s'il rencontre encore des difficultés dans ses pratiques de notation (voir Tableau 2).

Le Tableau 3 présente quelques indices clés du développement de sa LE.

Tableau 2

Évolution des compétences en évaluation chez Fernand

Fernand	Compétence A	Compétence B	Compétence C	Compétence D	Compétence E	Compétence F
Avant la recherche	Utilisation du plan d'études comme balise, parfois pas pris en compte pour l'évaluation, car il n'a « pas de sens »	Élaboration des tâches évaluatives en termes de situations qui « sollicitent l'intelligence des élèves et les outils qu'ils ont acquis »	Pondération très difficile : « est-ce correct d'opérer en fonction des gestes mathématiques »? Pourcentages de points jugés « harmonieux, naturels, mathématiques »	Construction de la note réalisée à l'aide du barème fédéral (barème standardisé) pour toutes les épreuves	-	-
Après la recherche E1 : Géométrie « Résolution de problèmes dans le plan et l'espace » E2 : Géométrie « Aires et périmètres »	Lecture du plan d'études au travers des contenus disciplinaires et des objectifs	Élaboration de tâches évaluatives complexes en lien avec les contenus disciplinaires	Attribution des points en référence aux contenus issus du plan d'études et traduits depuis des critères portant sur des contenus disciplinaires	Construction de la note en rapport avec les contenus du plan d'études, issue d'un raisonnement se référant aux unités d'apprentissages pour l'élaboration du seuil du 4. Puis, utilisation d'un barème standardisé pour les autres notes	Désalignements curriculaires encore difficiles à gérer au niveau de la notation	Choix le plus souvent assumés, même si le système critérié et une note se référant à l'apprentissage sont considérés comme incompatibles

Tableau 3

Indices clés du développement de la LE de Fernand

Dimensions de la LE (Fernand)	L'enseignant explique en quoi ses pratiques peuvent soutenir les apprentissages	L'enseignant explique en quoi la théorie sur l'évaluation (alignement curriculaire élargi) est au service de la cohérence de ses pratiques	L'enseignant exprime la justesse de ses pratiques au regard des attentes relatives aux apprentissages	L'enseignant accepte les compromis et/ou dilemmes éventuels jusqu'aux processus de notation
Indices de développement de la LE	Les démarches d'évaluation se centrent sur une recherche de qualité dans les épreuves : *« la volonté d›intégrer cette composante qualitative dans le processus a modifié la qualité de mon enseignement et m›a permis de lâcher certaines habitudes comme celle d›évaluer systématiquement les algorithmes de calcul dans tous les tests ».* Il fait ainsi un lien entre démarches évaluatives sommatives et enseignement	Le rapport au modèle est situé, permet d'opérationnaliser la recherche de cohérence : « *avec l'alignement curriculaire élargi, je suis conscient si l'objet d'étude fera partie de l'évaluation sommative ou si ce n'est qu'un élément pour donner plus de sens à ce que font les élèves; à chaque moment, je savais ainsi que mes élèves, ce qui serait évalué et ce qui ne le serait pas* »	Cette question d'être juste renvoie aux élèves. Fernand explique que « *l'épreuve doit évaluer la qualité, qu'on arrive à évaluer ce que l'élève sait faire. Je veux dire qu'en regardant, qu'à travers cette évaluation et la production de l'élève, j'ai des éléments qui m'indiquent clairement où il en est par rapport au thème* »	Fernand peut expliquer comment il détermine la note suffisante, mais reste en difficulté pour les autres : « *la contrainte, c'est des échelons en dessus et au-dessous… […] décrire chaque note est subjectif* »

En lien avec le développement de ses compétences, trois dimensions montrent que des indices de développement d'une LE sont bien présents chez Fernand, mettant au centre de ses réflexions et de ses pratiques un souci de privilégier des démarches évaluatives qualitatives, centrées sur les contenus disciplinaires clés et faisant sens pour les élèves. Toutefois, des compromis sont présents au niveau de la notation, soulignant une difficulté à gérer un côté subjectif pour l'établissement des notes.

Paul

Les pratiques évaluatives de Paul se réfèrent de plus en plus aux contenus disciplinaires et aux objectifs : l'utilisation du plan d'études a gagné en clarté; l'élaboration des tâches évaluatives, l'attribution des points et la construction de la note suivent davantage une réflexion centrée sur les apprentissages (voir Tableau 4).

Paul fait toutefois toujours face à des espaces d'inconfort au niveau de la construction de la note où il réalise des compromis peu convaincants à ses yeux, comme le fait d'utiliser toujours des barèmes standardisés. Le Tableau 5 présente le développement de sa LE.

Une capacité accrue à analyser et penser les phénomènes de cohérence et d'incohérence dans les démarches d'évaluation sommative sont des indices de développement d'une LE chez Paul. En revanche, les pratiques de notation n'ont pas évolué et restent difficiles à justifier, dans la mesure où elles se réfèrent à des outils standardisés et sont dès lors peu en lien avec les apprentissages. Là, nous observons les limites que rencontre Paul dans les négociations qu'il mène entre les contraintes liées aux outils et les caractéristiques des objets disciplinaires évalués.

Tableau 4

Évolution des compétences en évaluation chez Paul

Paul	Compétence A	Compétence B	Compétence C	Compétence D	Compétence E	Compétence F
Avant la recherche	Confusion dans la compréhension des différents objectifs du plan d'études et de leur rôle	Élaboration de toute épreuve avec 20-40% d'exercices techniques et le reste de problèmes	Attribution forte des points sur les exercices techniques La note 4 est atteignable si la technique est juste	Construction de la note à l'aide du barème fédéral (barème standardisé) pour toutes les épreuves Rôle important de la répartition des scores et de la moyenne Possibilité de « bouger les points » en cas de moyenne basse	-	-
Après la recherche E1 : Fonctions et algèbre « Résolution de problèmes » E2 : Calcul littéral « Résolution de problèmes »	Mise en lien des contenus disciplinaires avec les objectifs du plan d'études, dont la compréhension est parfois encore problématique	Élaboration de tâches évaluatives traduisant les contenus clés de la discipline, mais leur complexité prête encore à discussion	Attribution des points le plus souvent en référence aux contenus issus des objectifs du plan d'études	Construction de la note en lien avec la complexité des contenus disciplinaires, mais réalisée à l'aide d'un barème standardisé qui autorise les compensations et les désalignements curriculaires	Désalignements curriculaires omniprésents au niveau de la pondération et de la notation	Rapport critique au référentiel, choix des tâches assumés Utilisation de critères jugée impossible Notation sur la base de compromis acceptables, mais peu convaincants

Tableau 5

Indices clés du développement de la LE de Paul

Dimensions de la LE (Paul)	L'enseignant explique en quoi ses pratiques peuvent soutenir les apprentissages	L'enseignant explique en quoi la théorie sur l'évaluation (alignement curriculaire élargi) est au service de la cohérence de ses pratiques	L'enseignant exprime la justesse de ses pratiques au regard des attentes relatives aux apprentissages	L'enseignant accepte les compromis et/ou dilemmes éventuels jusqu'aux processus de notation
Indices de développement de la LE	Les épreuves sommatives peuvent donner lieu à des rétroactions ciblées sur des contenus à travailler : « *si je mets le doigt sur ce qui n'a pas été, je peux dire à l'élève " tu as perdu des points parce que tu ne sais pas factoriser " […] je pense que c'est très intéressant d'utiliser les évaluations autrement que juste pour mettre des notes* »	L'alignement curriculaire élargi permet de comprendre où se situent les désalignements fréquents : « *ça m'a aidé à plus prendre conscience que le fait de choisir une échelle peut former une cassure entre tout le travail que je me suis donné de la peine à faire avant* »	Des critères de justesse sont explicités : « *la démarche d'évaluation est cohérente par rapport à l'enseignement, au plan d'études, elle ne pénalise pas les élèves, elle leur laisse du temps […] et la distribution des scores est un indice de la qualité de mon enseignement* »	Des compromis peu satisfaisants sont réalisés au niveau de la notation : « *Je n'ai pas changé, parce que je n'ai pas trouvé de façon de faire autre qui soit satisfaisante à mes yeux, qui soit à la fois applicable, qui ne me pose pas de problème […] J'utilise le barème fédéral… encore… c'est ma base de travail […]. Je ne pense pas que c'est bien, mais il faut bien que je travaille avec quelque chose* »

Philémon

Au terme de la formation, le contenu du plan d'études détermine les attentes, l'élaboration de tâches évaluatives de complexité plus homogène et la construction du seuil de la note 4 sont rendus possibles par une réflexion sur les critères (voir Tableau 6).

Il subsiste toutefois encore des problèmes à construire les autres notes en référence à ces mêmes critères, Philémon étant toujours sceptique quant à la plus-value de pondérer sans points. Sur ce plan, il rencontre des difficultés à négocier des éléments formels (le nombre d'échelons de notation) avec leur qualification en matière d'apprentissage. Cela étant, il assume ses choix en matière d'évaluation. Le Tableau 7 présente le développement de sa LE.

Les réflexions de Philémon sur ses démarches évaluatives sont davantage modélisées et mises en lien avec les processus d'enseignement. Plusieurs indices de développement d'une LE chez lui sont ainsi présents, comme le fait de poser l'échelle de notation en amont de la passation de l'épreuve.

Tableau 6

Évolution des compétences en évaluation chez Philémon

Philémon	Compétence A	Compétence B	Compétence C	Compétence D	Compétence E	Compétence F
Avant la recherche	Considération du plan d'études comme un fil rouge insuffisamment exploité pour aller au-delà de messages à l'élève du type : « t'as fait 4,5-5 »	Élaboration de toute épreuve avec deux tiers d'exercices techniques et le reste de problèmes	Attribution de deux tiers des points sur la technique et un tiers à 50% des points sur les problèmes Attribution des points pouvant « jouer les vases communicants » en permettant la compensation	Construction du seuil de la note 4 en fonction des attentes relatives aux différents contenus évalués Échelles linéaires. Seuil des notes fixé en amont, mais pouvant bouger après correction	-	-
Après la recherche E1 : Géométrie « Résolution de problèmes dans le plan et l'espace » E2 : Géométrie « Aires et périmètres »	Lecture du plan d'études au travers des contenus disciplinaires et des objectifs	Élaboration de tâches évaluatives complexes en lien avec les contenus disciplinaires	Attribution des points en référence aux objectifs issus du plan d'études et traduits depuis des critères portant sur des contenus disciplinaires	Construction de la note 4 en rapport avec les contenus du plan d'études, issue d'un raisonnement se référant aux apprentissages Utilisation d'un barème standardisé pour les autres notes	Désalignements curriculaires encore présents et problématiques au niveau de la notation	Choix le plus souvent assumés, même si le système critérié et une note se référant à l'apprentissage sont considérés comme peu compatibles

Tableau 7

Indices clés du développement de la LE de Philémon

Dimensions de la LE (Philémon)	L'enseignant explique en quoi ses pratiques peuvent soutenir les apprentissages	L'enseignant explique en quoi la théorie sur l'évaluation (alignement curriculaire élargi) est au service de la cohérence de ses pratiques	L'enseignant exprime la justesse de ses pratiques au regard des attentes relatives aux apprentissages	L'enseignant accepte les compromis et/ou dilemmes éventuels jusqu'aux processus de notation
Indices de	Les épreuves sommatives peuvent donner lieu à des rétroactions ciblées sur les contenus encore en construction : « *je pense que c'est un apport durable dans la construction de mes évaluations, cette page de rétroaction qui passe par un tableau dans lequel les critères sont en lien avec les objectifs, l'enseignement, et la pondération* »; elles portent sur des apprentissages clés pour « *ne pas vouloir être dans le panachage à tout prix* »	L'alignement curriculaire élargi est un outil qui permet de comprendre la cohérence des épreuves sommatives : « *cet alignement curriculaire […] est une explication théorique qui met en évidence que, quand il y a rupture, quelque chose n'est pas cohérent entre ce qui a été prévu, pensé, et l'enseignement dispensé* » Le modèle permet aussi de justifier certaines approches : « *l'alignement curriculaire permet de me reposer sur un fondement théorique qui pourrait dire aux collègues "écoute, là non, ça ne va pas…"* »	Le rapport entre enseignement et évaluation est ici premier, notamment du point de vue des contenus fondamentaux que les élèves doivent s'approprier : « *une évaluation qui est alignée avec l'enseignement, et effectivement, la résultante de la construction de la note, en cohérence avec ce qui est plus ambitieux, ce qui est plus difficile à saisir, ce qui est plus enseigné* »	Malgré des difficultés à mettre en lien les notes avec l'apprentissage évalué, les échelles de notation sont construites en amont de la passation de l'épreuve : « *construire l'échelle en amont […] c'est un changement […] dans le sens où je me dis "qu'est-ce qui est utile pour moi ? À partir de quand est-ce que je peux dire que c'est réussi ? etc."* »

Les enseignantes de français

Alizée

À la fin de la formation, les pratiques évaluatives d'Alizée donnent plus de place aux contenus et aux objectifs se rapportant à la taxonomie d'Anderson et Krathwohl (2001). Son approche critériée lui permet d'assumer ses choix et de critiquer des prescriptions relatives à la notation considérées comme antipédagogiques, comme la linéarité des échelles (voir Tableau 8).

Alizée remet toutefois en question la transposition de telles pratiques à d'autres contenus, comme la compréhension de l'oral. Le Tableau 9 présente le développement de sa LE.

Globalement, les indices de développement de la LE chez Alizée soulignent une capacité à faire des choix centrés sur les apprentissages et à les expliciter. Cette explicitation lui permet notamment de souligner la complexité des pratiques évaluatives en développement, et de rendre visibles des tensions possibles, comme la conscientisation de l'impact de l'objet disciplinaire sur les possibilités de l'évaluer et de le noter à l'aide de critères.

Tableau 8

Évolution des compétences en évaluation chez Alizée

Alizée	Compétence A	Compétence B	Compétence C	Compétence D	Compétence E	Compétence F
Avant la recherche	Utilisation du plan d'études pour vérifier les contenus, par souci « d'être dans la cible » Choix dans le plan d'études des critères à évaluer (par exemple en rédaction)	Évaluation de la technique de la langue dans chaque épreuve parfois difficile à délimiter Habitude d'évaluer un objectif par exercice	Adaptation des critères et des échelles aux contenus, même si les grilles se ressemblent de fois en fois, car elles s'inspirent de grilles existantes Attribution de 50 % des points sur la technique et 50 % sur la production	Doutes sur les échelles linéaires, sur le moment de les créer et sur la variation des seuils de réussite Une importante question : « comment fixer mes échelles ? »	-	-
Après la recherche E1 : Compréhension de texte E2 : Expression écrite (le texte explicatif)	Lecture du plan d'études au travers des contenus, reformulations à l'aide d'une taxonomie comme outil analytique	Élaboration des tâches évaluatives traduisant des contenus clés de la discipline présents dans le plan d'études, le plus souvent complexes	Formulation des critères en référence aux contenus et objectifs issus du plan d'études	Construction de la note 4 et parfois de toutes les notes en rapport avec les objectifs issus du plan d'études, traduisant la complexité des contenus présents dans les tâches Accent mis sur la rétroaction donnée aux élèves	Absence de désalignements curriculaires	Choix des contenus disciplinaires et des tâches assumés et se référant à une taxonomie Adoption du système critérié Critique, voire refus des directives en matière de notation

Tableau 9

Indices clés du développement de la LE d'Alizée

Dimensions de la LE (Alizée)	L'enseignante explique en quoi ses pratiques peuvent soutenir les apprentissages	L'enseignante explique en quoi la théorie sur l'évaluation (alignement curriculaire élargi) est au service de la cohérence de ses pratiques	L'enseignante exprime la justesse de ses pratiques au regard des attentes relatives aux apprentissages	L'enseignante accepte les compromis et/ou dilemmes éventuels jusqu'aux processus de notation
Indices de développement de la LE	Il y a plus de clarté dans les finalités poursuivies : « *maintenant, ça me paraît clair, simple, enfin, […] je me dis "je vais faire un test sur quelque chose", je vois comment aller, je me sens cohérente* » Plus précisément, la démarche évaluative doit se concentrer sur l'essentiel : « *un changement, c'est de me dire que j'ai enseigné ce que je voulais, mais je ne vais peut-être pas tout tester… ou on peut faire l'année prochaine, ou… Je me suis donné, octroyé plus de libertés* »	Le modèle permet d'expliciter l'implicite : « *ça m'a rendue consciente de ce que je faisais ou pas. Au départ, je me suis dit "mais c'est tout simple, enfin…" et puis après, je me suis dit "ah mais non, ce n'est pas si facile"* » Toutefois, une cohérence a posé plus de défis : « *pour moi, le plus complexe, c'était entre les points et l'échelle* »	Le raisonnement est ici global. Il faut que l'épreuve « *teste ce que j'ai voulu transmettre à mes élèves, s'ils l'ont compris ou pas. Et que le résultat […], donc la note, puisse être un reflet de l'acquisition d'un objectif, ça je pense que c'est nouveau* »	La cohérence globale est assumée, même si sa complexité est relevée : « *maintenant, je me sens cohérente jusqu'à la note, en sachant que des fois c'est réduit à des points. Mais en me disant "j'ai fait cette réflexion, je sais pourquoi j'ai mis ça"* »

Anouck

Les changements dans les pratiques d'Anouck se situent à plusieurs niveaux : les contenus deviennent centraux dans sa manière d'envisager toute démarche d'évaluation sommative et la taxonomie est aussi une référence pour formuler et analyser les objectifs (voir Tableau 10).

Anouck arrive à éviter la compensation des points qui lui posait tant problème, tout en se questionnant encore sur la construction des notes autres que le 4 uniquement à l'aide de critères : par exemple, selon le contenu disciplinaire évalué, la distinction qualitative entre une très bonne note (5) et une note excellente (6) lui semble difficile à établir. Le Tableau 11 présente le développement de sa LE.

Les indices de développement de la LE font tous référence à une mise en évidence de la qualité de l'apprentissage évalué dans les pratiques d'Anouck. Cette qualité ne peut se concrétiser qu'au travers de raisonnements centrés sur les contenus et méthodes d'évaluation, dont la cohérence est primordiale, depuis les processus d'enseignement.

Tableau 10

Évolution des compétences en évaluation chez Anouck

Anouck	Compétence A	Compétence B	Compétence C	Compétence D	Compétence E	Compétence F
Avant la recherche	Faible utilisation du plan d'études pour construire l'évaluation : « Impression que ce sont toujours les mêmes objectifs »	Élaboration des épreuves portant toujours à la fois sur des éléments de technique de la langue et sur des compétences langagières Formulation des objectifs au travers des consignes	Insatisfaction quant aux critères produits Attribution des points aux questions ou aux critères en fonction de ce qui est attendu : « Plus c'est complexe, plus il y a de points »	Doutes sur le moment de créer l'échelle Construction de la note 4 en comptant les points montrant l'acquis de la matière, toujours entre 60 et 70 % du total Les notes : « le grand problème, du bricolage »	-	-
Après la recherche E1 : Compréhension de texte E2 : Expression écrite (le texte explicatif)	Lecture du plan d'études au travers des contenus, reformulations à l'aide d'une taxonomie comme outil analytique	Élaboration des tâches évaluatives traduisant des contenus clés de la discipline présents dans le plan d'études, le plus souvent complexes	Formulation des critères en référence aux contenus issus du plan d'études	Construction de la note 4 et parfois de toutes les notes en rapport avec les objectifs issus du plan d'études, traduisant la complexité des contenus présents dans les tâches Abandon de la logique de compensation	Absence de désalignements curriculaires	Choix des contenus disciplinaires et des tâches assumés et se référant à une taxonomie, adoption du système critérié Critique, voire refus des directives en matière de notation

Tableau 11

Indices clés du développement de la LE d'Anouck

Dimensions de la LE (Anouck)	L'enseignante explique en quoi ses pratiques peuvent soutenir les apprentissages	L'enseignante explique en quoi la théorie sur l'évaluation (alignement curriculaire élargi) est au service de la cohérence de ses pratiques	L'enseignante exprime la justesse de ses pratiques au regard des attentes relatives aux apprentissages	L'enseignante accepte les compromis et/ou dilemmes éventuels jusqu'aux processus de notation
Indices de développement de la LE	Les démarches sont qualitatives, centrées sur les contenus, notamment pour pouvoir clarifier ses attentes envers les élèves : « *c'est plus dans la qualité du test que dans la quantité, maintenant […] quand on évalue, qu'est-ce qui, pour nous, est important ? Qu'est-ce qu'on veut voir dans le test ? C'est une manière de faire ressortir le fondamental qui nous remet directement au cœur de la matière* »	Le modèle permet de penser, voire d'anticiper la cohérence de la démarche d'évaluation : « *à partir du moment où on choisit nos objectifs, qu'on décide clairement ce qu'on veut évaluer comme matière, je ne vois pas à quel moment ça pourrait nous poser problème d'aligner. Maintenant, l'enjeu, pour moi, il est aussi au début, enfin, justement, je l'ai dit plusieurs fois, au choix de la matière* »	La justesse est pensée de manière globale. Il est alors important que l'épreuve « *soit alignée du début, à la fin. Ça veut dire que j'arrive à me sentir bien, […] que ça soit aligné depuis la matière travaillée, des objectifs, jusqu'à la matière choisie dans le test. Et puis ensuite au résultat chiffré* »	Aucun doute n'est exprimé sur d'éventuels compromis à réaliser, au contraire : « *maintenant j'ai vraiment envie de rester dans le qualitatif, de ne plus passer par une autre manière de faire* », tout en soulignant que la matière va jouer un rôle prépondérant dans la faisabilité de telles pratiques : « *si on cherche à être plus qualitatif, ben forcément, on peut mettre moins d'éléments [dans l'épreuve], on va chercher à sélectionner ceux qui seront les plus pertinents, les plus importants* »

Samia

En fin de formation, sa référence pour évaluer est le plan d'études. Les tâches évaluatives se rapportent davantage aux contenus enseignés et appris, la note suffisante est établie en fonction d'attentes relatives aux apprentissages (voir Tableau 12).

Des doutes subsistent toutefois quant à l'opérationnalisation de la cohérence au sein de certaines épreuves, notamment pour définir les seuils des autres notes. Le Tableau 13 présente le développement de sa LE.

Les indices de développement de la LE chez Samia mettent en évidence que la centration des pratiques évaluatives sur des contenus jugés clés (choix des contenus) est à la base d'un regain de sens, notamment du point de vue des finalités poursuivies en évaluant. L'appropriation du modèle théorique et sa mise en œuvre ont joué un rôle non négligeable dans ces changements.

Tableau 12

Évolution des compétences en évaluation chez Samia

Samia	Compétence A	Compétence B	Compétence C	Compétence D	Compétence E	Compétence F
Avant la recherche	Référence aux moyens d'enseignement pour l'évaluation Plan d'études jugé « rébarbatif »	Élaboration des épreuves comme un mélange de tâches simples et complexes, portant toujours sur de la technique de la langue et sur une compétence langagière	Attribution des points en fonction de « ce qui est attendu minimalement »	Construction du seuil de la note 4 entre 60 et 70 % des points, en fonction de la complexité de l'épreuve. Puis, « équilibrage » des autres notes à l'aide de barèmes standardisés	-	-
Après la recherche E1 : Expression orale (le texte théâtral) E2 : Conjugaison	Analyse des contenus disciplinaires davantage en rapport avec le plan d'études Reformulation des objectifs à stabiliser	Élaboration des tâches évaluatives traduisant des contenus disciplinaires clés de la discipline, dont le niveau de complexité peut varier dans une même épreuve	Attribution des points en référence aux contenus issus du plan d'études	Construction de la note 4 en rapport avec les contenus issus du plan d'études, traduisant la complexité des contenus présents dans les tâches, issue d'un raisonnement se référant aux apprentissages Désalignements curriculaires attribuables à la compensation des points	Désalignements curriculaires évités jusqu'au choix des tâches évaluatives, probables au niveau de la pondération, puis omniprésents dans la notation	Interprétation du plan d'études proche de la source, choix assumés des contenus et des tâches, seuil de la note 4 en rapport avec les apprentissages Compensation des points omniprésente

Tableau 13

Indices clés du développement de la LE de Samia

Dimensions de la LE (Samia)	L'enseignante explique en quoi ses pratiques peuvent soutenir les apprentissages	L'enseignante explique en quoi la théorie sur l'évaluation (alignement curriculaire élargi) est au service de la cohérence de ses pratiques	L'enseignante exprime la justesse de ses pratiques au regard des attentes relatives aux apprentissages	L'enseignante accepte les compromis et/ou dilemmes éventuels jusqu'aux processus de notation
Indices de développement de la LE	Les pratiques se centrent sur l'identification sélective des apprentissages qui sont acquis et/ou en voie d'acquisition : « *la grosse chose qui a changé, c'est que je n'évalue plus tout. Il y a des choses sur lesquelles je ne mets pas de note, je vérifie juste que tous mes élèves aient acquis le minimum […]. Donc je vais faire en sorte que les points ne valident que l'objectif que j'ai choisi. De manière à vraiment pouvoir voir ce qui est compris, ce qui n'est pas compris* »	En permettant de repérer les désalignements curriculaires, le modèle a participé à donner du sens aux démarches évaluatives : « *en fabriquant mon évaluation, j'ai pu voir les cassures. Et c'est là que j'ai vraiment intégré et compris cet alignement. J'ai saisi, finalement, cette importance de donner du sens. Donner du sens en amont, c'était simplement "j'ai décidé d'enseigner ça donc j'évalue ça, et je travaille ce que j'ai décidé d'évaluer"* »	Le fait de se concentrer sur des contenus clés à évaluer est vécu comme « *un soulagement […], une pratique plus juste, parce que je me concentre sur ça, ça, ça et ça* »	Globalement, une posture rassurée émerge : « *j'ai regagné en professionnalisme, en autonomie, en liberté* » Il reste encore des difficultés à assumer la construction de notes en lien avec l'apprentissage sur quelques contenus disciplinaires

La synthèse

Nous observons chez tous les enseignants que le but de la démarche d'évaluation sommative gagne en clarté : ils expriment une volonté de cibler les contenus clés pour privilégier une recherche de qualité dans les informations recueillies dans les épreuves, ainsi qu'une conscience de la nécessité à formuler des rétroactions aux élèves. L'exploitation du modèle théorique leur permet également de mieux anticiper, comprendre et, dans une certaine mesure, éviter les désalignements curriculaires. Cette nouvelle perception de la démarche évaluative accroît le sentiment d'évaluer de manière cohérente, depuis l'enseignement, et pour soutenir les apprentissages des élèves. À ce titre, nous pouvons affirmer qu'un développement de LE a lieu chez chaque enseignant.

En revanche, si tous les enseignants sont aptes à assumer leurs choix dans la construction des épreuves, certains éprouvent encore de la difficulté à gérer des pratiques de notation pas toujours satisfaisantes. En effet, c'est sur ce plan que des facteurs comme la discipline, les prescriptions ou la collaboration viennent poser des problèmes perturbant la cohérence de leurs démarches, surtout chez les enseignants de mathématiques, ce dont discutera davantage la prochaine section.

Discussion

Relativement à notre question de recherche, nous avons montré que le développement de compétences en évaluation — rendu possible par l'appropriation de concepts théoriques ciblés — a permis aux enseignants d'améliorer la cohérence dans leurs pratiques, de trouver des compromis justifiables dans les situations délicates, rendant ainsi compte d'un développement de leur LE. Nous pouvons par ailleurs faire l'hypothèse forte que cette LE est ici située et non technique (DeLuca et al., 2019), dans la mesure où des questions sont toujours présentes chez la plupart d'entre eux, montrant la complexité de leurs négociations, notamment entre des dimensions contextuelles de niveaux micro, méso et parfois macro (Fulmer et al., 2015). Plusieurs éléments appuient ce constat :

1. Tous les enseignants soulignent les liens qui se sont intensifiés entre leurs pratiques d'évaluation sommative, et les processus d'enseignement et d'apprentissage mis en œuvre et effectifs. Leur souci de s'adapter ici aux contenus clés est saisi dans une préoccupation de faire progresser chaque élève pour l'évaluer au plus près de ses réelles acquisitions.

2. Les contenus disciplinaires jouent un rôle important dans la mise en œuvre de pratiques évaluatives sommatives au service des apprentissages, notamment lorsqu'il s'agit d'exploiter le plan d'études à des fins de certification. À ce titre, tous relèvent également la difficulté parfois de suivre l'organisation de cet outil et d'être cohérent jusqu'à la notation, en fonction des objets disciplinaires évalués.
3. Ils sont toutefois unanimes à préciser que le développement de telles pratiques dépendra de leur capacité à partager leurs connaissances et de faire comprendre leurs pratiques à leurs collègues de discipline. En mathématiques, Fernand dit être prêt à partager ses connaissances sur l'alignement curriculaire pour en « faire bénéficier le groupe, mais sans forcément prendre des risques et finir comme le mouton noir à qui on veut plus adresser la parole ». Paul désire tenter une approche en douceur : « je vais le montrer à mon collègue, quand même, parce que je trouve ça intéressant. Et puis je vais voir ce que ça donne ». Quant à Philémon, il considère le modèle théorique comme un outil aidant pour clarifier ses décisions en disant que « même dans des collaborations un peu plus tendues où on n'a pas la même approche, beaucoup peut être fait ». En français, les mêmes enjeux sont relevés. Alizée précise que des changements pourront avoir lieu au niveau de la notation « si mes collègues sont d'accord », et Anouck sait que la collaboration conditionnera la pérennité des nouvelles pratiques, car « on ne sait pas ce que ça va donner plus tard, comment on va tenir le coup. Il y a aussi toute la contrainte liée aux collègues ». Enfin, Samia voit une ouverture à la suite des premiers échanges qu'elle a eus avec sa collègue directe et reste positive : suivant les partages autour du modèle, « elle s'est remise en question… quelqu'un chez qui je m'y attendais pas du tout. Alors je l'ai encouragée […] donc on va regarder ensemble ».
4. Enfin, et c'est un point essentiel, nous observons que les quatre dimensions du développement de la LE que nous avons exploitées sont en interaction. Par exemple, les enseignants relèvent que l'appropriation du modèle théorique les a amenés à penser l'évaluation comme une démarche centrée premièrement sur les apprentissages et les contenus.

> Cette attention leur a permis de développer des outils et des épreuves à visée qualitative cohérentes, ce qui a donné un sens à l'évaluation sommative au-delà de sa stricte fonction de certification.

Ces éléments soulignent que le développement de la LE n'est jamais terminé. Il n'est dès lors pas pertinent de dire une fois pour toutes d'un enseignant qu'il est au bénéfice, ou non, d'une LE, cette dernière étant le fruit de perpétuelles négociations en fonction du contexte. Cela étant, et conformément à la littérature, ce dernier constat doit s'accompagner de deux réflexions importantes à nos yeux.

La première concerne le dispositif de formation. En effet, et comme le dit Anouck, celui que nous avons élaboré, bien qu'orienté sur les échanges et la création de démarches incarnées dans les réalités des participants, était limité dans le temps. Cela ne nous a pas permis de savoir comment la LE de chaque enseignant a évolué lors du retour dans leurs routines, leurs contextes, leurs environnements collaboratifs et hiérarchiques. Bien que porteur de développement, il serait enrichissant d'envisager un dispositif reprenant ces mêmes caractéristiques, encore plus ancré dans l'expérience des uns et des autres au sein des classes, mais à plus long terme (Guskey et Yoon, 2009).

La seconde réflexion porte sur le développement des pratiques de notation cohérentes. En effet, comme nous l'avons vu, c'est sur ce plan que le développement des compétences a rencontré le plus d'obstacles : les enseignants ont eu des difficultés à associer de manière satisfaisante une recherche de cohérence dans la construction des notes avec des spécificités disciplinaires, des prescriptions et des outils imposés, comme les échelles et les barèmes. Ce constat est intéressant à au moins deux titres. Tout d'abord, il souligne l'articulation entre pratiques de notation et pratiques d'évaluation sommative (McMillan et Nash, 2000) tout en mettant en évidence leurs spécificités respectives, puisque l'on observe, au travers de nos indices de développement de la LE, qu'un enseignant peut tout à fait justifier de la cohérence de ses pratiques évaluatives alors qu'il reste peu à l'aise avec ses pratiques de notation. Dans ce sens, nous pouvons faire l'hypothèse que l'appropriation de notre modèle théorique pour penser les phénomènes de cohérence jusqu'à la notation ne permet pas toujours de dépasser certains obstacles, car il subsiste pour l'enseignant une obligation de gérer des choix sur ce plan qui dépassent l'état de ses connaissances et ses compétences, en raison de leurs liens avec ses valeurs et ses croyances (Birenbaum, 2016). En effet, au-delà de signifier un résultat, la note reste un outil au service de multiples rôles qui circule hors

de la classe (parents, camarades, enseignants, direction d'école, milieux professionnels, comme ici au secondaire), et qu'à ce titre, les négociations des enseignants pour la rendre cohérente se réalisent probablement aussi en regard d'éléments de niveau macro (Brimi, 2011; Brookhart, 2011; McMillan, 2019; Pasquini et DeLuca, 2021), extérieurs à l'apprentissage, sur lesquels ils n'ont parfois que peu de prise. Autant de constats qui pousseraient, en référence à la productive recherche dans ce champ, à envisager une possible conceptualisation d'une littératie en évaluation articulée avec la notation.

Conclusion

Notre discussion a montré dans quelle mesure nous avons répondu à notre question de recherche qui se formulait ainsi : « Dans quelle mesure le développement de compétences en évaluation — rendu possible par l'appropriation de concepts théoriques — permet-il aux enseignants de résoudre des problèmes qu'ils identifient dans leurs pratiques évaluatives et d'ainsi rendre compte d'un développement d'une littératie en évaluation ? » Nous avons également pu valider nos deux hypothèses, en confirmant que l'appropriation de contenus théoriques est indispensable au développement d'une LE située. À ce titre, la LE nous semble réellement porteuse pour comprendre et décrire le développement de compétences en évaluation chez les enseignants de tous les degrés, notamment lorsqu'il s'agit pour eux d'expliciter, de justifier et d'assumer leurs décisions. Elle gagne encore plus en pertinence lorsque l'on observe que le développement de pratiques évaluatives au service des apprentissages est un processus complexe, long, irrégulier, qui ne va pas de soi. Cette complexité augmente probablement encore lorsqu'il s'agit de mettre en œuvre des démarches d'évaluation sommative au service des apprentissages (Laveault et Allal, 2016), cette dernière étant souvent considérée comme uniquement au service de la certification, et donc plutôt négative (Taras, 2005) ou bassement technique (Dubus, 2006). Le développement de la LE devrait dès lors permettre aux enseignants de dépasser une compréhension superficielle de l'approche de l'évaluation-soutien d'apprentissage — qui est souvent la cause de mises en œuvre inefficaces (Hayward et Spencer, 2010; Webb et Jones, 2009) — et d'éviter ainsi de réduire les pratiques à des procédures mécaniques (Klenowski, 2009; Moss, 2013). Nos propres travaux vont dans ce sens : nous avons fréquemment décrit des pratiques évaluatives sommatives et des pratiques de notation le

plus souvent pilotées par des habitudes et des procédures déconnectées de l'apprentissage (Pasquini, 2018, 2019, 2020, 2021), sans valeur ajoutée pour l'enseignant ou l'élève.

Toutefois, deux limites à notre étude sont à poser. La première est relative à la taille et à la nature du corpus d'enseignants avec qui nous avons collaboré : en effet, leur nombre réduit ne permet en aucun cas de généraliser nos résultats, et leur participation volontaire et engagée a pu se dérouler dans les meilleures conditions, notamment parce qu'ils avaient tous des questions importantes en évaluation qu'ils souhaitaient partager et résoudre. Il serait alors pertinent de penser les conditions nécessaires à la mise en œuvre de dispositifs similaires avec des enseignants moins concernés par cette question de cohérence en évaluation sommative, et surtout plus nombreux, tout en étant attentif au fait qu'il reste délicat de répliquer de tels processus de formation à large échelle (Black et al., 2003). La seconde limite concerne la difficulté que nous avons eue à engranger des données sur les négociations menées par les enseignants avec leurs collègues de même discipline (niveau méso), ou avec des éléments contextuels de niveau macro, comme la hiérarchie ou les parents, qui sont très influents sur le possible développement de pratiques évaluatives innovantes (Fulmer et al., 2015).

Il s'avère donc nécessaire d'approfondir, dans de futurs travaux, les axes de recherche évoqués sur le développement de la LE, en privilégiant une approche située qui tient également compte de dimensions plus larges. En effet, les dernières conceptualisations de la LE dont nous disposons (p. ex., DeLuca et al., 2019) soulignent que le développement des pratiques évaluatives des enseignants au service des apprentissages change non seulement leur manière de négocier les multiples facteurs complexes auxquels ils sont confrontés, mais aussi leur identité d'évaluateur. Cette dernière serait alors le fruit de subtiles interactions entre leurs expériences personnelles, leur contexte professionnel et leur environnement prescriptif externe (Mockler, 2011). L'identité de l'évaluateur, comme concept dynamique et en perpétuel mouvement, serait aussi le fruit de négociations chez l'enseignant, entre ce qu'il sait, ce qu'il perçoit, ce qu'il croit, ce en quoi il a confiance et la perception de son rôle (Looney et al., 2018). En élargissant ainsi la conceptualisation de la LE à l'analyse de l'évolution des dimensions identitaires et culturelles de l'enseignant, un champ ambitieux et porteur de connaissances pour la recherche et la pratique s'ouvre.

Références

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice, 20*(1), 91–106. https://doi.org/10.1080/0969594X.2011.650150

Airasian, P. W. et Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory Into Practice, 41*(4), 249–254. https://doi.org/10.1207/s15430421tip4104_8

Allal, L. et Laveault, D. (2009). Assessment for learning: évaluation-soutien d'apprentissage. *Mesure et évaluation en éducation, 32*(2), 99–106. https://doi.org/10.7202/1024956ar

Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory Into Practice, 41*(4), 255–260. https://doi.org/10.1207/s15430421tip4104_9

Anderson, L. W. et Krathwohl, D. R. (dir.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles. Research-based principles to guide classroom practice*. https://www.researchgate.net/publication/271849158

Bednarz, N. (dir.). (2013). *Recherche collaborative et pratique enseignante* : *regarder ensemble autrement*. L'Harmattan.

Bednarz, N. (2015). La recherche collaborative. *Carrefours de l'éducation, 39*(1), 171–184. https://doi.org/10.3917/cdle.039.0171

Biggs, J. (2003). *Aligning teaching for constructive learning*. The Higher Education Academy. https://www.researchgate.net/profile/John_Biggs3/publication/255583992

Birenbaum, M. (2016). Assessment culture versus testing culture: The impact on assessment for learning. Dans D. Laveault et L. Allal (dir.), *Assessment for learning: Meeting the challenge of implementation* (p. 275–292). Springer. https://doi.org/10.1007/978-3-319-39211-0_16

Black, P., Harrison, C., Lee, C., Marshall, B. et William, D. (2003). *Assessment for learning: Putting it into practice*. Open University Press.

Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. Dans J. H. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 87–106). SAGE Publications.

Braxmeyer, N., Guillaume, J.-C. et Levy, J.-F. (2004). *Les pratiques d'évaluation des enseignants au collège* (NE 04.13). Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, Direction de l'évaluation et de la prospective. https://archives-statistiques-depp.education.gouv.fr/Default/doc/SYRACUSE/10844/les-pratiques-d-evaluation-des-enseignants-au-college-nicole-braxmeyer-jean-claude-guillaume-et-jean?_lg=fr-FR

Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research et Evaluation*, *16*(16), article 17. https://doi.org/10.7275/j531-fz38

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brookhart, S. M. (2017). *How to use grading to improve learning*. ASCD.

Cross, L. H. et Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53–72. https://doi.org/10.1207/s15324818ame1201_4

DeLuca, C. et Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy &Practice, 24*(2), 121–126. https://doi.org/10.1080/0969594X.2017.1297010

DeLuca, C., Coombs, A., MacGregor, S. et Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education, 4*(94). https://doi.org/10.3389/feduc.2019.00094

Desgagné, S. (2005). *Récits exemplaires de pratique enseignante : analyse typologique*. Presses de l'Université du Québec.

Dionne, L., Lemyre, F. et Savoie-Zajc, L. (2010). Vers une définition englobante de la communauté d'apprentissage (CA) comme dispositif de développement professionnel. *Revue des sciences de l'éducation, 36*(1), 25–43. https://doi.org/10.7202/043985ar

Dogan, S., Pringle, R. et Mesa, J. (2016) The impacts of professional learning communities on science teachers' knowledge, practice and student learning: A review. *Professional Development in Education, 42*(4), 569–588. https://doi.org/10.1080/19415257.2015.1065899

Dubus, A. (2006). *La notation des élèves. Comment utiliser la docimologie pour une évaluation raisonnée*. Armand Colin.

Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural-historical perspective. *Assessment in Education: Principles, Policy & Practice, 22*(2), 224–246. https://doi.org/10.1080/0969594X.2015.1015403

Fulmer, G. W., Lee, I. C. H. et Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice, 22*(4), 475–494. https://doi.org/10.1080/0969594X.2015.1017445

Gauthier, C., Mellouki, M., Bissonnette, S. et Richard, M. (2005, avril). Écoles efficaces et réussite *scolaire des élèves à risque. Un état de la recherche* [Rapport de recherche]. CRIFPE, Université Laval. https://mobile.eduq.info/xmlui/bitstream/handle/11515/33103/W027996-Gauthier-ecoles-efficaces-CRIFPE-ULAVAL-avril-2005.pdf?sequence=1&isAllowed=y

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. The Falmer Press.

Good, R. (2011). Formative use of assessment information: It's a process, so let's say what we mean. *Practical Assessment, Research & Evaluation, 16*(3). https://doi.org/10.7275/3yvy-at83

Guskey, T. R. et Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades, *Assessment in Education: Principles, Policy & Practice, 26*(3), 303–320. https://doi.org/10.1080/0969594X.2018.1555515

Guskey, T. R. et Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500. https://doi.org/10.1177/003172170909000709

Harlen, W. (2005). Teacher's summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16*(2), 207–223. https://doi.org/10.1080/09585170500136093

Hayward, L. et Spencer, E. (2010). The complexities of change: Formative assessment in Scotland. *Curriculum Journal, 21*(2), 161–177. https://doi.org/10.1080/09585176.2010.480827

Kaufmann, J.-C. (2011). *L'entretien compréhensif* (3e éd.). Armand Colin.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16*(3), 263–268. https://doi.org/10.1080/09695940903319646

Klenowski, V. et Wyatt-Smith, C. (2014). *Assessment for education: Stadards, judgement and moderation.* SAGE Publications. https://dx.doi.org/10.4135/9781526401878

Klinger, D. A., McDivitt, P. J., Howard, B. B., Muñoz, M. A.Rogers, W. T. et Wylie, E. C. (2015). *Classroom assessment standards for PreK-12 teachers.* Kindle Direct Press.

Laveault, D. et Allal, L. (dir.). (2016). *Assessment for learning: Meeting the challenge of implementation*. Springer. http://dx.doi.org/10.1007/978-3-319-39211-0

Looney, A., Cumming, J., van Der Kleij, F. et Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice, 25*(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090

Lukin, L. E., Bandalos, D. L., Eckhout, T. J. et Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice, 23*(2), 26–32. http://dx.doi.org/10.1111/j.1745-3992.2004.tb00156.x

Martone, A. et Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research, 79*(4), 1332–1361. https://doi.org/10.3102/0034654309341375

McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. Dans T. R. Guskey et S. M. Brookhart (dir.), *What we know about grading: What works, what doesn't and what's next* (p. 84–112). ASCD.

McMillan, J. H. et Nash, S. (2000, 27 avril). *Teacher classroom assessment and grading practices decision making* [Communication orale]. 2000 NCME Annual Meeting, New Orleans. https://files.eric.ed.gov/fulltext/ED447195.pdf

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

Mockler, N. (2011) Beyond 'what works': Understanding teacher identity as a practical and political tool. *Teachers and Teaching, 17*(5), 517–528. https://doi.org/10.1080/13540602.2011.602059

Moss, C. M. (2013). Research on classroom summative assessment. Dans J. H. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 235–256). SAGE Publications.

Mottier Lopez, L. et Tessaro, W. (dir.).(2016). *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages.* Peter Lang.

Mottier Lopez, L. et Morales Villabona, F. (2018). Quand des enseignants de l'école primaire évaluent des résolutions de problèmes additifs : études des incidents critiques en cours de jugement. *Mesure et évaluation en éducation, 41*(1), 125–161. https://doi.org/10.7202/1055899ar

Nisbet, I. et Shaw, S. (2020). *Is assessment fair?* SAGE Publications.

Organisation de coopération et de développement économiques (OCDE). (2000). *La littératie à l'ère de l'information : rapport final de l'enquête internationale sur la littératie des adultes*. http://www.oecd.org/fr/edu/innovation-education/39438013.pdf

Organisation for Economic Cooperation and Development (OECD). (2012). Grade Expectations : How marks and education policies shape student's ambitions. http://dx.doi.org/10.1787/9789264187528-en

Paquay, L., Van Nieuwenhoven, C. et Wouters, P. (dir.). (2010). *L'évaluation, levier du développement professionnel ? Tensions, dispositifs, perspectives.* De Boeck.

Pasquini, R. (2013). *Quand les récits de pratique enseignante parlent d'apprentissage*. Presses de l'Université Laval.

Pasquini, R. (2016). Et si, pour mieux saisir la complexité de la pratique, nous passions par l'écriture ? *Didactiques en pratique, 2*, 21–29. http://hdl.handle.net/20.500.12162/2423

Pasquini, R. (2018). *Le modèle théorique de l'alignement curriculaire élargi pour étudier des pratiques évaluatives sommatives d'enseignants de mathématiques et de français du secondaire : enjeux conceptuels et pragmatiques* [Thèse de doctorat, Université de Genève]. Archive ouverte UNIGE. https://doi.org/10.13097/archive-ouverte/unige:106442

Pasquini, R. (2019). Élargir conceptuellement le modèle de l'alignement curriculaire pour comprendre la cohérence des pratiques évaluatives sommatives notées des enseignants : enjeux et perspectives. *Mesure et évaluation en éducation, 42*(1), 63–92. https://doi.org/10.7202/1066598ar

Pasquini, R. (2020). Et si la recherche collaborative participait au développement de pratiques évaluatives sommatives cohérentes ? *Mesure et évaluation en éducation, 43*(1), 33–66. https://doi.org/10.7202/1076965ar

Pasquini, R. (2021). *Quand la note devient constructive. Évaluer pour certifier et soutenir les apprentissages*. Presses de l'Université Laval.

Pasquini, R. et DeLuca, C. (2021). Grading in a dilemmatic space: An exploratory cross-cultural analysis of mathematics and language secondary teachers. *Comparative and International Education*, *49*(2), 51–70. https://doi.org/10.5206/cieeci.v50i1.14133

Popham, W. J. (2004). All about accountability/why assessment illiteracy is professional suicide. *Educational Leadership, 62*(1), 82–83. https://www.ascd.org/el/articles/why-assessment-illiteracy-is-professional-suicide

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice, 48*(1), 4–11. https://doi.org/10.1080/00405840802577536

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238–245.

Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Tierney, R. D. (2006). Changing practices: influences on classroom assessment. *Assessment in Education: Principles, Policy & Practice, 13*(3), 239–264. https://doi.org/10.1080/09695940601035387

Tierney, R. D., Simon, M. et Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum, 75*(3), 210–227. https://doi.org/10.1080/00131725.2011.577669

Timperley, H., Wilson, A., Barrar, H. et Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. New Zealand Ministry of Education. https://www.educationcounts.govt.nz/__data/assets/pdf_file/0017/16901/TPLandDBESentireWeb.pdf

Volante, L. et Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749–770. https://journals.sfu.ca/cje/index.php/cjerce/article/view/2973

Webb, M. et Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice, 16*(2), 165–184. https://doi.org/10.1080/09695940903075925

Wiliam, D. (2000, août). *Integrating formative and summative functions of assessment* [communication orale]. International Congress on Mathematics Education, Makuhari, Tokyo. https://discovery.ucl.ac.uk/id/eprint/10001151/1/Wiliam2000IntergratingAEA-E_2000_keynoteaddress.pdf

Willis, J., Adie, L. et Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher, 40*(2), 241–256. https://doi.org/10.1007/s13384-013-0089-9

Xu, Y. et Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*(1), 149–162. https://doi.org/10.1016/j.tate.2016.05.010