

Les épreuves externes comme soutien au développement professionnel des enseignants dans leurs pratiques d'évaluation certificative : les apports d'une recherche collaborative

Walther Tessaro

Université de Genève

Walther.Tessaro@unige.ch

Résumé

Parmi les repères qui peuvent participer aux systèmes de référents pour l'évaluation des apprentissages des élèves, les épreuves externes sont des instruments significatifs mais dont les effets sont encore peu connus. A partir de documents produits par des enseignants et d'outils proposés par des chercheurs-formateurs, ainsi que des verbalisations en cours de correction et des échanges lors de dispositifs de modération sociale (Mottier Lopez, Tessaro & Filliettaz, 2012), nous avons examiné le rôle de ce type d'épreuves lors de la construction d'épreuves certificatives par les enseignants et lors de la correction de productions aboutissant à une notation. Nous nous sommes aussi penchés sur les formes que peuvent prendre les contributions des enseignants à notre recherche collaborative, reposant sur une relation réciproque entre acteurs de terrain et chercheurs et visant respectivement le développement professionnel et la production de savoirs scientifiques.

Si les épreuves externes semblent soutenir l'amélioration des pratiques d'évaluation des enseignants, il reste à réfléchir à un usage adéquat dans des dispositifs de formation continue. La modération sociale pratiquée au sein d'un établissement apparaît dès lors comme une modalité et un niveau d'action intéressants. Les consensus qui ont émergé entre les enseignants, dont certains ont été directement ou indirectement induits par les épreuves externes, semblent avoir favorisé leur développement professionnel et devraient permettre un réinvestissement dans les pratiques ainsi que l'intériorisation de nouveaux savoirs professionnels.

Mots-clés

pratiques d'évaluation certificative, recherche collaborative, épreuves externes, alignement, soutien au développement professionnel

Pour citer cet article : Tessaro, W. (2015). Les épreuves externes comme soutien au développement professionnel des enseignants dans leurs pratiques d'évaluation certificative : les apports d'une recherche collaborative. *Evaluer. Journal international de Recherche en Education et Formation*, 1(2), pp. 49-65.

1. Introduction

Bien que l'évaluation des apprentissages des élèves soit une des difficultés majeures rencontrées par les enseignants novices (De Stercke, 2010), elle reste généralement peu considérée dans les programmes de formation initiale et continue. Outre les apports théoriques du champ de l'évaluation et de la régulation, les formateurs ont à prendre en compte les aspects contextuels des pratiques évaluatives des enseignants, qui participent à une culture commune. En font partie les prescriptions de l'institution scolaire, mais aussi de nombreux autres repères auxquels les enseignants peuvent se référer et qui soutiennent les manières de faire admises ou partagées (Morrissette, 2010).

Parmi ces repères qui peuvent participer aux systèmes de référents pour l'évaluation des apprentissages des élèves (Figari, 2006) lors de l'appréciation et de l'interprétation des informations recueillies par les enseignants, les épreuves externes sont des sources significatives. En effet, comme le souligne Rosenkvist (2010), si les résultats des élèves à ces épreuves permettent d'identifier les besoins d'apprentissage, ils sont aussi souvent utilisés comme des indicateurs pour améliorer les pratiques de classe.

Dans les systèmes scolaires européens et nord-américains, l'augmentation progressive des épreuves externes a été caractérisée par une attention aussi grandissante pour les conséquences de leur passation sur l'enseignement. Cet intérêt porte généralement sur les effets des résultats des élèves – notamment dans les régulations formatives qui peuvent en découler en classe – et sur une adaptation du curriculum. De façon plus ambitieuse, des chercheurs comme Barneveld (2008) et Slomp (2008) ont par ailleurs mis en évidence, dans des démarches compréhensives, les dimensions des pratiques d'évaluation des enseignants sur lesquelles les épreuves externes produites par le système scolaire peuvent agir. La recherche de Cooksey, Freebody et Wyatt-Smith (2007) ainsi que celle de Wyatt-Smith et Castleton (2005) se sont quant à elles penchées sur la façon dont le jugement évaluatif articule les composantes du curriculum officiel -dont les épreuves externes- avec d'autres connaissances internes à la classe.

Les effets spécifiques des épreuves externes sur les pratiques d'évaluation certificative des enseignants, conçues comme l'établissement et la communication d'un bilan qui atteste du niveau des acquisitions de l'élève (Wegmuller & Allal, 1997), restent néanmoins encore peu connus. Les seuls résultats de celles-ci, bien qu'encore sous-exploités (Ungerleider, 2007), apparaissent comme insuffisants. Leurs contenus, comprenant les documents destinés aux élèves mais aussi (et surtout) ceux élaborés pour les enseignants, nous semblent cependant d'une grande richesse pour la formation des enseignants et leur développement professionnel. Si les items des épreuves illustrent les attentes institutionnelles, les documents destinés aux enseignants présentent en effet des informations intéressantes telles que les objectifs d'apprentissage évalués, les critères de correction, les consignes de passation, le seuil de réussite et le barème. Ces épreuves peuvent en outre contribuer à une harmonisation des pratiques, sur un plan local ou plus général (Wyatt-Smith, Klenowski & Gunn, 2010).

L'évaluation par les enseignants des comportements et des compétences des élèves en classe étant le meilleur prédicteur des trajectoires de formation des élèves (Guimard, Cosnefroy & Florin, 2007), il nous semble indispensable de proposer un accompagnement qui permette aux enseignants de renforcer des compétences professionnelles à formuler des jugements d'évaluation fiables et équitables, en réduisant notamment les différences entre enseignants et entre écoles. C'est dans ce but que les épreuves externes peuvent offrir selon nous des outils intéressants.

A partir de documents produits par les enseignants et d'outils proposés par des chercheurs-formateurs lors d'une recherche menée dans le cadre d'un fonds national suisse¹ (FNS) (Mottier Lopez et al., 2012), ainsi que des verbalisations en cours de correction et des échanges lors de dispositifs de modération sociale (Wyatt-Smith et al., 2010), nous avons examiné le rôle des épreuves externes du canton de Genève comme possibles référents dans le choix des objectifs et des types de tâches proposés par les enseignants dans l'élaboration de leurs propres épreuves certificatives, ainsi que lors des processus de jugements évaluatifs débouchant sur la mise de notes. Les épreuves externes genevoises sont élaborées par des commissions regroupant des enseignants, des experts disciplinaires et des spécialistes de l'évaluation. Elles évaluent le niveau d'atteinte des objectifs d'apprentissage définis dans le plan d'études romand (PER). Administrées dans toutes les classes à la fin des degrés 2 (7-8 ans), 4 (9-10 ans) et 6 (11-12 ans), elles contribuent à l'évaluation certificative des élèves en français et en mathématiques, ainsi qu'en allemand dans les deux derniers degrés 4 et 6. Dans les degrés 4 et 6, le résultat compte pour un tiers de la note du 3^{ème} trimestre. Les contenus, les dates, la durée et les conditions de passation, les modalités de correction et les barèmes sont uniformisés. Ces épreuves sont cependant administrées et corrigées par les titulaires de classe.

Dans cet article, nous nous penchons aussi sur les formes que peuvent prendre les contributions des enseignants à la recherche FNS, reposant sur une relation réciproque entre acteurs de terrain et chercheurs (St-Arnaud, 2003) et visant respectivement le développement professionnel et la production de savoirs scientifiques. Le développement professionnel s'exprime à travers le rapport à la pratique, questionné tout au long de la démarche collective et qui doit permettre aux enseignants d'atteindre un nouveau degré de compréhension de leur travail (Mukamurera, 2014). En considérant les enseignants comme à la fois objets et acteurs dans leur travail sur eux-mêmes et sur les objets de la recherche (Dubost & Lévy, 2002), nous avons cherché à mieux comprendre comment leurs propres préoccupations ont orienté le processus de recherche.

2. Les épreuves externes comme source de régulations.

Alors que de nombreux systèmes scolaires ont adopté la tradition anglo-saxonne de l'évaluation externe, le cadre théorique qui concerne ce type d'épreuves est encore limité (Mons, 2009). Les buts poursuivis par l'administration scolaire à ce sujet ont toujours été un sujet de controverses. En effet, les fonctions varient d'un pays à l'autre (Eurydice, 2009) et, dans un pays décentralisé comme la Suisse, d'un canton à l'autre. Selon certains auteurs (e.g. Bain, 1996 ; Blais, 2005 ; Resnick, 1994), on peut ramener à trois les principales fonctions de ces épreuves: la régulation des pratiques des enseignants, l'évaluation sommative des élèves et le monitoring du système d'enseignement. D'autres (Allal, 2002 ; Weiss, 2002) ont proposé de distinguer ce type d'épreuves selon deux logiques : celle de l'apprenant, ayant pour visée la gestion du cursus de l'élève, et celle de l'organisation, contribuant principalement au pilotage du système.

La définition des objectifs dans les plans d'études étant souvent trop générale pour guider les enseignants dans l'évaluation des apprentissages (Monseur & Demeuse, 2005 ; Scallon, 2004), la traduction en tâches susceptibles d'en révéler le niveau d'atteinte par les élèves (Roegiers, 2010) reste relativement délicate. Si les épreuves externes auxquelles nous nous référons dans cet article évaluent les connaissances et compétences des élèves en fin d'année

¹ Fonds National Suisse de la Recherche Scientifique, octroi n°100013-143453/1.

scolaire à trois reprises à l'école primaire, elles indiquent aussi aux enseignants les niveaux attendus à ces moments du cursus.

En observant les pratiques des pays européens en matière d'évaluation externe (Morris, 2011) il apparaît que les épreuves régionales ou nationales remplissent parfois plusieurs fonctions simultanément. En Suisse romande, celles-ci constituent une forme hybride d'évaluation scolaire, située entre l'évaluation des apprentissages des élèves assurée par les enseignants et l'évaluation du système scolaire (Ntamakiliro & Tessaro, 2010). Ces épreuves contribuent, d'une manière ou d'une autre, à une régulation des pratiques d'enseignement, l'exploitation des résultats pouvant avoir lieu potentiellement à plusieurs niveaux (Tessaro, 2008). Au sein de la classe, les régulations qui résultent des épreuves externes certificatives passées en fin d'année se réalisent essentiellement par une adaptation de l'enseignement pour la volée d'élèves suivante. Outre les effets sur les pratiques d'évaluation (Brenan, 2006), elles permettraient une meilleure identification des besoins d'apprentissage des élèves et une plus grande différenciation de l'enseignement, ainsi que le développement d'un sentiment accru d'efficacité chez l'élève (Barneveld, 2008). Elles peuvent aussi entraîner une modification de la transposition curriculaire ainsi que des types de tâches proposés et de leur articulation (Mottier Lopez & Tessaro, 2010 ; Shepard, 2006). Pour certains auteurs, les épreuves externes servent parfois de « levier didactique » et favorisent des changements de pratiques (Bodin, Delory, Lafontaine & Wolf, 2001 ; Yerly, 2012), notamment en contribuant à l'implémentation de nouvelles méthodes pédagogiques ou d'un nouveau plan d'études.

Les recherches menées en Suisse romande auprès des enseignants de l'école primaire (Dierendonck, 2008 ; Longchamp, 2010 ; Ntamakiliro & Tessaro, 2002; Soussi, Guilley, Guignard & Nidegger, 2009) indiquent que les épreuves externes sont perçues comme complémentaires à leurs propres évaluations certificatives et qu'elles proposent des repères susceptibles d'engager une régulation des pratiques. Les enseignants se disent soucieux de la correspondance entre les résultats aux épreuves externes et leurs propres appréciations. Certains planifient leur enseignement en veillant à ce que les compétences évaluées dans les épreuves externes soient suffisamment apprises en classe, tout en évitant le « rétrécissement curriculaire » (Harlen, 2007 ; Slomp, 2008), c'est-à-dire un enseignement sur les seuls contenus évalués.

Si une étude a montré que la majorité des enseignants considèrent que les épreuves externes ont de réelles retombées positives sur le travail quotidien en classe (Lafontaine, Soussi & Nidegger, 2009), d'autres auteurs suggèrent que les retombées sur les pratiques ne sont pas toujours significatives (e.g. Duru-Bellat, 2009), notamment à cause du manque de formation dans l'analyse et l'utilisation des résultats des épreuves (Hirsch, Sioberg & Germuth, 2009 ; Ungerleider, 2007 ; Wayman & Stringfield, 2006).

En complétant les résultats des épreuves internes à la classe – conçues, administrées et corrigées par les enseignants – l'évaluation externe assurerait une plus grande équité. Selon Allal (2009), si cette équité s'estime par des indicateurs statistiques à l'échelle d'un système, elle se réalise pour l'enseignant, dans sa classe, en effectuant l'évaluation la plus juste possible compte tenu des informations souvent multiples qu'il a à sa disposition.

3. L'évaluation des apprentissages en classe : un jugement professionnel

Si les épreuves externes s'inscrivent dans le paradigme de la mesure et bénéficient des nombreux mais parfois anciens travaux docimologiques, les pratiques internes à la classe sont, elles, caractérisées par le jugement professionnel de l'enseignant, considéré à la fois comme un acte cognitif et une pratique sociale située (Mottier Lopez & Allal, 2008). Plusieurs

études dans le contexte scolaire ont montré que le jugement évaluatif de l'enseignant est un processus complexe, que sa construction est progressive et dynamique (Tourmen, 2014). Il est considéré comme « professionnel » (Lafortune & Allal, 2008 ; Laveault, 2008 ; Mottier Lopez & Tessaro, à paraître) lorsque son exercice reflète des caractéristiques telles que la rigueur, la cohérence et la transparence. Multidimensionnel, le jugement professionnel intègre des sources d'informations diverses, de nature qualitative et quantitative, combinées entre elles, ce qui ne correspond pas à un jugement mécanique visant à mesurer un écart entre des objectifs et la performance évaluée (Cooksey et al., 2007; Wyatt-Smith et al., 2010).

Dans cet article, nous nous penchons sur la conception de contrôles écrits certificatifs et sur l'appréciation des productions d'élèves, qui constituent deux des trois niveaux de jugement professionnel en matière d'évaluation définis par Allal (2013) – le troisième étant le bilan périodique combinant les résultats de plusieurs évaluations certificatives. Pour les enseignants participant à la recherche FNS dont nous exploitons quelques données, ces deux niveaux concernent 1) la construction d'un outil d'évaluation certificative des apprentissages des élèves, comportant le cadre de référence, les items, des types de tâches différents, la cotation, le barème critérié ; 2) l'appréciation des productions des élèves de leur classe respective sous forme de *jugements évaluatifs* tels qu'ils se manifestent pendant la correction et des *décisions* qui en résultent sous forme d'attribution de points, de notes et de commentaires qualitatifs (Mottier Lopez, Tessaro, Dechamboux, Morales Villabona & Serry, 2013).

Dans le contexte de la classe, l'instrument d'évaluation certificative participe à une évaluation ponctuelle suite à une séquence d'enseignement-apprentissage. Les types de tâches proposés – spécifiques ou complexes (Allal, Béatrix Kohler, Rieben, Rouiller Barbey, Saada-Robert & Wegmuller, 2001) – et la nature des réponses plus ou moins élaborées que les élèves doivent alors formuler ont une influence sur le jugement évaluatif de l'enseignant et sa variabilité. En effet, plus les tâches d'évaluation laissent place à de multiples et parfois imprévisibles démarches de résolution, plus elles sont sources de possibles divergences d'interprétation, donnant lieu à différentes décisions de la part des enseignants.

Le jugement de l'enseignant tel qu'il se réalise dans l'activité évaluative est prescrit par des plans d'études, des règlements, des directives institutionnelles, des épreuves externes, des manuels d'enseignement, mais il est aussi influencé par la réalité de la classe (Mottier Lopez et al., 2013). Comme l'a montré Tourmen (2009), la production du jugement final, ici la notation, peut être considérée comme une multitude d'évaluations provisoires qui s'ajustent progressivement.

4. La modération sociale pour améliorer la qualité des pratiques évaluatives

La modération sociale désigne des pratiques collectives de confrontation de « jugements humains » (Linn, 1993) sur des travaux d'élèves, entre enseignants et formateurs-chercheurs. Elle vise à construire collectivement une représentation commune des attentes officielles traduites dans les référents externes et à déboucher sur des consensus à propos des pratiques d'évaluation des apprentissages des élèves internes à la classe (Wyatt-Smith et al., 2010).

Les buts visés par la modération sociale sont d'améliorer la fiabilité, la comparabilité et la validité des évaluations certificatives et pronostiques réalisées par les enseignants (Harlen, 2007 ; Matters, 2006), ainsi que d'assurer un meilleur alignement (Biggs, 1999) entre le curriculum prescrit, les activités didactiques menées en classe et les situations d'évaluation (Laveault, 2009).

En cherchant à développer les compétences professionnelles des enseignants à produire des jugements d'évaluation équitables et à contribuer ainsi à améliorer le système éducatif, la recherche à partir de laquelle sont tirées nos données répond à l'une des deux fonctions de la modération sociale définies par Maxwell (2002), *for improvement*, qui reste, selon cet auteur, encore trop souvent une conséquence du second type (*for accountability*) alors qu'elle devrait correspondre à une attente délibérée. Considérée comme centrale, la pratique de modération sociale, dite aussi de *consensus*, offre des opportunités aux enseignants d'intégrer à leurs jugements ceux des collègues et, ce faisant, de créer une représentation partagée des critères et des attentes (Maxwell, 2007).

5. Un dispositif souple malgré les contraintes

La recherche FNS dans laquelle s'inscrit notre étude a concerné deux cohortes d'enseignants de l'école primaire (total : 28), suivies chacune pendant deux années scolaires consécutives. Ces enseignants étaient caractérisés par une diversité d'expériences professionnelles, notamment par le type de formation, le nombre d'années d'expérience et le parcours professionnel. Ils provenaient de plusieurs établissements scolaires de niveaux socioéconomiques contrastés, avaient une expérience d'au moins trois ans dans les degrés concernés et s'engageaient à suivre leurs élèves pendant deux ans (Mottier Lopez, Tessaro, Dechamboux & Morales Villabona, 2012).

Les trois volets de cette recherche sont (1) des observations individuelles de pratiques de correction de copies en français (expression écrite) et en mathématiques, assorties à des entretiens semi-structurés, le recueil des outils d'évaluation élaborés par les enseignants ; (2) des réunions de modération sociale visant à confronter les différentes évaluations du volet 1 et à construire une culture d'évaluation commune ; (3) la réplication du volet 1. Comme indiqué, ce dispositif a été utilisé deux fois : avec des enseignants de degrés 3 et 4 (élèves de 8-10 ans), puis avec les degrés 5 et 6 (élèves de 10-12 ans). Toutes les rencontres ont été enregistrées.

Lors des volets 1 et 3, des entretiens semi-dirigés ont permis de recueillir des informations à propos, notamment, des séquences didactiques menées en amont, des objectifs d'apprentissage évalués, du choix des items et du barème établi. Chaque enseignant a ensuite corrigé quatre copies d'élèves considérés comme « susceptibles de rencontrer des difficultés », tout en verbalisant les interprétations et raisonnements émergeant en cours de correction, selon la méthodologie du « *Think-aloud* » (Charters, 2003).

Le volet 2 proposait un dispositif de modération sociale dans un but explicite de développement professionnel associé à la démarche de recherche scientifique. Les séances ont alterné des échanges en sous-groupes et des mises en commun. L'ensemble des interactions verbales a été enregistré et toutes les traces écrites ont été recueillies. Le but des échanges en petits groupes et en plénière était de dégager les aspects *partagés* et *admis* du jugement professionnel en évaluation (Morrissette, 2010).

Pour cet article, nous n'avons analysé qu'une partie du corpus de recherche : des retranscriptions du volet 1 des deux cohortes et des échanges de modération sociale en sous-groupes et en plénière (volet 2).

6. Des choix négociés et un développement professionnel contextualisé

Dans la recherche FNS, des rencontres préalables au recueil de données ont permis de définir avec les enseignants les domaines évalués au cours du premier volet ainsi que d'autres aspects du dispositif. Pour la première cohorte, il a, par exemple, été décidé conjointement de retenir l'expression écrite en français et des problèmes additifs en mathématiques.

Pendant le déroulement de la recherche, la nature et l'intensité de la collaboration entre les participants ont été différentes (Landry, Anadon & Savoie-Zajc, 1996). Lors des réunions du volet 2 de la première cohorte, des outils pour soutenir les dimensions de la modération sociale ont été proposés par les chercheurs, à partir de la littérature de recherche sur l'évaluation des apprentissages, dans une perspective d'analyse réflexive des pratiques. Ainsi, les « manières de faire » (Morrissette, 2010) ont été croisées avec les catégories propres à la construction et à la correction des contrôles écrits certificatifs que les enseignants devaient analyser. Les interactions sociales qui en ont résulté, entre enseignants puis entre enseignants et chercheur-formateur, ont représenté un vecteur privilégié d'apprentissage par la confrontation des points de vue et constitué des opportunités de changement de représentations (Paquay, Van Nieuwenhoven & Wouters, 2010). A plusieurs reprises, les outils ont été ajustés afin qu'ils soient en adéquation avec la réalité des pratiques des enseignants et qu'ils puissent continuer à soutenir les confrontations et le partage du savoir d'expérience (Desgagné & Larouche, 2010) d'une part, la construction de consensus d'autre part (Mottier Lopez, Tessaro, Dechamboux, Morales Villabona & Serry, 2014).

Le processus collaboratif pour atteindre ces deux buts a été révélateur des postures des uns et des autres. Du temps a été consacré à la description des pratiques et à leur mise en relation avec des observables (Beckers & Simons, 2010) que sont les instruments d'évaluation (leurs propres contrôles écrits, mais aussi d'autres exemples de pratiques). L'ambition était que ce partage du savoir d'expérience parvienne à un statut de savoir de recherche par la théorisation, tout en comportant une dimension émancipatoire pour les praticiens et servant de levier de développement professionnel (Desgagné, 2001).

Dans le dispositif décrit, il s'agissait de réfléchir à la façon dont les choix des praticiens peuvent orienter le processus et les objets de recherche, de façon dialectique et constructive. Cette réciprocité entre les deux catégories d'acteurs, caractéristique de la recherche collaborative (Bélaïr, 2009), permet d'installer conjointement une *zone proximale de recherche* au sein de laquelle les partenaires coconstruisent un sens partagé (Bourassa, Phillion & Chevalier, 2007). Les activités menées dans les différentes rencontres de modération sociale ont été l'occasion de développement professionnel et l'occasion de recherche et participent à une double vraisemblance, tant pour les acteurs du terrain que pour les chercheurs-formateurs (Desgagné, 1998).

L'atteinte de ces deux buts questionne le rôle classique du chercheur, qui est à la fois organisateur et participant dans la mise en place d'un espace réflexif, à travers le choix et l'ajustement des activités favorables aux discussions, ainsi que dans les régulations des interactions (Bednarz, 2013). La réflexion sur l'action, ses effets et les consensus auxquels devaient aboutir les enseignants ont demandé un accompagnement pour permettre la construction des savoirs. Cet accompagnement, caractérisé par des ajustements permanents (Perréard Vité & Tessaro, à paraître), a confronté les participants à l'image traditionnelle de l'expert – celui qui sait – amenant certains à adopter comme dans d'autres formations une attitude d'« exécutant » (Bednarz, Desgagné, Maheux & Savoie Zajc, 2012).

Dans le dispositif de modération sociale, la « montée en généralité » à partir des singularités des acteurs (Quéré, 1984) a pris en compte les contraintes institutionnelles locales qui, même si elles laissent une certaine autonomie et une marge d'interprétation (Touraine, 1984), obligent le chercheur-formateur à maintenir un esprit critique sans être prescriptif, tout en ne perdant pas de vue l'interprétation de l'objet de recherche (Larouche, 2005). Le double rôle de formateur et de chercheur a ainsi représenté un équilibre difficile à maintenir, veillant à la fois à faire émerger le savoir partagé par les acteurs en leur offrant un lieu pour échanger en toute liberté et se souciant de rendre transparente sa propre interprétation en accédant à la discussion (Dubet, 1994). Mais certainement que le rapprochement, voire la médiation entre les deux cultures (Bourassa, Bélaïr & Chevalier, 2007) sont à ce prix.

7. Des repères pour le jugement évaluatif

L'observation des pratiques d'évaluation des enseignants questionne, d'une façon ou d'une autre, l'adéquation entre celles-ci et les politiques scolaires (Laveault, Dionne, Lafontaine, Tessaro & Allal, 2014). Les premières analyses des entretiens et des groupes de modération sociale effectuées par l'équipe de recherche EReD¹ semblent indiquer à ce propos que bien qu'il soit mentionné dans les directives comme un repère essentiel pour l'évaluation, le nouveau plan d'études romand n'oriente la pratique que de façon indirecte (Mottier Lopez et al., 2012). Que les enseignants avouent une difficulté à utiliser ce document curriculaire pour l'évaluation des apprentissages des élèves n'est pas surprenant: la progression est présentée sous forme d'objectifs généraux qui sont à spécifier par les enseignants. Ces derniers soulèvent cependant l'importance d'un alignement entre ce document et les activités didactiques réalisées en classe. En définitive, ce sont, selon eux, ces activités qui constitueraient leur référent externe principal lorsqu'ils élaborent un contrôle écrit. Ils s'assurent ainsi d'évaluer ce qui a été vraiment enseigné, tout en respectant le prescrit.

Comme nous l'avons montré ailleurs (Serry & Tessaro, à paraître), les contrôles écrits certificatifs analysés montrent une homogénéité très forte concernant les en-têtes placés en page de garde. Ces derniers présentent notamment les objectifs d'apprentissage retenus pour l'évaluation, les items correspondants, la cotation, le barème et la note. On y découvre aussi les attentes minimales – les seuils de réussite précisant l'atteinte des objectifs – qui se construisent à partir de chaque exercice, objectif, composante ou de façon globale.

7.1 Une harmonisation progressive des pratiques

Intégrées aux guides de correction des enseignants, les tables de spécification des épreuves externes du canton de Genève comportent des informations importantes pour l'analyse des productions des élèves. Nos analyses montrent qu'une partie de ces informations sont présentes dans les instruments élaborés par les enseignants. Cette similitude semble indiquer que, sur ce point, le référent externe contribue à l'élaboration des évaluations certificatives par les enseignants.

Lors de la modération sociale, les groupes ont engagé des discussions à propos de ces tables de spécification pour en établir les liens avec leurs propres pratiques et une éventuelle adéquation. Dans l'extrait de verbatim ci-dessous, le trio d'enseignantes tente de cerner dans quelle mesure les visées et les contenus sont semblables, au-delà de la forme qui peut varier.

¹Evaluation, Régulation et Différenciation des apprentissages (<http://www.unige.ch/fapse/ered/index.html>)

- Mar* Enfin la différence c'est que ça (la table de spécification de l'épreuve externe) c'est pas donné aux parents ni à l'élève.
- Pat* Ouais, parce que nous elle (l'en-tête du contrôle écrit) est quand même à destination des parents.
- Kat* Mais est-ce que c'est une table de spécification ?
- Mar* Oui
- Kat* Oui ?
- Pat* Enfin, qu'on appelle autrement.
- Mar* Regarde (lit) prendre une mesure à l'aide d'un instrument ... être capable de lister ... c'est pareil.
- Pat* Parce que moi j'appelle ça un objectif.
- Kat* Moi aussi, objectif spécifique.
- ...
- Mar* Je crois que c'est pareil.
- ...
- Pat* Moi je mets pas de table de spécification. En voyant ça ... je me dis peut-être que je devrais faire une table de spécification.
- ...
- Kat* ... On est d'accord pour la table de spécification destinée uniquement à l'enseignant ou à l'élève ?
- Mar* Est-ce que nous on dit qu'elle DOIT être sur l'épreuve ? Il me semble que oui.

Cet extrait tiré du deuxième volet de la recherche FNS illustre une confrontation sociale visant à construire une représentation commune des attentes officielles traduites dans les épreuves externes et à déboucher sur un consensus à propos de leurs propres pratiques d'évaluation (Mottier Lopez et al., 2012). Il apparaît que le référent externe contribue ici à orienter et harmoniser des pratiques enseignantes différentes entre elles, tout en favorisant une discussion sur « ce qui est acceptable » pour l'une et pour l'autre, ainsi que pour l'institution. L'outil proposé par le chercheur-formateur provoque dans ce groupe une discussion à propos des contenus qui doivent être présents dans les contrôles écrits outre les items, mais aussi à propos des destinataires concernés. S'il s'agit de faire un bilan des apprentissages des élèves, l'instrument d'évaluation a en effet aussi une visée informative pour les parents.

Dans le second extrait, les interactions concernent les attentes minimales des enseignants face aux productions des élèves, représentées dans les instruments d'évaluation par le *seuil de réussite*. Celui-ci est fixé institutionnellement à la note 4 (sur 6). L'enjeu de cet échange est de mieux comprendre comment il peut être établi, toujours à la lumière de ce qui est indiqué dans l'épreuve externe.

Mar Alors toi t'as fait quoi ? T'es allée aux deux tiers (des points) ou t'as regardé par exercice ?

Pat Moi ... je prends le total des points et je divise par trois. Je fais fois deux et puis là je me suis dit que c'est une évaluation un petit peu difficile alors j'arrondis vers le bas.

Mar D'accord. Et toi tu regardes par exercice ?

Kat Moi je regarde par exercice effectivement... Après je regarde quand même où est-ce que ça se situe par rapport à ce fameux deux tiers. Je regarde si je suis dedans si je suis beaucoup plus haut ou beaucoup plus bas.

Mar Moi je fais par exercice puis je regarde. On tombe dessus (les deux-tiers des points) plus ou moins.

...

Kat Donc on est d'accord de ne pas mettre le barème (automatiquement) aux deux tiers.

...

Mar Je me pose une question quand même : quand on fait par exercice comme on fait nous (dans notre école), est-ce qu'on n'a pas déjà le deux tiers dans la tête ?

Pat Ouais je pense.

Kat Disons que si t'as un exercice où t'as six points et ils réussissent seulement la moitié. ... Est-ce qu'en réussissant seulement la moitié de l'exercice je peux considérer qu'il a compris l'objectif ? Tu te dis non, je trouve que c'est (trop) facile et qu'il faudrait un tout petit peu plus.

Pat Donc tu mets deux tiers.

Les interactions entre les trois enseignants rendent compte de la diversité de leurs pratiques, tout en convergeant sur le fait que les attentes concernant l'atteinte minimale des objectifs sont à définir à chaque nouvelle situation d'évaluation – selon la difficulté, le travail effectué, le moment – et qu'elles ne sont donc pas dépendantes d'une logique arithmétique (2/3 des points). A ce propos, si les concepteurs des épreuves externes adoptent des procédures différentes selon la discipline, ils n'établissent en effet jamais les seuils de réussite de façon proportionnelle. Ceux-ci sont définis systématiquement selon les caractéristiques de l'épreuve. Cependant, aucune indication n'est donnée concernant le processus d'élaboration, laissant les enseignants parfois démunis dans leurs pratiques.

7.2 La situation complexe en question

L'analyse des contrôles écrits montre que les épreuves externes soutiennent l'harmonisation des pratiques sur d'autres plans. Elles valorisent par exemple, depuis une quinzaine d'années, la présence d'items complexes en mathématiques, visant l'élaboration par l'élève d'une compétence dans une situation relativement ouverte (Allal et al., 2001). Selon certains enseignants, cette orientation a constitué un vecteur de changement des pratiques d'enseignement et servi ainsi de « levier didactique » (Bodin et al., 2001). Cependant, les pratiques d'évaluation en classe semblent résister. En moyenne, nous n'avons trouvé qu'environ 40% de points attribués aux tâches complexes, alors que les épreuves externes des sept dernières années en comptabilisent entre 61% et 70 % selon le degré et l'année scolaire.

Lors de réunions de modération sociale, des enseignants ont expliqué la présence d'une majorité de tâches d'application par le fait que celles-ci sont plus faciles à évaluer d'une part,

ne discriminent pas comme le font les situations-problèmes et contribuent donc à la réussite des élèves d'autre part.

En expression écrite, les productions analysées contiennent toujours un guide à l'usage de l'élève, qui indique la forme et les contenus caractérisant le genre textuel évalué. Ce guide fait aussi office de grille d'évaluation pour l'enseignant, voire d'auto-évaluation de l'élève. Outil polyvalent, il est issu des différents modules abordés lors de la séquence didactique, donc dans une perspective d'enseignement/apprentissage, et sert de support à l'évaluation certificative.

Le recours à un tel outil questionne la notion de « situation complexe » (Allal et al., 2001 ; Beckers, 2002). A titre illustratif, voici ce que dit Karen, l'une des participantes :

« On a fait pour l'évaluation une grille qu'était vraiment très complète pour aider les élèves pour aller pas à pas. Ça nous aide nous aussi pour la correction... Je me dis maintenant que peut-être la grille était un petit peu trop longue mais en même temps elle est très très claire et très précise... Si vraiment ils lisent ils ont quand même facilement les points, mais après quand on regarde quand même globalement ça mérite pas toujours (la note attribuée) ... ».

Avec un tel outil, la situation complexe et authentique de production de l'écrit est découpée en autant de ressources nécessaires pour la réaliser (Beckers, 2001 ; Roegiers, 2010). Les élèves n'ont ainsi pas à mobiliser celles-ci, mais essentiellement à les traiter, de façon souvent chronologique d'ailleurs. L'épreuve externe a participé à la création d'une culture commune en véhiculant auprès des enseignants la nécessité d'assurer un étayage lors d'une production textuelle. Il s'agit dès lors pour les élèves d'accomplir des procédures, mettant en jeu une compétence élémentaire (Rey, Carrette, Defrance & Kahn, 2006). L'usage d'un tel outil en situation d'évaluation certificative suscite débat : tant auprès des scientifiques que chez certains enseignants, qui estiment que l'évaluation certificative de cette composante de la langue n'est par conséquent pas significative.

8. Soutenir la régulation des pratiques évaluatives : une préoccupation partagée

Les pratiques d'évaluation des enseignants semblant généralement proches de celles promues par les épreuves externes du canton, ces dernières contribueraient ainsi à l'alignement et à une cohésion interne (Laveault et al., 2014) de façon descendante. La modération sociale utilisée dans la recherche FNS pourrait cependant participer aussi à une perspective ascendante (Mottier Lopez & Tessaro, 2010), dès lors que l'harmonisation des pratiques issue des échanges concourt à la formulation de politiques d'évaluation (encore) plus adéquates. Les collaborations d'établissement que nous avons observées à propos de la construction et de la correction des contrôles écrits favorisent de façon nouvelle cette harmonisation ainsi que la création d'une culture partagée de l'évaluation (Morales Villabona & Mottier Lopez, à paraître).

Ce constat s'établit malgré plusieurs changements conséquents, effectués ces dernières années dans les contenus et la forme des épreuves externes du canton de Genève. Seule exception notoire : la proportion des tâches complexes dans les contrôles écrits en mathématiques reste inférieure dans les pratiques d'évaluation en classe. Ce type de tâches reste difficile à évaluer, ce qui peut expliquer en partie son nombre réduit. Le maintien d'une majorité de situations d'application, voire de mémorisation, et donc de tâches spécifiques, renvoie dès lors à un maintien de pratiques d'enseignement plus « traditionnelles » qui privilégient l'acquisition de connaissances déclaratives ou procédurales chez les apprenants,

mais qui n'intègrent pas (ou insuffisamment) ces composantes cognitives dans des situations plus larges (Allal, 1999). La mobilisation adéquate de ces ressources apparaît pourtant comme un aspect essentiel dans le développement d'une compétence (Jonnaert, Ettayebi & Defise, 2009).

Si les épreuves externes peuvent soutenir l'amélioration des pratiques d'évaluation des enseignants genevois, il reste (à continuer) à réfléchir à un usage adéquat dans les dispositifs de formation continue, de manière à les considérer comme des repères significatifs. La modération sociale pratiquée au sein d'un établissement apparaît dès lors comme une modalité et un niveau d'action intéressants et constitue ainsi une perspective prometteuse. La confrontation des épreuves externes avec celles élaborées par les enseignants s'y inscrit dans une conception contextualisée du développement professionnel, qui n'est pas détaché de la pratique dans lequel il prend place (Bednarz & Barry, 2010). Les consensus qui ont émergé entre les enseignants, dont certains ont été directement ou indirectement induits par les épreuves externes, devraient permettre un réinvestissement dans les pratiques ainsi que l'intériorisation de nouveaux savoirs professionnels.

Par ailleurs, la présentation des résultats des premières analyses effectuées par l'équipe EReD pourrait également avoir bénéficié au développement professionnel des enseignants. Avec la première cohorte, les constats généraux effectués à la suite du premier volet de recherche ont été présentés et discutés lors de la modération sociale mise en place avec tous les participants. Avec la seconde cohorte, les résultats partiels ont été aussi présentés en plénière, suscitant de la part des participants des réactions variées sur les pratiques de collègues et questionnant parallèlement leurs propres usages. Cette présentation a contribué à la démarche collective de recherche de consensus à propos du jugement évaluatif et à la verbalisation de pratiques d'arrangements et d'ajustements (Mottier Lopez et al., 2013).

Le dispositif de la recherche FNS, en partie co-élaboré avec les participants, avait également pour intention de promouvoir le développement professionnel des enseignants des établissements scolaires dans lesquels des formations continues ont été proposées. En effet, les modérations sociales soutenant l'élaboration de consensus sont d'autant plus significatives dans ce contexte qu'elles mettent en jeu les conceptions de chacun dans la réalisation collégiale de démarches authentiques – de la construction d'un contrôle écrit certificatif à la notation. L'expérience réalisée illustre un aspect moins étudié de la recherche collaborative. Ici, l'*interfécondation* (Desgagné, 1997) se réalise dans une temporalité et des modalités différentes de celles présentées habituellement: elle concerne deux cohortes de participants à la recherche ainsi que des enseignants impliqués dans des formations en établissement. Savoirs scientifiques et de la pratique devraient ainsi bénéficier des apports mutuels dans une progression que l'on pourrait qualifier de « spiralaire », à l'image de certaines formes de curriculum (Allal et al., 2001).

9. Références

- Allal, L. (1999). *Acquisition et évaluation des compétences en situation scolaire*. In J. Dolz & E. Ollagnier. (Ed.), *L'énigme de la compétence en éducation* (pp. 77-94). Bruxelles : De Boeck Université.
- Allal, L. (2002). Intervention au symposium « Epreuves externes : évaluation de la qualité des apprentissages ou des curricula ? » Congrès de l'Admée-Europe. Université de Lausanne.
- Allal, L. (2009). Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur les systèmes scolaires. In L. Mottier Lopez & M. Crahay (Ed.), *Évaluations en tension : entre la régulation des apprentissages et le pilotage des systèmes* (pp. 29-45). Bruxelles : De Boeck.
- Allal, L. (2013). Teachers' professional judgement in assessment: a cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, (special issue: Moderation practice and teacher judgement), 20, 20-34.
- Allal, L., Béatrix Kohler, D., Rieben, L., Rouiller Barbey, Y., Saada-Robert, M. & Wegmuller, E. (2001). *Apprendre l'orthographe en produisant des textes*. Fribourg : Editions Universitaires.
- Bain, D. (1996). Introduction au thème du colloque de l'ADMEE-Europe 1994. In D. Bain (Ed.) *Fonctionnement de l'évaluation dans la formation, évaluation du fonctionnement de la formation* (pp.3-26). Genève : Centre de recherches psychopédagogiques.
- Barneveld, C. (2008). Using Data to Improve Student Achievement, Research Monograph No. 15, the Literacy and Numeracy Secretariat, Ontario www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/Using_Data.pdf.
- Beckers, J. (2001). Aider les élèves à développer des compétences à l'école : Révolution ou continuité ? *Puzzle*, 10, 2-10.
- Beckers, J. (2002). Pistes de réflexion pour développer, à l'école, les compétences de tous les élèves. *L'École valdôtaine*, 55, 4-11.
- Beckers, J. & Simons, G. (2010) Analyse rétrospective de programmes de recherches collaboratives en langues modernes à l'Université de Liège. *Recherches en Education*, (1), 31-46.
- Bednarz, N. (2013). Regarder ensemble autrement : ancrage et développement des recherches collaboratives en éducation au Québec. In N. Bednarz (Ed.), *Recherche collaborative et pratique enseignante. Regarder ensemble autrement* (pp. 13-29). Paris : L'Harmattan.
- Bednarz, N. & Barry, S. (2010). Recherches collaboratives en enseignement des mathématiques comme soutien au développement professionnel des enseignants. In C. Couture et L. Dionne (Ed.), *La formation et le développement professionnel des enseignants en sciences, technologie et mathématiques* (pp. 225-253). Ottawa : Les Presses de l'Université d'Ottawa.
- Bednarz, N., Desgagné, S., Maheux, J.-F. & Savoie-Zajc, L. (2012). La mise au jour d'un contrat réflexif comme régulateur de démarches de recherche participative : le cas d'une recherche-action et d'une recherche collaborative. *Recherches en éducation*, 14, 129-152.
- Bélaïr, L. (2009). *Réflexions autour de la recherche collaborative en développement professionnel*. Discussion du symposium "Recherches collaboratives: Enjeux et modalités" organisé par L. Mottier Lopez. Actes du 21ème colloque de l'ADMEE-Europe. Université catholique de Louvain. Louvain-La-Neuve.
- Biggs, J. (1999). *Teaching for Quality Learning at University : What the Student Does*. The Society for Research into Higher Education and Open University Press. Buckingham.
- Blais, J.-G. (2005). L'obligation de résultats à la lumière de l'interaction entre le quantitatif et le social. In C. Lessard & P. Meirieu (Ed.), *L'obligation de résultats en éducation. Evolution, perspectives et enjeux internationaux* (pp.123-144). Bruxelles : De Boeck.
- Bodin, A., Delory, C., Lafontaine, D. & Wolf, J.-L. (2001). Quelle évaluation externe des élèves pour quels socles de compétences ? *Pédagogies*, 14, 56-65.
- Bourassa, M., Bélaïr, L. & Chevalier, J. (2007). Les outils de la recherche participative. *Education et Francophonie*, XXXV (2), 1-11.

- Bourassa, M., Phillion, R. & Chevalier, J. (2007). L'analyse de construits, une coconstruction de groupe. *Education et Francophonie*, XXXV (2), 78-116.
- Brennan, R.L. (2006). Perspectives on the evolution and future of educational measurement. In R.L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 1-16). Praeger Publishers. Westport.
- Charters, E. (2003). The Use of Think-aloud Methods in Qualitative Research. An Introduction to Think-aloud Methods. *Brock Education*, 12(2), 68-82.
- Cooksey, R., Freebody, P. & Wyatt-Smith, C.M. (2007). Assessment as judgement-in-context: Analysing how teachers evaluate students writing. *Educational Research and Evaluation*, 13(5), 401-434.
- 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.
- Desgagné, S. (1998). La position du chercheur en recherche collaborative : illustration d'une démarche de médiation entre culture universitaire et culture scolaire. *Recherches qualitatives*, 18, 77-105.
- Desgagné, S. (2001). La recherche collaborative : nouvelle dynamique de recherche en éducation. In M. Anadon (Ed.), *Des nouvelles dynamiques de recherche en éducation* (pp. 51-76). Québec : Presses de l'université de Laval.
- Desgagné, S. et H. Larouche (2010). Quand la collaboration de recherche sert la légitimation d'un savoir d'expérience. *Recherches en Education*, hors série, 7-18.
- Dierendonck, C. (2008). *Comment les évaluations externes des acquis des élèves sont-elles perçues par les enseignants du primaire dans les cantons de Neuchâtel, Vaud et Fribourg ?* Neuchâtel : IRDP.
- De Stercke, J. (2010). *Le mentorat en enseignement*. Mémoire de Master en sciences de l'éducation. Université de Mons.
- Dubet, F. (1994). *Sociologie de l'expérience*. Seuil.
- Dubost, J. & Lévy, A. (2002). Recherche-action et intervention. In J. Barus-Michel, E. Enriquez et A. Lévy, *Vocabulaire de Psychologie, Références et positions* (pp. 391-416). Paris : Eres.
- Duru-Bellat, M. (2009). Evaluation externe du système éducatif et ressenti des acteurs. In L. Mottier Lopez & M. Crahay (Ed.), *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (pp. 47- 59). Bruxelles : De Boeck.
- Eurydice (2009). *Les évaluations standardisées des élèves en Europe : objectifs, organisation et utilisation des résultats*. Bruxelles: Agence exécutive «Éducation, audiovisuel et culture» (EACEA P9 Eurydice).
- Figari, G. (2006). L'activité évaluative entre cognition et réponse sociale : nouveaux défis pour les évaluateurs. *Mesure et Evaluation en Education*, 29(1), 5-18.
- Guimard, P., Cosnefroy, O. & Florin, A. (2007). Evaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances et des parcours à l'école élémentaire et au collège. *L'Orientation scolaire et professionnelle*, 36(2), 179-202.
- Harlen, W. (2007). *Assessment of learning*. London : SAGE.
- Hirsch, E., A. Sioberg and A. Germuth (2009), TELL Colorado: Creating Supportive School Environments to Enhance Teacher Effectiveness, the New Teacher Center.
- Jonnaert, P., Ettayebi, M. & Defise, R. (2009). *Curriculum et compétences : un cadre opérationnel*. Bruxelles : De Boeck.
- Lafontaine, L., Soussi, A. & Nidegger, C. (2009). Evaluations internationales et/ou épreuves nationales: tensions et changement de pratiques. In L. Mottier Lopez et M. Crahay (Ed.), *Evaluations en tension* (pp. 61-80). Bruxelles : De Boeck.
- Lafortune, L. & Allal, L. (2008). *Jugement professionnel en évaluation, Pratiques enseignantes au Québec et à Genève*. Québec : Presses de l'Université du Québec.
- Landry, C., Anadón, M. & Savoie-Zajc, L. (1996). Du discours politique à celui des acteurs : le partenariat en éducation, une notion en construction. *Apprentissage et Socialisation*, 17(3), 9-28.
- Larouche, H. (2005). Le double rôle de formatrice et de chercheure : un point de rencontre pour concilier les besoins de la pratique et de la recherche. *Revue des Sciences de l'Education*, 31(2), 335-354.

- Laveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Revue Suisse des Sciences de l'Éducation*, 30 (3), 483-500.
- Laveault, D. (2009). L'amélioration de l'efficacité du système éducatif : sur quels indicateurs s'appuyer ? In X. Dumay & V. Dupriez (Ed.), *L'efficacité dans l'enseignement : promesses et zones d'ombre* (pp. 177-194). Bruxelles: De Boeck.
- Laveault, D., Dionne, E., Lafontaine, D., Tessaro, W. & Allal, L. (2014). L'évaluation des compétences en salle de classe : régulations entre politiques et pratiques. In C. Dierendonck, E. Loarer & B. Rey (Ed.), *L'évaluation des compétences en milieu scolaire et en milieu professionnel*. Bruxelles : De Boeck.
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83-102.
- Longchamp, A. L. (2010). *Epreuves cantonales de référence de mathématiques en 6e année. Point de vue d'enseignants sur l'utilisation du dispositif*. Lausanne : URSP.
- Matters, G. (2006). Good data, bad news, good policy making. *Queensland Teachers Union Professional Magazine*. October. Brisbane : QTU.
- Maxwell, G. (2002). *Moderation of Teacher Judgments in Student Assessment*. Brisbane : Queensland School Curriculum Council.
- Maxwell, G. (2007). *Implications for moderation of proposed changes to senior secondary school syllabuses*. Brisbane: Queensland Studies Authority.
- Mons, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue française de pédagogie. Recherches en Éducation*, 169, 99-140.
- Monseur, C. & Demeuse, M. (2005). Les évaluations externes permettent-elles une régulation efficace ? In M. Demeuse, A. Baye, M.H. Straeten, J. Nicaise & A. Matoul (Ed.), *Vers une école juste et efficace* (pp. 489-517). Bruxelles : De Boeck.
- Morales Villabona, F. & Mottier Lopez, L. (à paraître). Quelle évaluation collaborative dans la modération sociale entre enseignants? In L. Mottier Lopez & W. Tessaro (Ed.), *Le jugement professionnel, au coeur de l'évaluation et de la régulation des apprentissages*. Berne : Peter Lang.
- Morris, A. (2011) *Student standardised testing : current practices in OECD countries and a literature review*. OECD Education Working Paper No. 65.
- Morissette, J. (2010). *Manières de faire l'évaluation formative des apprentissages : analyse interactionniste du savoir-faire d'enseignantes du primaire*. Sarrebruck: Les Éditions Universitaires Européennes.
- Mottier Lopez, L. (2008). *Apprentissage situé : la microculture de classe en mathématiques*. Berne: Peter Lang.
- Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue Suisse des Sciences de l'Éducation*, 30(3), 465-482.
- Mottier Lopez, L. & Tessaro, W. (2010). Evaluation des compétences à l'école primaire genevoise : entre prescriptions et pratiques. *Mesure et Evaluation en Éducation*, 33(3), 29-53.
- Mottier Lopez, L. & Tessaro, W. (à paraître). *Le jugement professionnel, au coeur de l'évaluation et de la régulation des apprentissages*. Berne : Peter Lang.
- Mottier Lopez, L., Tessaro, W. & Fillietz, L. (2012). *Étude des jugements d'évaluation des apprentissages des élèves et des pratiques de modération sociale*. Requête acceptée par le FNS de la recherche scientifique : 100013_143453/1.
- Mottier Lopez, L., Tessaro, W., Dechamboux, L. & Morales Villabona, F. (2012). La modération sociale : un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves. *Questions Vives*, 18, 159-175.
- Mottier Lopez, L., Tessaro, W., Dechamboux, L., Morales Villabona, F. & Serry, S. (2013). *Jugements évaluatifs-en-acte : analyse des discours et des pratiques d'enseignants de l'école primaire genevoise*. Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF). Université de Montpellier.
- Mottier Lopez, L., Tessaro, W., Dechamboux, L., Morales Villabona, F. & Serry, S. (2014). *Modération sociale et outils de la recherche dans une approche participative : quelle médiation des processus de construction d'une culture commune de l'évaluation des apprentissages des élèves ?* Actes du colloque de l'ADMEE-Europe. Marrakech.

- Mukamurera, J. (2014). Le développement professionnel et la persévérance en enseignement. Éclairage théorique et état des lieux. In L. Portelance, S. Martineau, S. & J. Mukamurera. (Ed.), *Développement et persévérance professionnels en enseignement. Oui mais comment?* Québec : Presses de l'Université du Québec.
- Ntamakiliro, L. & Tessaro, W. (2002). *La perception de l'évaluation externe des élèves par les enseignants primaires genevois*. Actes du congrès de l'Admée-Europe. Université de Lausanne.
- Ntamakiliro, L. & Tessaro, W. (2010). Vers plus d'épreuves externes: une standardisation de l'évaluation scolaire? In P. Gilliéron Giroud & L. Ntamakiliro (Ed.), *Réformer l'évaluation scolaire: mission impossible ?* Berne: Peter Lang.
- Paquay, L., Van Nieuwenhoven, C. & Wouters, P. (2010). *L'évaluation, levier de développement professionnel*. Bruxelles : De Boeck.
- Perréard Vité, A. & Tessaro, W. (à paraître). Entre formation et accompagnement en formation continue : analyse d'une expérience au secondaire I. In E. Charlier, R. Wittorski, A. Perréard Vité & S. Boucenna (Ed.), *L'accompagnement et l'analyse des pratiques professionnelles comme vecteur de professionnalisation*. Toulouse : Octares.
- Quéré, L. (1984). *La sociologie à l'épreuve de l'herméneutique*. Paris : L'Harmattan.
- Resnick, L. (1994). Performance puzzles. *American Journal of Education*, 102, 511-526.
- Rey, B., Carette, V., Defrance, A. & Kahn, S. (2006). *Les compétences à l'école. Apprentissage et évaluation*. Bruxelles: De Boeck.
- Roegiers, X. (2010). *L'école et l'évaluation*. Bruxelles : De Boeck.
- Rosenkvist, M.A. (2010) *Using student test results for accountability and improvement: A literature review*. OECD Education Working Paper No. 54.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles : De Boeck.
- Serry, S. & Tessaro, W. (à paraître). Construction d'instruments d'évaluation certificative à l'école primaire: discours et pratiques d'enseignants genevois. In L. Mottier Lopez et W. Tessaro (Ed.), *Le jugement professionnel, au coeur de l'évaluation et de la régulation des apprentissages*. Berne : Peter Lang.
- Shepard, L.A. (2006). Classroom assessment. In R.L. Brennan (Ed.), *Educational measurement* (pp. 623–646). Westport : Praeger Publishers.
- Slomp, D. (2008). Harming not Helping: The Impact of a Canadian Standardized Writing Assessment on Curriculum and Pedagogy. *Assessing Writing*, 13(3), 180-200.
- Soussi, A., Guilley, E., Guignard & Nidegger, C. (2009). *Evaluation des acquis des élèves à l'école obligatoire. L'évaluation cantonale et commune : état de situation du dispositif existant, points forts et points faibles, perspectives d'avenir*. Genève : SRED.
- St-Arnaud, Y. (2003). *L'interaction professionnelle : efficacité et coopération*. Montréal : Les Presses de l'Université de Montréal.
- Tessaro, W. (2008). *Epreuves externes élaborées par les systèmes de formation: quelle utilité pour les enseignants ?* Introduction à la table ronde. Actes du 20ème colloque international de l'ADMEE-Europe, Genève.
- Touraine, A. (1984). *Le retour de l'acteur : essai de sociologie*. Paris : Fayard.
- Tourmen, C. (2009). L'activité évaluative et la construction progressive du jugement. *Les Dossiers des Sciences de l'Education*, 22, 101-119.
- Tourmen, C. (2014). Contributions des sciences de l'éducation à la compréhension de la pratique évaluative. *Politiques et Management Public*, 31(1), 69-85.
- Ungerleider, C. (2007). *Evaluation of the Ontario Ministry of Education's Student Success Strategy, Phase 1 Report*. Ottawa: Canadian Council on Learning.
- Wayman, J.C. & Stringfield, S. (2006). Technology-supported Involvement of Entire Faculties in Examination of Student Data for Instructional Improvement. *American Journal of Education*, 112(4), 549-571.
- Wegmuller, E. & Allal, L. (1997). LEX-EVAL. Lexique d'évaluation dans le cadre de la scolarité obligatoire. *Résonances*, 3, 14-15.

- Weiss, J. (2002). *L'évaluation externe des curricula, sa spécificité fonctionnelle dans un concept général et cohérent d'évaluation*. Communication au symposium « Epreuves externes : évaluation de la qualité des apprentissages ou des curricula ? » Congrès de l'Admée-Europe. Université de Lausanne.
- Wyatt-Smith, C. & Castleton, G. (2005). Examining how teachers judge student writing: An Australian case study. *Journal of Curriculum Studies*, 37(2), 131-54.
- Wyatt-Smith, C., Klenowski, V. & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: a study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.
- Yerly, G. (2012). L'évaluation externe standardisée : un levier didactique vers l'approche par compétences ? Actes du 24e colloque de l'ADMEE-Europe.