Document generated on 07/19/2025 12:52 a.m.

Mesure et évaluation en éducation

Et si la recherche collaborative participait au développement de pratiques évaluatives sommatives cohérentes ?

Raphaël Pasquini

Volume 43, Number 1, 2020

Recherches collaboratives et évaluation : de l'objet au processus

URI: https://id.erudit.org/iderudit/1076965ar
DOI: https://doi.org/10.7202/1076965ar

See table of contents

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)
2368-2000 (digital)

Explore this journal

Cite this article

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

Article abstract

Numerous studies show that teachers' summative assessment practices often lack coherence in all contexts. This article describes how six secondary school teachers of French and mathematics in French-speaking Switzerland improved this coherence by participating in a collaborative research in which reflective activity was the subject of particular attention, taking into account their questions, context and discipline. Conducted with the help of conceptualizing categories, the qualitative analysis of their tests, comprehensive interviews, writing tales and exchanges during work sessions show that six skills evolve positively in each of them in a dynamic perspective, at the service of their professional development, and that this evolution is notably allowed by the theoretical contents proposed within the scheme.

© ADMEE-Canada - Université Laval, 2021

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/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Et si la recherche collaborative participait au développement de pratiques évaluatives sommatives cohérentes ?

Raphaël Pasquini
Haute école pédagogique du canton de Vaud, Lausanne

Mots clés : recherche collaborative, activité réflexive, évaluation sommative, compétences en évaluation, développement professionnel

De multiples études montrent que les pratiques évaluatives sommatives des enseignants manquent souvent de cohérence, tous contextes confondus. Cet article décrit comment six enseignants de français et de mathématiques suisses romands du secondaire ont amélioré cette cohérence en participant à une recherche collaborative[1], dans laquelle l'activité réflexive a fait l'objet d'une attention particulière en prenant en compte leurs questions, leur contexte et leur discipline. Menée à l'aide de catégories conceptualisantes, l'analyse qualitative de leurs épreuves, d'entretiens compréhensifs, d'écrits de pratique et des échanges en séances de travail montre que six compétences évoluent positivement chez chacun d'entre eux dans une perspective dynamique, au service de leur développement professionnel, et que cette évolution est notamment permise par les contenus théoriques proposés au sein du dispositif.

Keywords : collaborative research, reflective activity, summative assessment, assessment skills, professional development

Numerous studies show that teachers' summative assessment practices often lack coherence in all contexts. This article describes how six secondary school teachers of French and mathematics in French-speaking Switzerland improved this coherence by participating in a collaborative research in which reflective activity was the subject of particular attention, taking into account their questions, context and discipline. Conducted with the help of conceptualizing categories, the qualitative analysis of their tests, comprehensive interviews, writing tales and exchanges during work sessions show that six skills evolve positively in each of them in a dynamic perspective, at the service of their professional development, and that this evolution is notably allowed by the theoretical contents proposed within the scheme.

Palavras-chave : investigação colaborativa, atividade reflexiva, avaliação sumativa, competências em avaliação, desenvolvimento profissional

Vários estudos mostram que as práticas de avaliação sumativa dos professores carecem muitas vezes de coerência em todos os contextos. Este artigo descreve como seis professores de francês e de matemática da Suíça francófona do ensino secundário melhoraram essa coerência ao participarem numa investigação colaborativa, na qual a atividade reflexiva foi objeto de uma atenção particular, tendo em consideração as suas questões, o seu contexto e a sua disciplina. Realizada por meio de categorias conceituais, a análise qualitativa dos seus testes, entrevistas abrangentes, narrativas da prática e as partilhas em sessões de trabalho mostra que seis competências evoluem positivamente em todos numa perspectiva dinâmica, ao serviço do seu desenvolvimento profissional, e que esta evolução é significativamente permitida pelos conteúdos teóricos oferecidos no âmbito do dispositivo.

Note des auteures : La correspondance liée à cet article peut être adressée à [raphael.pasquini@hepl.ch].

Problématique

Depuis une vingtaine d'années, des travaux sur le développement professionnel en éducation soulignent une démotivation grandissante des enseignants[2], due notamment à une dépossession de leur pouvoir d'action par la noosphère, qui les assimilerait davantage à des techniciens qualifiés de l'éducation qu'à des praticiens réflexifs et responsables (Osborn, 2002 ; Rayou, 2008). Ce phénomène a été spécifiquement pointé par des recherches qui postulent que les enseignants se sentent dépossédés de leurs compétences à évaluer leurs élèves lorsqu'ils font face à de nombreuses et récurrentes modalités d'évaluations externes, comme c'est aujourd'hui le cas dans la plupart des systèmes scolaires occidentaux (Jorro, 2015 ; Koretz, 2017 ; Maroy et Cattonar, 2002 ; Osborn, 2006).

Conjointement, des études dans le domaine de l'évaluation soulignent qu'un pourcentage restreint d'enseignants toutes années d'études confondues met en œuvre des démarches d'évaluation sommative[3] centrées sur les apprentissages et les progrès des élèves (p. ex., Bateman et al., 2009 ; Moss, 2013 ; Mottier Lopez, 2014). Nusche, Radinger, Santiago et Shewbridge (2013) confirment ce constat dans une méta-analyse menée à large échelle qui montre que, dans la plupart des pays de l'Organisation de coopération et de développement économiques (OCDE), les évaluations sommatives sont centrées sur un nombre limité d'apprentissages le plus souvent peu complexes, réduisant ainsi le contenu du curriculum. Mottier Lopez (2014) va dans le même sens en décrivant des pratiques qualifiées de « traditionnelles » parce qu'axées fortement sur les notes et peu informatives.

D'autres travaux, complémentaires, pointent de forts écarts existant entre les représentations que se font les enseignants de pratiques d'évaluation sommative cohérentes et la description de leurs propres pratiques (Braxmeyer et al., 2005 ; Martinez et al., 2009 ; Rieg, 2007). Selon Moss (2013), ces écarts s'expliqueraient notamment par le fait que les enseignants utilisent régulièrement une variété de méthodes d'évaluation sommative alors que leurs compétences en la matière sont faibles. Cette auteure fait alors l'hypothèse qu'une des causes de ces problèmes viendrait notamment d'un manque de formation ou encore d'actions de formation

insuffisamment centrées sur des contenus pédagogiques et disciplinaires (Guskey et Yoon, 2009 ; Timperley, 2011). Les pratiques d'évaluation sommative, tous contextes et années d'études confondus, seraient donc souvent davantage le fruit d'habitudes peu remises en question ou d'essais-erreurs que de pratiques modélisées et cohérentes.

Étant donné que le développement professionnel est optimal lorsque « les enseignants deviennent coapprenants ou apprenants autorégulés par un processus d'enquête situé dans l'environnement immédiat de leur pratique » (Timperley, 2011, p. 34), la conjugaison de ces éléments encourage à mettre en œuvre des dispositifs de recherche en évaluation offrant aux enseignants la possibilité de « s'orienter vers la résolution de problèmes, d'être autonomes dans le choix des stratégies pédagogiques » (Perrenoud, 1996, p. 554). Cette nécessité serait d'autant plus grande relativement à l'évaluation sommative pour que cette dernière puisse être en lien avec les apprentissages des élèves, et non dépendante d'autres facteurs (Hadji, 2016, 2017 ; Pasquini, 2021).

Ces constats nous amènent à formuler deux hypothèses :

1. Amener les enseignants à améliorer la cohérence de leurs pratiques évaluatives sommatives demande à travailler leur capacité à mobiliser des concepts issus de la recherche (Altet, 2000) étant susceptibles d'être « des instruments du pouvoir d'agir des acteurs de terrain » (Vinatier et Morrissette, 2015, p. 142) ;
2. Avec l'idée que des dispositifs de formation de type descendant (*top-down*) ont peu d'impact positif sur les pratiques effectives (Jorro, 2013), le développement des compétences en évaluation doit alors être envisagé selon deux perspectives : a) la considération des pratiques réelles avec une attention portée aux « questions vives » de la pratique des enseignants inscrites dans leur réalité de classe (Pasquini, 2017) et b) la prise en compte des spécificités disciplinaires et des contextes (p. ex., type et niveau des élèves, prescriptions en matière d'évaluation).

Au regard de ces éléments de problématisation, notre question de recherche est la suivante : *En quoi une recherche collaborative peut-elle contribuer au développement de compétences en évaluation sommative au service de démarches de certification cohérentes ?*

Pour la traiter, nous commençons par dresser notre cadre conceptuel en y définissant les éléments théoriques clés de notre étude. Ensuite, la méthodologie présente les éléments de notre recherche collaborative, 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 entre le début et la fin de la recherche, dans une perspective visant à souligner la dimension dynamique du processus. Une discussion relève ensuite quelques tensions en vue de futures problématisations. Enfin, nous terminons avec une conclusion mettant notamment en évidence les atouts et limites du présent travail.

Cadre conceptuel

L'évaluation sommative

L'évaluation sommative se définit comme « une évaluation qui a pour objet un inventaire. Son usage social est de vérifier (probation), sa fonction sociale de certifier. Elle a comme fonctions annexes classer, situer ou encore informer » (Hadji, 1989, cité dans Mottier Lopez, 2015, p. 23). Relativement à notre problématique, trois raisons nous ont poussé à suivre cette définition :

1. Elle renvoie à la nécessité d'obtenir des preuves des apprentissages essentiels des élèves (*to sum up* ; Brookhart, 2017), qui induit la question « qu'évaluer ? » et qui interroge les modalités permettant une évaluation référée aux apprentissages et aux contenus ;
2. Elle met en évidence les fonctions informatives de toute démarche sommative. La note validant un résultat peut dès lors, à certaines conditions, renseigner l'élève et le corps enseignant sur les apprentissages évalués, au-delà du seul chiffre ;
3. Enfin, elle considère que l'évaluation sommative peut, dans une certaine mesure, poursuivre d'autres fonctions, comme classer ou situer les élèves, deux visées possiblement contrastées qui ont un impact important sur le rôle de la note dans leur parcours.

Ainsi, nous pouvons envisager que la mise en œuvre d'une démarche d'évaluation sommative devrait être qualitative, car centrée sur les apprentissages réellement menés, ce qui permet à l'enseignant de situer les élèves par rapport à l'atteinte d'un objectif et de les informer sur leurs performances au-delà d'une note chiffrée. La littérature s'accorde sur le fait que

la conceptualisation d'une cohérence en évaluation sommative (Anderson, 2002 ; Pasquini, 2018 ; Sadler, 2009) doit faire appel à des modèles théoriques (Mottier Lopez et Figari, 2012).

Un modèle théorique pour saisir la cohérence des pratiques évaluatives sommatives

Pour comprendre et analyser les phénomènes de cohérence en évaluation sommative, nous avons exploité avec les enseignants le modèle théorique de l'alignement curriculaire élargi (Pasquini, 2018, 2019). Conceptualisé à partir du modèle de l'alignement curriculaire (Anderson, 2002 ; Gauthier et al., 2005), il désigne, en 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).

Plus précisément, et comme le montre la figure 1, la *référenciation restreinte* décrit « ce qui fait foi » chez les enseignants quand ils opèrent des choix relativement aux objectifs d'apprentissages à évaluer d'un point de vue sommatif dans une épreuve (Vial, 2012). Le « *design* » (Wiggins et McTighe, 2005) explicite les processus de construction des tâches évaluatives complexes ou des items simples, du point de vue des contenus et des habiletés cognitives sur lesquels ils portent. La *pondération* renvoie au poids alloué aux apprentissages et aux outils exploités pour ce faire (des critères, des points ou des critères appariés à des points comme modalité mixte ; Jönsson, 2014). Enfin, la *notation* cible les démarches de construction de la note chiffrée en lien avec l'apprentissage à l'aide d'outils comme les échelles et les barèmes (Dubus, 2006 ; Marzano et Heflebower, 2011). Ce sont alors cinq relations entre les différentes composantes qui devraient entretenir entre elles des liens de cohérence au sein de toute épreuve, possiblement traduisibles par le biais de questions non exhaustives comme :

- Les objectifs formulés sont-ils évalués au travers des tâches ? Les tâches rendent-elles compte des objectifs évalués ? (relation A) ;
- Les objectifs formulés se retrouvent-ils dans les critères ? Les critères décrivent-ils les objectifs évalués ? (relation B) ;
- Les tâches donnent-elles lieu à un poids alloué à l'apprentissage cohérent ? Le poids des apprentissages évalués correspond-il à ce que les tâches évaluent quant à la complexité des contenus et des habiletés cognitives ? (relation C) ;

– Les contenus et habiletés présents dans les tâches se retrouvent-ils dans les descripteurs des échelles ? Les échelons des échelles donnent-ils à voir les contenus et habiletés issus des tâches ? (relation D) ;
– Relativement à la notation, les critères permettent-ils l'établissement des échelons, surtout le seuil de suffisance ? Les échelons et leur progression sont-ils en rapport avec les critères ? (relation E)

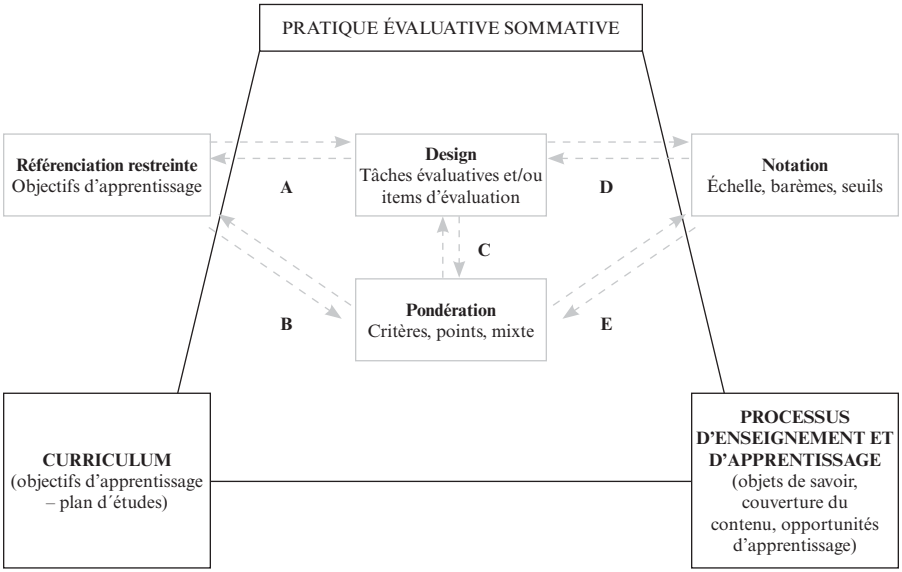

Figure 1. L'alignement curriculaire élargi (Pasquini, 2018, 2019)

Ici, la cohérence entre les différentes composantes du modèle est lue à l'aune de la taxonomie d'Anderson et Krathwohl (2001), qui propose de formuler des objectifs d'apprentissage en appariant des habiletés cognitives à des contenus (p. ex., en français : rédiger le portrait physique et moral d'un héros).

C'est au travers de ce questionnement permis par le modèle que nous avons parlé avec les enseignants de cohérence curriculaire élargie de l'épreuve en cas de cohérence globale (relations A, B, C, D et E) et de désalignements curriculaires (Airasian et Miranda, 2002 ; Biggs, 2003) sur

une ou plusieurs de ces relations quand cette cohérence se perdait, par exemple lorsque les tâches évaluatives ne permettaient pas ou partiellement de certifier les objectifs évalués (Pasquini, 2018). Ce questionnement itératif, dès le début du processus de recherche, a permis une explicitation et une objectivation des phénomènes de cohérence – ou d'incohérence – curriculaire au sein des pratiques, dans diverses situations. C'est également à l'aune des composantes de ce modèle que nous avons construit nos catégories conceptualisantes pour documenter le développement progressif des compétences en évaluation chez les enseignants, comme nous le détaillons dans notre méthodologie.

Les compétences en évaluation

Une importante revue de littérature (Pasquini, 2018) nous a permis de recenser les connaissances jugées indispensables au développement de compétences en évaluation (Paquay, 2012). Nous avons considéré ces connaissances comme des unités d'apprentissage mobilisées par les acteurs sous forme de compétences (Paquay et al., 2010). Nous avons dès lors postulé que le développement des compétences relevait d'une subtile mobilisation de connaissances en contexte en vue de résoudre des situations complexes (Paquay et al., 2010). Dans ce sens, nous avons suivi Le Boterf (2011), pour qui agir avec compétence revient à « mettre en œuvre une pratique professionnelle pertinente par rapport aux exigences de la situation, tout en mobilisant une combinatoire appropriée de ressources (connaissances, habiletés, comportements, aptitudes) » (p. 28).

Nous avons fait l'hypothèse selon laquelle l'amélioration de la cohérence des pratiques évaluatives des enseignants et leur capacité à progressivement les justifier en fonction de leurs discipline, contraintes et contexte seraient des signes de développement de leurs compétences, et que ce processus d'apprentissage serait rendu possible en partie depuis les connaissances issues de notre modèle théorique. La finalité poursuivie était d'amener chacun à pouvoir répondre à ses questions évaluatives, à trouver des stratégies pour les affronter et à faire face aux dilemmes verbalisés en début et en fin de processus (Perrenoud, 2004). Nous avons alors souhaité que les apports puissent être incarnés dans une manière de penser la cohérence évaluative sommative du point de vue des pratiques réelles, selon une perspective dynamique et disciplinaire.

Dès lors, en référence aux différentes composantes de notre modèle théorique, et en articulant les apports de la recherche sur les compétences clés en évaluation (Brookhart, 2011 ; Popham, 2009) avec les dimensions contextuelles des enseignants, nous avons formulé six compétences :

A. 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 ;

B. 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) ;

C. Adopter des systèmes de pondération permettant de mettre en évidence le poids donné à l'apprentissage référé aux objectifs et aux contenus en privilégiant, autant que possible, une pondération critériée, notamment pour permettre du feedback aux élèves à visée de régulation ;

D. Construire des notes de manière à ce qu'elles rendent compte des acquis et lacunes des apprentissages des élèves en élaborant ses propres échelles et barèmes en référence au système de pondération, à la complexité des tâches et aux objectifs évalués ;

E. Penser ses pratiques évaluatives sommatives en termes 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 ;

F. 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.

Méthodologie

Notre recherche collaborative

Pour favoriser un développement professionnel des enseignants sur leurs compétences en évaluation (Jorro, 2013 ; Van Nieuwenhoven et Colognesi, 2015), nous avons modélisé une recherche collaborative (voir Figure 2) en nous inspirant des travaux de Desgagné (1997) et de Desgagné, Bednarz, Lebuis, Poirier et Couture (2001).

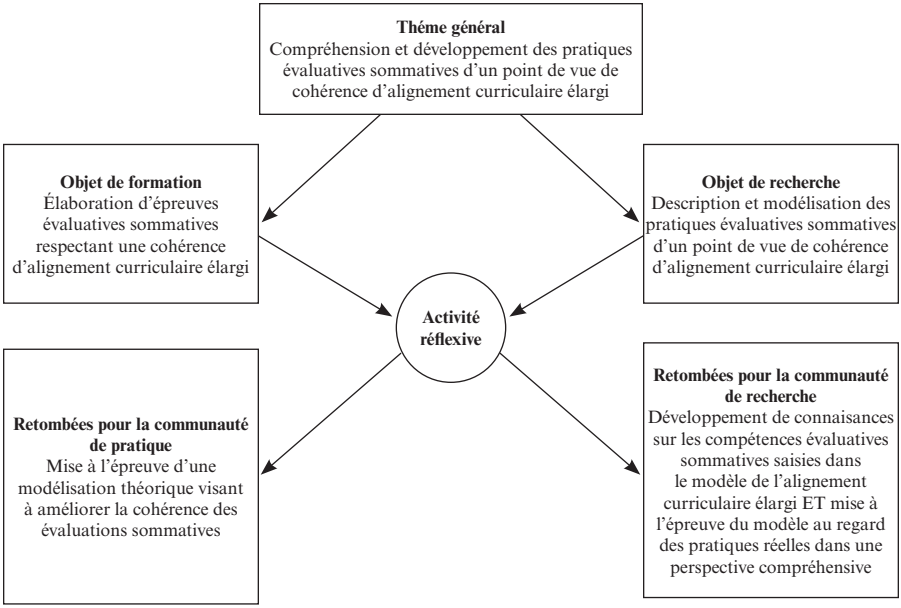

Figure 2. Modélisation de notre recherche collaborative

Plus spécifiquement, nous avons souhaité que l'activité réflexive soit présente dans tous les temps de notre dispositif décrit plus bas, dans la mesure où elle constitue le pivot de la recherche collaborative (Bednarz, 2013) en tant que « zone interprétative partagée entre chercheurs et praticiens où les argumentations et les ressources des uns et des autres sont mobilisées, où les praticiens en collaboration avec des chercheurs viennent éclairer un certain objet lié à cette pratique » (Bednarz, 2015, p. 174). D'un point de vue méthodologique, la mise en œuvre de cette activité a été

rendue possible par la création d'espaces dans lesquels les connaissances sur des objets liés à la pratique évaluative se sont coconstruites, issues du croisement entre notre logique de chercheur et celle des enseignants. Aussi avons-nous invité ces derniers à expliciter les problèmes et à formuler leurs questions dès le début du processus, dans un souci de rendre compte de leurs préoccupations, tout en étant attachés aux nôtres afin de construire des connaissances pour la pratique et la recherche.

Ce critère de double vraisemblance du dispositif (Bednarz, 2013 ; Desgagné, 2007) s'est décliné en fonction des trois phases de la recherche : 1) une phase de cosituation, où il s'est agi de faire en sorte que la problématique choisie soit pertinente d'un point de vue de préoccupation de recherche, mais aussi relativement aux questions des enseignants ; 2) une phase de coopération, où ont lieu les interactions entre les enseignants et où nous avons favorisé un espace de collecte de données en rapport avec la méthodologie, tout en laissant ouvert un espace de questionnement de la pratique ; et 3) une phase de coproduction, qui a permis une double fécondité des résultats sur le plan professionnel (pour la pratique) et sur le plan de la recherche, en lien avec le questionnement évolutif des participants. Ces phases ont traduit la validité du dispositif : « double pertinence sociale lorsque se coconstruit le projet, double rigueur méthodologique dans la coactivité autour des pratiques […] et double fécondité des résultats » (Bednarz, 2015, p. 181).

Un rôle d'accompagnateur pour le chercheur

Dans la mesure où il s'est avéré nécessaire de guider les praticiens dans leurs questionnements et démarches de développement, nous avons également investi de manière articulée un rôle de formateur qui « encadre la démarche de recherche des enseignants, prise au sens informel du terme » (Desgagné et al., 2001, p. 39). Ce double investissement de postures, considérées parfois comme peu compatibles (Bednarz, 2015), a pu se réaliser aux deux conditions suivantes : nous avons gardé constamment en vue l'objet de la recherche et nous sommes resté fidèle au processus d'investigation requis par la recherche. Nous avons alors adopté une posture d'accompagnateur, c'est-à-dire d'un chercheur qui s'implique dans le projet pédagogique qu'il accompagne (Cuche et al., 2012), dans la mesure où nous avons aidé chaque enseignant à cheminer dans son parcours individuel au regard de ses questions, mais aussi à gérer les tensions, craintes, voire découragements auxquels chacun a fait face.

Cette posture s'est avérée nécessaire vu l'expérimentation collaborative qu'ont menée les enseignants dans le développement de leurs pratiques évaluatives, et dans le souci d'approfondir le « pour quoi » et le « comment » de ces pratiques d'un point de vue spontané, puis progressivement à l'aide d'un modèle théorique (Boucenna et Charlier, 2012) puisque « pour être formatrice, la pratique doit être théorisée » (Charlier, 2012, p. 130).

Les enseignants participant à notre recherche collaborative

Trois enseignants de mathématiques et trois de français du secondaire (élèves de 13 à 16 ans) ont pris part à cette recherche collaborative[4]. Ils ont été recrutés sur une base de volontariat en contactant divers établissements scolaires où nous avions mené des actions de formation continue. Leur expérience variait de 3 à 15 ans, mais aucun d'entre eux n'avait participé à une recherche de ce type. Tous étaient titulaires d'un titre d'enseignement reconnu, mais personne n'avait suivi de formation continue récente en évaluation.

Les enseignantes de langue travaillaient dans le même établissement. Deux enseignants de mathématiques exerçaient dans un établissement différent, tandis que le troisième provenait d'une autre école présentant des caractéristiques socioéconomiques similaires. Du point de vue de l'évaluation sommative, ils étaient légalement tenus d'évaluer leurs élèves par le biais d'épreuves complexes, puis de les certifier à l'aide de notes allant de 1 à 6 avec demi-notes, la note 4 traduisant le seuil de suffisance.

Étant donné que les enseignants ont été invités à partager au plus vite leurs questions avec leurs pairs et le chercheur, notre recherche collaborative s'est ouverte à une complexité du réel exprimée du point de vue de la pratique, qui « est au cœur d'une tension entre un souci de rigueur pour toute démarche méthodologique […] et la nécessité de maintenir une certaine souplesse pour mieux sentir le terrain » (Morrissette et Desgagné, 2009, p. 139). Il s'est agi ici de saisir la pratique dans toutes ses dimensions et, donc, de reconnaître aux praticiens un champ de compétences spécifique à mettre au service de l'objet de recherche (Desgagné et Larouche, 2010). Dès lors, leur développement de compétences a dépendu également en partie de connaissances et de compétences dont ils étaient porteurs, dès le départ, comme nous le montrons dans nos résultats.

Le dispositif et les données

Huit temps ont marqué notre dispositif, lesquels sont synthétisés dans le tableau 1.

Tableau 1
Dispositif de recherche

Temps	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Séances	Entretien individuel 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	Entretien individuel 2
Données	Transcription de l'entretien	1er écrit sur les pratiques	1re épreuve sommative créée par les enseignants Enregistrement des séances	2e écrit sur les pratiques	Enregistrement des discussions des épreuves sommatives réalisées 2e épreuve sommative créée par les enseignants	Enregistrement des discussions des épreuves sommatives réalisées	3e écrit sur les pratiques	Transcription de l'entretien

Une première rencontre a permis de recenser les problèmes que souhaitaient traiter les enseignants dans leurs pratiques évaluatives, notamment du point de vue des contenus disciplinaires en mathématiques et en français. Le temps 1 a été consacré à des entretiens compréhensifs (Kaufmann, 2011) individuels permettant de documenter comment les participants parlaient de leurs pratiques évaluatives sommatives en contexte avant d'entrer dans la recherche. Le temps 2 a concrétisé le premier acte d'écriture par la production d'un récit exemplaire de pratique (Desgagné, 2005 ; Pasquini, 2013, 2016).

Cette démarche réflexive s'est reproduite pendant la recherche (temps 4) pour décrire l'évolution de leur questionnement et de leurs acquisitions et à la suite de celle-ci (temps 7) afin de dresser un bilan de leur parcours. Lors du temps 3, les enseignants ont modélisé la cohérence d'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 en respectant leurs habitudes de travail et leur programme.

La deuxième rencontre (temps 5) a eu lieu six semaines plus tard, une fois entre enseignantes de français, une autre entre enseignants de mathématiques, afin d'approfondir les questions liées aux contenus des deux disciplines. Ce laps de temps a permis aux enseignants de terminer la conception de leur épreuve et, parfois, de la mettre en œuvre dans leur classe. Durant cette deuxième séance, la première épreuve a été discutée entre le chercheur et les enseignants, puis ces derniers ont commencé à construire une seconde épreuve.

Deux mois plus tard, le temps 6 a consisté en une dernière rencontre, en grand groupe, durant laquelle a été réitéré le processus de discussion de chaque épreuve. Parallèlement au dernier écrit de pratique (temps 7), les enseignants ont été conviés au second entretien compréhensif (temps 8) centré sur l'évolution de leurs pratiques depuis le départ.

L'entièreté du dispositif a duré sept mois. Quatre types de données ont ainsi été récoltés : les épreuves sommatives réalisées, les enregistrements des échanges lors des discussions de ces épreuves pendant les rencontres, deux entretiens compréhensifs et trois écrits exemplaires de pratique.

La méthode d'analyse

L'analyse des données issues de notre dispositif a permis de documenter et de comprendre l'impact d'un modèle théorique sur le développement des compétences en évaluation des enseignants, dans des situations d'évaluation réelles référées à des pratiques de classe contextualisées.

Afin de dégager des éléments qualitatifs de nos données, nous avons opté pour une analyse à l'aide de catégories conceptualisantes, définies comme « une production textuelle se présentant sous la forme d'une brève expression et permettant de dénommer un phénomène perceptible à travers une lecture conceptuelle d'un matériau de recherche » (Paillé et Mucchielli, 2012, p. 316). Toute catégorie désignant un phénomène, nous avons considéré les composantes de notre modèle théorique comme étant proches de catégories. Par exemple, la tâche évaluative joue un rôle structurant dans la compréhension des phénomènes évaluatifs par le praticien, tout en demandant au chercheur d'incarner l'attribution de signification des phénomènes relatifs au concept de tâche pour comprendre les pratiques. Aussi, nous avons mis en dialogue les composantes de notre modèle théorique avec nos données, ce qui nous a permis de construire cinq catégories : rapport au référentiel, expression du contenu, élaboration des tâches évaluatives, pratiques de pondération et pratiques de notation.

Nous avons alors croisé nos cinq catégories avec les six compétences élaborées au regard des apports de la recherche, des questions des praticiens et de leur contexte. Les catégories conceptualisantes ont ainsi permis de documenter l'évolution du développement des différentes compétences. En complément, nous avons adopté une démarche de triangulation méthodologique (Silverman, 2009) pour contrôler les biais éventuels d'interprétation, dans une recherche de validité. Nous avons alors considéré la triangulation comme une stratégie inductive de recherche, car elle privilégie et valorise le fait de partir d'un phénomène particulier issu des pratiques pour le décrire et le comprendre, ce qui est pertinent avec la logique des trois phases de la recherche collaborative décrites auparavant (cosituation, coopération et coproduction).

Dans cette idée, la triangulation s'est assimilée à une procédure de va-et-vient dynamique entre opérations de recueil, d'analyse et de conceptualisation – qui renvoie à l'analyse des données par le biais des catégories conceptualisantes – et au rapport de cohérence que cette analyse a

entretenu avec nos données qualitatives. Cette triangulation s'est dès lors basée sur le principe de réinjection, prenant lui-même la forme de questionnements, d'hypothèses et d'observations.

Résultats

Nous présentons les résultats qui documentent le développement des compétences en évaluation de chaque enseignant dans son parcours individuel. Pour chacun d'entre eux, un tableau (voir Tableaux 2 à 7) synthétise l'évolution des six compétences avant et après la recherche, en fonction de la discipline. Les éléments significatifs ont été sélectionnés et exemplifiés à l'aide d'extraits de protocoles, signalés entre guillemets dans les tableaux. Les contenus des épreuves qu'ils ont conçues sont symbolisés par les codes E1 et E2. De par le fait que les compétences E et F portaient sur la mobilisation des apports proposés, les indices de leur développement n'apparaissent que dans la seconde partie de tableau. Quelques lignes résument ensuite les changements saillants chez chaque participant en précisant sur quoi portait son questionnement de départ. Une synthèse clôt cette section en mettant en évidence les points clés.

Les enseignants de mathématiques

Fernand

Au début de la recherche, Fernand explique que son plus grand problème réside dans la pondération : « *Dans une évaluation, quel poids donner sur chaque partie, sur chaque objectif que j'évalue ? Pour moi, c'est le plus difficile.* » À la fin, nous observons que le plan d'études prend davantage de sens avec un rôle central donné aux contenus disciplinaires et aux objectifs, qui déterminent par ailleurs l'élaboration des tâches évaluatives. L'attribution des points ne suit plus une logique mathématique, mais fait référence à l'importance des apprentissages évalués à partir de critères. Le seuil de suffisance (note = 4) est établi en référence aux apprentissages, mais le barème fédéral[5] est toujours utilisé pour établir le seuil des autres notes. Fernand peut justifier la cohérence de ses épreuves, même s'il peine encore à éviter des désalignements de la notation et s'il ne parvient pas à apparier une évaluation critériée et une note référée à l'apprentissage : « *Encore dans l'évaluation d'hier, j'ai essayé de mettre quelques critères, et puis je les ai gardés pour moi… Comme on regarde un tableau de maître qu'on a dans son coffre (rire).* »

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 référé aux unités d'apprentissage pour l'élaboration du seuil (note = 4). Puis, utilisation d'un barème standardisé pour les autres notes	Désalignements curriculaires encore difficiles à gérer en ce qui a trait la notation	Choix le plus souvent assumé, même si le système critérié et une note référée à l'apprentissage sont considérés comme incompatibles

Paul

Le problème principal que Paul soulève dans son premier entretien et son premier écrit est celui des barèmes de notation, notamment en matière d'exploitation du barème fédéral : « *Cet aspect mécanique de l'évaluation ne me plaît pas du tout* […]*, car on utilise rarement l'évaluation pour que ça serve à l'élève.* » En fin de recherche, de manière générale, nous observons cependant que ses pratiques évaluatives se réfèrent de plus en plus aux contenus disciplinaires et aux objectifs : le plan d'études a gagné en clarté, et l'élaboration des tâches évaluatives, l'attribution des points et la construction de la note suivent davantage une réflexion centrée sur les apprentissages. Malgré un gain d'assurance dans ses choix, il subsiste toutefois chez Paul des espaces d'inconfort concernant la construction de la note, où des compromis peu convaincants sont toujours réalisés : « *Le barème fédéral, 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

Le souhait de Philémon serait d'arriver à des évaluations plus « justes » en se concentrant sur la pondération, jugée selon lui délicate à opérationnaliser : « *Est-ce que je suis vraiment en train d'évaluer chez l'élève une acquisition de connaissances ou est-ce que je suis en train d'évaluer tout le temps les mêmes choses ?* » Au terme du processus, les objectifs et les contenus prennent plus d'importance dans ses pratiques évaluatives : le plan d'études devient une référence pour déterminer les attentes, et c'est autour d'une réflexion sur les critères que l'élaboration des tâches évaluatives – de complexité plus homogène – et la construction de la note suffisante 4 ont lieu. Il subsiste toutefois encore des problèmes à construire les autres notes en référence à ces mêmes critères : « *Je ne suis pas convaincu que faire fi des points, ça va me permettre d'être meilleur dans ma cohérence.* » Cela dit, Philémon assume ses choix en matière d'évaluation.

Les enseignantes de français

Alizée

Alizée éprouve trois problèmes. Le premier est l'évaluation de certains contenus comme l'expression orale, tandis que le deuxième consiste en l'exploitation des grilles de critères : « *Quel poids on met à chaque partie* […]*, quel équilibre on donne à chaque partie ?* » Le troisième problème est relatif à la notation : « *Comment et quand fixer nos échelles ?* » À la fin de la recherche, en référence aux deux épreuves construites avec sa collègue

Tableau 3
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 en ce qui a trait à la pondération et la notation	Rapport critique au référentiel, choix des tâches assumé En revanche, utilisation de critères jugée impossible Notation sur la base de compromis acceptables, mais peu convaincants

Tableau 4
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 : « tu 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 (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 du seuil (note = 4) en rapport avec les contenus du plan d'études issue d'un raisonnement référé aux apprentissages Utilisation d'un barème standardisé pour les autres notes	Désalignements curriculaires encore présents et problèmes en matière de notation	Choix le plus souvent assumés, même si le système critérié et une note référée à l'apprentissage sont considérés comme peu compatibles

Tableau 5
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 (p. ex., 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 du seuil (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 assumé et référé à une taxonomie Adoption du système critérié Critique, voire refus des directives en matière de notation

directe Anouck, les changements majeurs se situent dans la place prépondérante que prennent désormais les contenus et les objectifs, référés à la taxonomie d'Anderson et Krathwohl (2001). Son approche critériée, totalement cohérente d'un point de vue d'alignement curriculaire élargi, lui permet d'assumer ses choix et de refuser certaines directives relatives à la notation considérées comme apédagogiques, par exemple la linéarité des échelles. Toutefois, elle reste toujours en questionnement sur la transposition de telles pratiques à d'autres contenus : « *La compréhension orale… je ne sais toujours pas comment l'évaluer.* »

Anouck

Le questionnement initial d'Anouck se situe par rapport aux contenus à évaluer : « *Je me rends compte qu'on ne sait pas toujours ce qu'on attend, finement, quand on fait un test. Enfin, qu'est-ce qu'on veut ? Si on veut juste qu'ils sachent écrire, ou bien s'ils doivent avoir atteint des objectifs plus précis à l'intérieur ?* » Au terme de notre collaboration, les changements dans ses pratiques sont de plusieurs ordres : les contenus sont désormais centraux dans sa manière d'envisager toute démarche d'évaluation sommative, et la taxonomie est aussi une référence pour formuler et analyser les objectifs. Anouck se sent enfin délivrée du phénomène de compensation des points qui lui posait problème, mais elle reste en questionnement sur la construction des notes autres que le seuil de suffisance (note = 4) : « *La grande question que j'ai encore, c'est celle de l'échelle. […] Je me sens légitimée dans ma manière de placer le 4. Je le fais facilement. Après, il reste le problème d'en dessus et d'en dessous.* »

Samia

Dès la première rencontre, Samia exprime un malaise dans ses pratiques qui la font se demander si elle est « juste » avec les élèves : « *Des textes relativement pauvres en imagination ou intérêt obtiennent quand même la note 4 grâce aux différents objectifs. Et parfois des textes assez mauvais en syntaxe et structuration obtiennent le 4 grâce à la richesse des idées ou simplement au fait d'avoir suivi scrupuleusement les objectifs.* » Cela étant, en fin de recherche, des changements clés ont eu lieu : le plan d'études est devenu une référence pour évaluer, et les tâches évaluatives se réfèrent aux contenus et ne dépendent plus d'une habitude mélangeant technique et compétences langagières. La note suffisante est établie en fonction d'attentes relatives aux apprentissages menés. En revanche, des doutes subsistent quant à l'opérationnalisation de la cohérence au sein des

Tableau 6
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 du seuil (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 du seuil (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 assumé et référé à une taxonomie Adoption du système critérié Critique, voire refus des directives en matière de notation

Tableau 7
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 la technique de la langue et sur une compétence langagière	Attribution des points en fonction de « ce qui est attendu minimalement »	Construction du seuil (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 du seuil (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 référé aux apprentissages Désalignements curriculaires dus à la compensation des points	Désalignements curriculaires évités jusqu'au choix des tâches évaluatives, probables dans la pondération, puis omniprésents dans la notation	Interprétation du plan d'études proche de la source, choix assumé des contenus et des tâches, seuil (note = 4) en rapport avec les apprentissages Compensation des points omniprésente

épreuves : « *Il y a encore une difficulté. Quand je fais mes évaluations, ça me prend toujours beaucoup de temps parce que je dois encore ressortir (rire) le schéma de l'alignement…* »

Synthèse

Au début de la recherche, les enseignants des deux disciplines sont aux prises avec des problèmes variés, mais parfois partagés : le plan d'études est au mieux exploité de loin comme fil rouge, au pire est mis de côté, car considéré comme peu lisible pour évaluer. Les tâches évaluatives des épreuves obéissent à des principes établis, mais rarement étayés : mélange d'exercices techniques et de problèmes en mathématiques, évaluation systématique de la technique de la langue dans des épreuves de production et de compréhension écrites en français. Toutefois, leurs interrogations se rejoignent pleinement sur deux sujets : 1) la pondération en points souffre de manque de réflexion sur les apprentissages et débouche systématiquement sur une logique de compensation perçue comme délicate et 2) l'utilisation de critères est problématique ; la notation s'opère à l'aide de barèmes standardisés ou de manière à correspondre à certains pourcentages de points, obéissant ainsi à des procédures jugées peu satisfaisantes.

À la fin de la recherche, tous ont développé des compétences sur plusieurs plans. Relativement à la compétence A, les enseignants sont capables de lire et d'exploiter le curriculum de manière critique en donnant une importance aux contenus et à la complexité des objectifs. Tous les enseignants, sauf dans une certaine mesure Samia, élaborent désormais des tâches évaluatives de niveau de complexité homogène qui sont alignées avec la complexité des objectifs évalués (compétence B). Il est ensuite intéressant de différencier le développement des compétences C et D en fonction de la discipline relativement aux questions plus prégnantes. Alors que les enseignants de mathématiques restent toujours en difficulté pour pondérer explicitement leurs épreuves à l'aide de critères, Anouck et Alizée, en français, y parviennent sur des épreuves de production écrite. Logiquement, pour une majorité d'entre eux, une construction de note référée à l'apprentissage par le biais de ces critères et alignée curriculairement reste toutefois délicate, voire inconcevable (compétence E) : nous observons que les enseignants de mathématiques sont tous aux prises avec une impossibilité d'éviter des désalignements curriculaires dans la notation, ce qui n'est pas le cas de deux enseignantes de français. Enfin, relativement à la compétence F, nous voyons que tous les participants

sont en mesure de justifier leurs choix évaluatifs, et ce, même si certaines questions de départ n'ont toujours pas de réponse, comme chez Fernand, Paul, Anouck et Alizée.

Discussion

Notre question de recherche visait à comprendre en quoi notre recherche collaborative pouvait contribuer au développement de compétences en évaluation sommative au service de démarches de certification cohérentes. Nous observons qu'un développement de compétences est présent chez tous les enseignants et qu'il est favorisé par les multiples espaces dédiés à l'activité réflexive (Bednarz, 2015). À partir des compétences de départ des uns et des autres, de nouvelles réflexions ont émergé, comme si le questionnement permis par le modèle théorique proposé, inscrit dans de multiples possibilités de travailler à partir des pratiques réelles, permettait une explicitation d'interrogations ancrées souvent dans des habitudes relativement figées (Pasquini, 2018).

Le développement des compétences que ces réflexions ont rendu possible – bien qu'irrégulier, non linéaire et dépendant de multiples éléments contextuels et disciplinaires – est selon nous significatif. En effet, nous observons qu'il existe une corrélation entre la discussion progressive de la cohérence des pratiques entre les enseignants à la lumière du modèle théorique et le développement des six compétences. Même si des questions restent en suspens et mériteraient plus de travail, ces résultats tendent à confirmer que l'activité réflexive gagne à s'appuyer sur des éléments théoriques possiblement mobilisables en contexte (Charlier, 2012).

Plus précisément, nous observons que c'est le développement des compétences A et B qui est le plus marquant chez tous les participants : leurs réflexions donnent désormais une place prépondérante aux contenus et aux objectifs dans une perspective systémique, soulignant ainsi la validité de leurs pratiques évaluatives (Bonner, 2013).

En revanche, la compétence C se développe de manière contrastée en fonction des disciplines et des enseignants. Cela confirmerait que les contenus jouent un rôle déterminant dans la formulation des critères d'évaluation (p. ex., Jönsson, 2014 ; Mottier Lopez et Pasquini, 2017 ; Prøitz, 2013), jusque dans les possibilités, ou non, de construire une note référée à l'apprentissage, ce qui encourage à privilégier des dispositifs de recherche

donnant une place prépondérante aux spécificités disciplinaires. À ce titre et relativement à la compétence D, notons que tous les enseignants, en début de recherche, se demandaient comment noter « juste » et que cette question reste délicate chez bon nombre d'entre eux à la fin du processus. Serait-ce à dire, comme le suggèrent par exemple McMillan et Nash (2000), que les pratiques de notation méritent une attention particulière en recherche ?

Quant aux compétences E et F, plus réflexives, nous observons que la majorité des enseignants sont en mesure de justifier la cohérence de leurs pratiques avec des arguments pédagogiques, même s'ils sont toujours aux prises avec des compromis (Carless, 2014), notamment lorsqu'il est question de faire cohabiter une recherche de cohérence curriculaire avec des spécificités disciplinaires ou des prescriptions ressenties parfois comme (trop) fortes.

Ces divers résultats, bien que difficiles à généraliser, soulignent la dimension dynamique du développement des compétences saisi dans notre recherche collaborative ainsi que quelques axes de recherche à développer. Nous y revenons en fin de texte.

Conclusion

Notre recherche collaborative a poursuivi une finalité heuristique dans un souci de décrire et de comprendre les pratiques. Toutefois, elle a également été fidèle à un souci pragmatique en proposant aux enseignants de nouvelles ressources en vue d'amener leurs pratiques évaluatives vers plus de cohérence, puis de développer leurs compétences (Lefeuvre et al., 2009). Ces finalités traduisent une perspective professionnalisante de développement professionnel, entendu comme :

> un processus graduel d'acquisition et de transformation des compétences et des composantes identitaires conduisant progressivement les individus et les collectivités à améliorer, enrichir et actualiser leur pratique, à agir avec efficacité et efficience dans les différents rôles et responsabilités professionnelles qui leur incombent, à atteindre un nouveau degré de compréhension de leur travail et à s'y sentir à l'aise (Mukamurera, 2014, p. 12).

Plus précisément, notre dispositif a suivi trois principes déterminants pour un développement professionnel efficace : le développement professionnel s'est inscrit dans une démarche d'accompagnement reposant

sur un travail collaboratif ; il a été distribué dans le temps en fonction des possibles des participants ; et les activités proposées aux enseignants ont été étayées par des apports de la recherche (Richard et al., 2017). Sa perspective professionnalisante a relevé d'un processus d'apprentissage provoqué par une intention de formation explicite (Lefeuvre et al., 2009 ; Timperley, 2011), processus concrétisé par la référence au modèle théorique de l'alignement curriculaire élargi pour penser la cohérence des pratiques évaluatives (Pasquini, 2018, 2019). Ce choix a inscrit notre recherche collaborative dans une épistémologie qui a grandement influencé la nature et le rôle de l'activité réflexive. Notons également qu'en nous appuyant sur le principe de triangulation, nous avons pu en approfondir la conceptualisation en exploitant nos quatre types de données.

Nos deux hypothèses de départ se voient ainsi confirmées. Toutefois, au moins deux réflexions subsistent, ce qui souligne les limites de notre recherche. La première interroge le rôle du modèle théorique sur l'activité réflexive et sur le développement des compétences. En quoi le modèle les a-t-il orientés ? Dans quelle mesure le recours à l'alignement curriculaire élargi comme outil compréhensif des pratiques a-t-il induit un effet de désirabilité sociale (Leclerc et al., 2011), poussant les enseignants à se conformer à nos attentes implicites ? En d'autres termes, quel aurait été l'impact d'un accompagnement où les apports auraient été coconstruits avec les participants au fur et à mesure de leur cheminement sur le développement de leurs compétences ?

La seconde réflexion, qui prolonge la précédente, revient à penser l'activité réflexive dans une autre méthode de recherche visant également un développement professionnel (Dionne et al., 2010) et, surtout, comme nous l'avons montré, lorsque certains problèmes rencontrés par les enseignants « résistent » aux cadres interprétatifs offerts par les modèles théoriques et aux espaces de réflexivité proposés. Nous pensons par exemple aux communautés d'apprentissage (Savoie-Zajc, 2010), où les durées sont plus conséquentes et les activités réflexives encore plus articulées avec les réalités de classes, donc plus propices au développement des compétences (Guskey et Yoon, 2009).

L'impact des méthodes de recherche sur le développement professionnel ne peut être jugé de manière univoque (Richard et al., 2017 ; Timperley, 2011). Toutefois, ces réflexions nous semblent importantes pour penser des dispositifs de recherche en évaluation tenant compte de facteurs

clés qui ont fortement un impact sur les pratiques comme les contextes, les prescriptions ou les contenus disciplinaires, et pour encourager des formes collectives de développement professionnel permettant de créer des espaces de collaboration et de coapprentissage (Dionne et al., 2010). Notre recherche est modestement allée dans ce sens. Or, beaucoup de travail reste à faire quant aux pratiques d'évaluation sommative, dont la cohérence demeure trop souvent insuffisante (Moss, 2013). Ce travail semble encore plus urgent en matière de pratiques de notation, qui constituent un espace d'important questionnement professionnel lorsqu'il s'agit de construire une note référée à l'apprentissage (Brookhart, 2017).

Réception : 7 janvier 2020

Version finale : 1er décembre 2020

Acceptation : 4 décembre 2020

NOTES

1. Cet article est issu d'une recherche doctorale menée à terme.
2. Dans le présent article, les mots de genre masculin appliqués aux personnes désignent les hommes et les femmes.
3. La définition de l'évaluation sommative que nous suivons est précisée dans le cadre conceptuel.
4. Nous les présentons à l'aide de prénoms d'emprunt.
5. Barème standardisé correspondant au calcul suivant : *(n^bre de pts cumulés / n^bre de pts total x 5) + 1.*

RÉFÉRENCES

Airasian, W., & 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

Altet, M. (2000). L'analyse de pratiques, une démarche de formation professionnalisante ? *Recherche et formation*, *35*, 25-41. https://www.persee.fr/doc/refor_0988-1824_2000_num_35_1_1668

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

Bateman, D., Taylor, S., Janik, E., & Logan, A. (2009). Curriculum coherence and student success. *Pédagogie collégiale*, *22*(5), 8-18. https://aqpc.qc.ca/sites/default/files/revue/Bateman.pdf

Bednarz, N. (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, avril). Aligning teaching and assessing to course objectives [communication orale]. *Teaching and Learning in Higher Education : New Trends and Innovations*, Aveiro, Portugal. https://www.dkit.ie/system/files/Aligning_Reaching_and_Assessing_to_Course_Objectives_John_Biggs.pdf

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

Boucenna, S. et Charlier, E. (2012). Accompagnement d'équipes d'enseignants, entre recherche et développement professionnel. Dans E. Charlier et S. Biemar (dir.), *Accompagner : un agir professionnel* (p. 127-138). De Boeck.

Braxmeyer, N., Guillaume, J.-C. et Levy, J.-F. (2005). Enseignement scolaire : les pratiques d'évaluation des enseignants en collège, *Les dossiers*, *160*. Ministère de l'*Éducation* nationale, de l'Enseignement supérieur et de la Recherche de France, Direction de l'évaluation et de la prospective. http://media.education.gouv.fr/file/50/7/2507.pdf

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. (2017). *How to use grading to improve learning*. ASCD.

Carless, D. (2014). Exploring learning-oriented assessment processes. *Higher Education*, *69*, 963-976. https://doi.org/10.1007/s10734-014-9816-z

Charlier, E. (2012). Former des enseignants professionnels pour une formation continuée articulée à la pratique. Dans L. Paquay, M. Altet, E. Charlier et P. Perrenoud (dir.), *Former des enseignants professionnels : quelles stratégies ? quelles compétences ?* (p. 119-143). De Boeck.

Cuche, C., Canivet, C. et Donnay, J. (2012). L'accompagnement, source de création de savoirs. Dans E. Charlier et S. Biemar (dir.), *Accompagner un agir professionnel* (p. 139-152). De Boeck.

Desgagné, S. (1997). Le concept de recherche collaborative : l'idée d'un rapprochement entre chercheurs universitaires et praticiens enseignants. *Revue des sciences de l'éducation*, *23*(2), 371-393. https://doi.org/10.7202/031921ar

Desgagné, S. (2005). *Récits exemplaires de pratique enseignante : analyse typologique*. PUQ.

Desgagné, S. (2007). Le défi de coproduction de « savoir » en recherche collaborative : autour d'une démarche de reconstruction et d'analyse de récits de pratique enseignante. Dans M. Anadon (dir.), *La recherche participative : multiples regards* (p. 89-121). PUQ.

Desgagné, S., Bednarz, N., Lebuis, P., Poirier, L. et Couture, C. (2001). L'approche collaborative de recherche en éducation : un rapport nouveau à établir entre recherche et formation. *Revue des sciences de l'éducation*, *27*(1), 33-64. https://doi.org/10.7202/000305ar

Desgagné, S. et Larouche, H. (2010). Quand la collaboration de recherche sert la légitimation d'un savoir d'expérience. *Recherches en éducation*, *1*, 7-18. http://www.unige.ch/fapse/SSE/teachers/barthassat/REE-HS-no1.pdf

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

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

Gauthier, C., Mellouki, M., Bissonnette, S. et Richard, M. (2005). Écoles efficaces et réussite scolaire des élèves à risque *: un état de la recherche* [Rapport de recherche]. CRIFPE, Université Laval.

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

Hadji, C. (2016). Le système éducatif peut-il passer de l'évaluation normative à une évaluation constructive ? Dans B. Bocquet (dir.), *La fièvre de l'*évaluation*: quels symptômes ? quels traitements ?* (p. 113-135). Presses universitaires du Septentrion.

Hadji, C. (2017). Savoir mettre en œuvre une évaluation constructive. Dans A. Bentolila (dir.), *L'essentiel de la pédagogie* (p. 229-250). Nathan.

Jönsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment and Evaluation in Higher Education*, *39*(7), 840-852. https://doi.org/10.1080/02602938.2013.875117

Jorro, A. (2013). L'accompagnement des enseignants dans l'activité évaluative face à des situations de production écrite. *Revue française de linguistique appliquée*, *XVIII*, 107-116. https://doi.org/10.3917/rfla.181.0107

Jorro, A. (2015, 4 septembre). *Des compétences aux gestes évaluatifs des enseignants* [vidéo]. Vimeo. https://www.hepl.ch/cms/accueil/actualites-et-agenda/actu-hep/les-conferences-du-colloque-geva.html

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

Koretz, D. (2017). *The testing charade: Pretending to make schools better*. University of Chicago Press.

Le Boterf, G. (2011). Qu'est-ce qu'un professionnel compétent ? Comment développer son professionnalisme ? *Pédagogie collégiale*, *24*(2), 27-31. https://aqpc.qc.ca/sites/default/files/revue/LeBoterf-24-2.pdf

Leclerc, C., Bourassa, B., Picard, F. et Courcy, F. (2011). Du groupe focalisé à la recherche collaborative : avantages, défis et stratégies. *Recherches qualitatives*, *29*(3), 145-167. http://www.recherchequalitative.qc.ca/documents/files/revue/edition_reguliere/numero29(3)/RQ_29(3)_Leclerc_et_al.pdf

Lefeuvre, G., Garcia, A. et Namolovan, L. (2009). Les indicateurs de développement professionnel. *Questions vives : Recherches en éducation*, *5*, 277-314. https://doi.org/10.4000/questionsvives.627

Maroy, C. et Cattonar, B. (2002). Professionnalisation ou déprofessionnalisation des enseignants ? Le cas de la communauté française de Belgique. *Cahier de recherche en éducation et formation*, *18*, 4-26. https://halshs.archives-ouvertes.fr/halshs-00603557/document

Martinez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics : Evidence in the ECLS. *Educational Assessment*, *14*, 78-102. https://doi.org/10.1080/10627190903039429

Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Effective Grading Practices*, *69*(3), 34-39.

McMillan, J. H., & Nash, S. (2000, avril). *Teacher classroom assessment and grading practices decision making*. Communication présentée au NCME Annual Meeting, Nouvelle-Orléans, LA. https://files.eric.ed.gov/fulltext/ED447195.pdf

Morrissette, J. et Desgagné, S. (2009). Les jeux des positions de savoir en recherche collaborative : une analyse des points de vue négociés d'un groupe d'enseignantes du primaire. *Recherches qualitatives*, *28*(2), 118-144. http://www.recherche-qualitative.qc.ca/documents/files/revue/edition_reguliere/numero28(2)/morrissette(28)2.pdf

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

Mottier Lopez, L. (2014). L'évaluation pédagogique va-t-elle enfin marcher sur ses deux pieds ? Les enseignements de l'histoire récente de l'école genevoise. Éducation *et francophonie*, *XLII*(3), 85-101. https://doi.org/10.7202/1027407ar

Mottier Lopez, L. (2015). *Évaluations formative et certificative des apprentissages : enjeux pour l'enseignement*. De Boeck.

Mottier Lopez, L. et Figari, G. (dir.). (2012). *Modélisations de l'évaluation en éducation.* De Boeck.

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

Mukamurera, J. (2014). Le développement professionnel et la persévérance en enseignement : éclairage théorique et état des lieux. Dans L. Portelance, S. Martineau et J. Mukamurera (dir.), *Développement et persévérance professionnels en enseignement : oui, mais comment ?* (p. 23-48). PUQ.

Nusche, D., Radinger, T., Santiago, P., & Shewbridge C. (2013). *Synergies for better learning : An international perspective on evaluation and assessment*. OECD Reviews of Evaluation and Assessment in Education. OCDE.

Osborn, M. (2002). Le métier d'enseignant en période de réforme : le cas anglais. *Revue internationale d'éducation*, *30*, 27-39. https://doi.org/10.4000/ries.1852

Osborn, M. (2006). Changing the context of teachers' work and professional development : An European perspective. *International Journal of Educational Research*, *45*, 242-253. https://doi.org/10.1016/j.ijer.2007.02.008

Paillé, P. et Mucchielli, A. (2012). *L'analyse qualitative en sciences humaines et sociales*. Armand Colin.

Paquay, L. (2012). Continuité et avancées dans la recherche sur la formation des enseignants. *Les cahiers de recherche du GIRSEF*, *90*, 1-36. https://hal.archives-ouvertes.fr/hal-00978949/document

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

Pasquini, R. (2013). *Quand les récits de pratique enseignante parlent d'apprentissage*. PUL. http://hdl.handle.net/20.500.12162/1601

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. (2017). Évaluation : quatrième question vive de la pratique. Éducateur, *7*, 34-35. https://orfee.hepl.ch/bitstream/handle/20.500.12162/1607/Pasquini.2017.07.profession-evaluation.pdf ?sequence=1&isAllowed=y

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]. Archives ouvertes. https://archive-ouverte.unige.ch/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. (2021). *Quand la note devient constructive : évaluer pour certifier et soutenir les apprentissages*. PUL.

Perrenoud, P. (1996). Le métier d'enseignant entre prolétarisation et professionnalisation : deux modèles du changement. *Perspectives*, *XXVI*(3), 543-562. https://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_1996/1996_16.html

Perrenoud, P. (2004). Évaluer des compétences. *Éducateur*, 8-11. http://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_2004/2004_01.pdf

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

Prøitz, T. S. (2013). Variations in grading practice : Subjects matter. *Education Inquiry*, *4*(3), 555-575. https://doi.org/10.3402/edui.v4i3.22629

Rayou, P. (2008). Ni guerre, ni paix : tensions et malentendus dans la formation. Dans P. Perrenoud, M. Altet, C. Lessard, et L. Paquay (dir.), *Conflits de savoirs en formation des enseignants : entre savoirs issus de la recherche et savoirs issus de l'expérience* (p. 77-89). De Boeck.

Richard, M., Carignan, I., Gauthier, C. et Bissonnette, S. (2017). *Quels sont les modèles de formation continue les plus efficaces pour l'enseignement de la lecture et de l'écriture chez les élèves du préscolaire, du primaire et du secondaire ? Une synthèse des connaissances* (publication nº 2015-AP-187763). Gouvernement du Québec, MEES, Fonds de recherche du Québec Société et culture. http://www.frqsc.gouv.qc.ca/documents/11326/448958/AP_20142015_Richard.M_rapport_lecture-ecriture.pdf.pdf/70d9d12a-d844-4c6f-9239-5e3e3b4bb787

Rieg, S. A. (2007). Classroom assessment strategies : What do students at-risk and teachers perceive as effective and useful? *Journal of Instructional Psychology*, *34*(4), 214-225. https://www.questia.com/library/journal/1G1-173375578/classroom-assessment-strategies-what-do-students

Sadler, R. D. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. Dans G. Joughin (dir.), *Assessment, learning and judgement in higher education* (p. 1-15). Springer.

Savoie-Zajc, L. (2010). Les dynamiques d'accompagnement dans la mise en place de communautés d'apprentissage de personnels scolaires. Éducation *et formation*, *293*, 9-20.

Silverman, D. (2009). *Doing qualitative research*. SAGE Publications.

Timperley, H. (2011). Le développement professionnel des enseignants et ses effets positifs sur les apprentissages des élèves. *Revue française de pédagogie*, *174*, 31-40. https://doi.org/10.4000/rfp.2910

Van Nieuwenhoven, C. et Colognesi, S. (2015). Une recherche collaborative sur l'accompagnement des futurs instituteurs : un levier de développement professionnel pour les maîtres de stage. *Évaluer : Journal international de recherche en éducation et formation*, *1*(2), 103-121. https://dial.uclouvain.be/pr/boreal/object/boreal:165239/datastream/PDF_01/view

Vial, M. (2012). *Se repérer dans les modèles de l'évaluation*. De Boeck.

Vinatier, I. et Morrissette, J. (2015). Les recherches collaboratives : enjeux et perspectives. *Carrefours de l'éducation*, *1*(39), 137-170. https://doi.org/10.3917/cdle.039.0137

Wiggins, G., & McTighe, J. M. (2005). *Understanding by design*. ASCD.