

# À quelles conditions développer des pratiques de notation au service des décisions des enseignants et des apprentissages des élèves ?<sup>1</sup>



**RAPHAËL PASQUINI**

HAUTE ÉCOLE PÉDAGOGIQUE DU CANTON DE VAUD (SUISSE)

## Introduction

L'étude « *Teaching and Learning International Survey* » (TALIS) de 2018 a montré qu'une immense majorité des élèves qui fréquentent les écoles des pays dits industrialisés voient leurs apprentissages essentiellement évalués au moyen d'épreuves sommatives<sup>2</sup> notées, élaborées par l'enseignant, passées en classe ou données en devoirs à la maison, sous une forme traditionnelle « papier-crayon » (Quittre *et al.*, 2019). En fonction du système et des degrés scolaires, le moment de l'introduction de la note varie et ses modes de structuration diffèrent : les échelles utilisées, la fixation des seuils de réussite ou le nombre de notes qualifiant l'échec peuvent changer. Toutefois, malgré ces spécificités contextuelles, la note est généralisée dans tous les systèmes scolaires (*Organisation for Economic Cooperation and Development* [OECD], 2012). Il devient dès lors raisonnable d'admettre, qu'à certaines conditions, elle représente l'information la plus concrète que les élèves peuvent obtenir sur les bilans de leurs apprentissages (Guskey & Link, 2019).

La question séculaire visant à se positionner « pour » ou « contre » la note est donc pour nous surannée et peu en lien avec la réalité des systèmes éducatifs. À ce titre, rappelons l'inefficacité de toutes les réformes mises en œuvre depuis vingt ans en francophonie à large échelle pour remplacer la note par d'autres codes<sup>3</sup>. Des recherches ont en effet mis en évidence qu'avec des appréciations lettrées (par exemple, « atteint »), avec des codes de couleurs ou d'autres symboles, les enseignants fonctionnent comme avec des pourcentages ou des notes chiffrées (Kohn, 2011). L'universalité de la note tiendrait notamment au fait que le corps enseignant entretient avec cette dernière une relation ambiguë : d'un côté, imposée par l'institution-École, elle est le principal vecteur par lequel il lui est demandé de « rendre compte » de son travail, de documenter l'avancée dans les programmes et de pronostiquer la suite du parcours de chaque élève. De l'autre, la note, au quotidien, prend une place

<sup>1</sup> Pour citer ce document, merci d'utiliser la référence suivante : Pasquini, R. (2023) À quelles conditions développer des pratiques de notation au service des décisions des enseignants et des apprentissages des élèves ? Dans Cnesco, *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (pp. 136-147). Cnesco-Cnam.

<sup>2</sup> L'évaluation sommative intervient à la fin de l'étude d'un sujet ou d'une séquence d'enseignement en dressant un bilan des apprentissages réalisés, sa fonction première étant de les certifier ou de les attester. Elle fait l'objet de l'attribution d'une note le plus souvent chiffrée. Elle est parfois synonyme d'évaluation certificative.

<sup>3</sup> Nous pensons par exemple à la réforme ministérielle au Québec en 2011, à la réforme École Vaudoise en Mutation dans le canton de Vaud à la fin des années 1990 ou à la décision du canton de Zürich de réintroduire les notes à la rentrée 2022.

conséquence dans le temps scolaire au détriment parfois de l'apprentissage : nombre d'enseignants disent ne faire que noter et s'en plaignent (Brookhart, 2013 ; Dubus, 2006 ; Kitabgi, 2009). Cette « course d'obstacles » que la note peut parfois provoquer a des effets sur le stress des élèves, sur leur motivation et sur leur rapport au savoir, et elle impacte également les parents qui n'ont comme balises de la réussite de leur enfant que ses notes et ses moyennes.

Tout en considérant les effets néfastes que la note peut provoquer et la logique de sélection qu'elle induit, et prenant acte que les pratiques de notation relèvent souvent d'habitudes peu questionnées (Brookhart *et al.*, 2016 ; Moss, 2013), nos travaux étudient les conditions nécessaires pour que la note se réfère aux apprentissages enseignés, réalisés et évalués, afin qu'elle puisse informer l'enseignant et l'élève de ce qui a été réussi et de ce qui doit encore être travaillé. L'enjeu est de pouvoir offrir des possibilités de rétroactions aux élèves afin de soutenir leurs apprentissages à l'aide de notes à haute valeur informative. En cela, nous parlons de note constructive (Pasquini, 2021).

## I. Notre démarche de formation pour favoriser le développement de pratiques de notation au service des apprentissages des élèves

---

Nos travaux et nos actions de formation s'appuient sur un nombre conséquent d'études réalisées dans la recherche anglophone, qui cherche depuis plusieurs décennies à mieux comprendre les conditions nécessaires au développement de pratiques de notation constructives (Brookhart, 2022 ; Guskey & Brookhart, 2019). Nous en déclinons ici les principaux apports.

### A. L'alignement curriculaire et son acception élargie

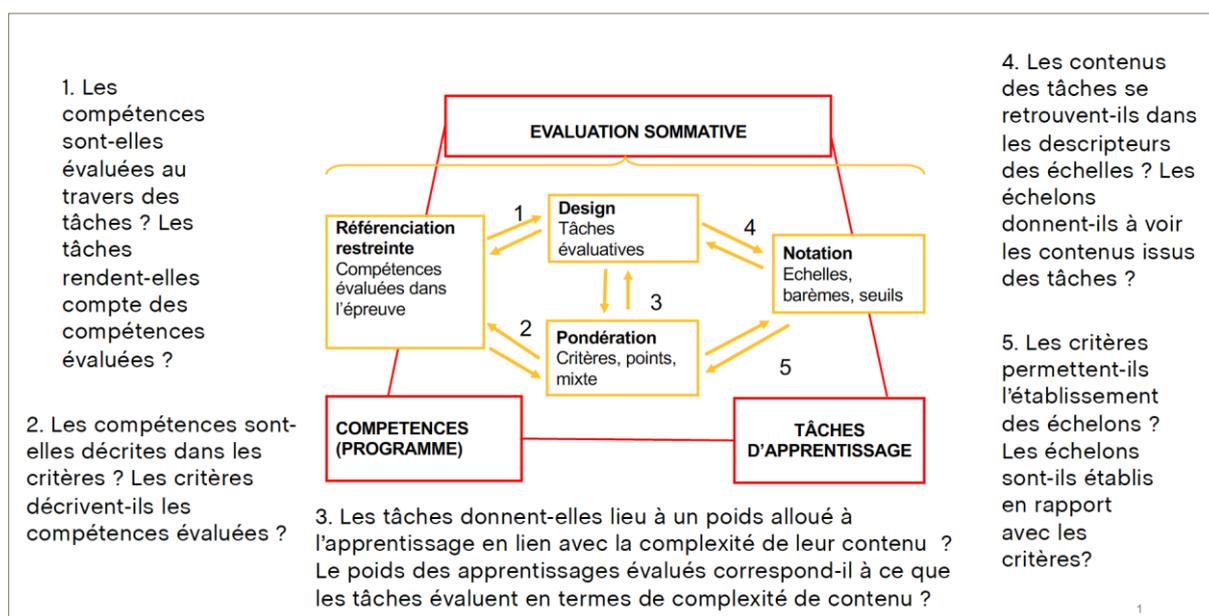
Anderson (2002) définit **l'alignement curriculaire comme étant « le niveau de cohérence entre les compétences du programme et l'évaluation, entre les compétences du programme et les activités d'apprentissage, et enfin entre l'évaluation et les activités d'apprentissage »** (p. 257, traduction libre). Les modalités d'enseignement ainsi que les opportunités d'apprentissage offertes aux élèves au travers de tâches d'apprentissage font également partie du modèle. Lorsque c'est le cas, nous parlons **d'alignement constructif** (Biggs, 2003). La notion d'alignement relève d'une conceptualisation de la cohérence à l'aune d'une taxonomie d'habiletés cognitives appariées à des contenus (Anderson & Krathwohl, 2001)<sup>4</sup>. La dimension curriculaire renvoie au fait que cette cohérence est pensée depuis un programme. L'élaboration d'épreuves évaluatives sommatives notées est ici conçue comme un processus complexe, dynamique, contextuel, centré sur les apprentissages (par ex., Biggs, 1999, 2003 ; Gauthier *et al.*, 2005 ; Martone & Sireci, 2009).

Face à certaines incomplétudes du modèle de départ qui ne conceptualise pas suffisamment les composantes spécifiques à l'évaluation sommative que sont la pondération et la notation, nous avons élaboré le modèle de **l'alignement curriculaire élargi** en nous basant sur la littérature et en observant les pratiques effectives (Figure 1- Pasquini, 2018, 2019a, 2020, 2021). C'est à l'aide de ce modèle que nous travaillons avec les enseignants depuis quatre ans, en formation initiale et continue, au primaire, au secondaire et au post-obligatoire, dans toutes les disciplines.

---

<sup>4</sup> Cette taxonomie permet de formuler – ou d'analyser – des objectifs d'apprentissage (ou compétences) en croisant six niveaux d'habiletés cognitives croissant en complexité (se souvenir, comprendre, appliquer, analyser, évaluer et créer) avec quatre contenus de nature différente (les faits, les concepts, les procédures et les démarches métacognitives).

**Figure 1 : L'alignement curriculaire élargi et les questions qu'il suscite (Pasquini, 2018, 2019a)**



Le modèle ainsi conceptualisé s'inscrit au cœur du modèle premier où interagissent en cohérence les tâches d'apprentissage, les compétences issues des programmes et les pratiques d'évaluation, que nous centrons volontairement sur l'évaluation sommative (traits rouges dans la Figure 1). Cette dernière déploie ensuite ses quatre composantes (acolade en jaune) en les mettant en liens réciproques (doubles flèches en jaune). Ainsi, la référenciation restreinte décrit les choix réalisés dans les compétences à évaluer d'un point de vue sommatif. Le *design* (Wiggins & McTighe, 2005) explicite les processus de construction des tâches évaluatives, du point de vue des contenus sur lesquels elles portent et des habiletés cognitives que les élèves doivent mobiliser. 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 (Andrade, 2005 ; 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 (Brookhart, 2017 ; Marzano & Heflebower, 2011). Les questions proposées, que nous abordons avec les enseignants, concrétisent les liens de cohérence d'un point de vue taxonomique et de contenu entre chaque composante. Le modèle ainsi élargi – qui désigne le fort lien de cohérence systémique existant entre les compétences évaluées (référenciation restreinte), les tâches évaluatives (*design*), les points et/ou les critères attribués (pondération) et l'échelle (notation) (Pasquini, 2018, 2019a) - permet d'appréhender les conditions concrètes nécessaires à l'élaboration d'une note constructive.

## B. La planification à rebours

Pragmatiquement, nous favorisons l'appropriation de ce modèle en travaillant avec les enseignants sur la planification à rebours (Gauthier *et al.*, 2005 ; Tomlinson & McTighe, 2010) dont les étapes sont les suivantes :

- déterminer les compétences du programme à atteindre ;
- identifier les critères clés caractérisant l'atteinte de ces compétences ;

- construire des situations et des tâches d'apprentissage alignées avec les compétences et les critères ;
- collecter des informations sur les apprentissages des élèves tout au long de la séquence en lien avec les attentes formulées en pensant aux attentes relatives à la certification.

Ce travail demande de la flexibilité, s'adapte à la discipline et aux contextes ainsi qu'aux publics d'élèves. Il se réalise idéalement en équipe, jusqu'à l'élaboration détaillée de l'épreuve critériée finale qui vient à la fin du processus, en se basant sur la réflexion menée sur l'apprentissage en amont (Pasquini, 2019b).

### C. L'approche critériée

Dans un récent article théorique recensant un grand nombre d'études anglophones, Brookhart (2022) montre que les pratiques de notation constructives doivent s'appuyer sur des critères. Étant qualitatifs, les critères représentent l'unique outil permettant de développer des pratiques de notation à haute valeur informative. Dans le cadre des formations que nous menons, nous discutons quatre de leurs caractéristiques incontournables avec les enseignants (Balan et Jönsson, 2018) :

- En se centrant sur les apprentissages des élèves, l'évaluation critériée permet de formuler des rétroactions (ou *feedback*) soutenant l'apprentissage. Il est de fait plus constructif de dire à un élève sur quel critère il a par exemple échoué, plutôt que de lui décompter le nombre de points qui lui ont manqué pour obtenir une note satisfaisante ou supérieure.
- Référés à des compétences, les critères situent chaque élève par rapport à la maîtrise de ces derniers. La comparaison entre élèves est inexistante, ce qui évite de faire jouer à l'évaluation sommative un rôle de comparaison. La certification des apprentissages reprend donc ici le premier rôle, donnant à la note bilan une valeur pédagogique.
- Les critères étant communiqués dès que possible, les élèves sont rapidement et fréquemment confrontés aux contenus et aux apprentissages travaillés et évalués. Ils sont alors petit à petit en mesure de s'approprier une partie de la démarche évaluative. En cela, les critères deviennent des outils pour soutenir les apprentissages des élèves déjà pendant l'enseignement.
- La rétroaction fondée sur des critères permet à l'enseignant de mettre en évidence, à côté de la note chiffrée, ce que les élèves ont réussi, mais surtout ce qu'ils vont devoir encore travailler sous son guidage. La rétroaction est donc ici aussi constructive et qualitative, dans la mesure où elle se centre sur l'apprentissage (« nous allons reprendre la notion de schéma narratif », par exemple en français) et non sur la personne de l'élève (« tu aurais pu te donner un peu plus de peine »), ni sur un total de points (« avec deux points de plus, tu aurais eu 12 »), ces deux derniers commentaires n'ayant aucun impact sur les apprentissages des élèves (De Ketele & Freres, 2009).

En s'appuyant sur des critères, nous observons que l'enseignant, pour établir des seuils de notes, change son questionnement, qui passe de « à quel pourcentage de points met-on le seuil de suffisance ? » à « quelles preuves d'apprentissage l'élève doit-il fournir dans cette situation pour obtenir telle note, synonyme d'apprentissages en construction ou réalisés ? ». Dans une telle

perspective, les points peuvent toujours être exploités, mais ils s'apparient à des critères qui renvoient aux apprentissages évalués (Jönsson, 2014).

#### **D. Une critique des échelles de notation arithmétiques et des procédures de mesure**

Comme le dit McMillan (2019), « les pratiques de notation font référence à la manière dont les enseignants utilisent les informations provenant des épreuves sommatives pour déterminer et communiquer les notes des élèves » (traduction libre, p. 85). En cela, toute pratique de notation constructive est qualitative puisqu'elle cherche à identifier des indices de réussite. Dans cette perspective, nous amenons les enseignants à construire des barèmes de notation en référence à des critères, adaptés aux situations d'enseignement-apprentissage-évaluation, et qui tiennent compte des spécificités de la discipline, du temps d'enseignement et d'apprentissage effectif, du public d'élèves, de la complexité du contenu et de l'épreuve. En parallèle, le fait de partager les travaux de DeLandsheere (1992) avec les enseignants leur permet de devenir progressivement critiques face aux échelles numériques basées uniquement sur des points, car elles ont quatre faiblesses qui empêchent l'élaboration de notes constructives :

1. *L'échelle n'a ni zéro ni maximum naturels ou absolus.* Elle commence et finit n'importe où, au choix de celui ou celle qui l'utilise.
2. *Les degrés, ou échelons, ne sont pas de la même grandeur à l'intérieur d'une même échelle.* D'un point de vue d'apprentissage, il est effectivement impossible de démontrer que l'écart séparant la bonne de la très bonne connaissance d'un sujet est le même que l'écart séparant la connaissance moyenne de la bonne.
3. *Des degrés ou échelons correspondant d'une échelle à l'autre pour une même discipline ne sont pas de même grandeur.* Par exemple, malgré l'usage d'échelles arithmétiques, « bon » en expression orale en français pour une enseignante ou un enseignant n'est pas quantitativement égal à « bon » en expression orale en français pour un autre.
4. Des échelles portant sur des disciplines différentes ne sont pas comparables.

Guskey (2013) met également en évidence quelques problèmes des échelles élaborées uniquement sur la base de points, que nous mettons également en discussion (p. 70-72) :

- Par convention, les enseignants qui utilisent de telles échelles posent le plus souvent le seuil de suffisance autour de 60 % à 65 % du total des points. Le résultat est une échelle qui identifie potentiellement deux tiers des points pour décrire l'échec et seulement un tiers pour la réussite. Quel message donne-t-on aux élèves avec de tels outils ? Est-ce à dire qu'il est plus facile de différencier l'échec que la réussite ?
- Dans cette perspective, plus le nombre de points est important dans l'épreuve, plus le phénomène est ample. Comme DeLandsheere, l'auteur observe que de nombreux enseignants n'utilisent pas toutes les notes insuffisantes, ni celles qui décrivent l'excellence. Pourquoi alors en avoir autant ?
- Une croyance très répandue est que plus il y a de points à disposition pour calculer la note, plus cette dernière est précise et objective. Pourtant, dans une perspective qualitative, il est

impossible de différencier l'accumulation de 45 et de 50 points. Augmenter la gradation d'une échelle en pourcentages est donc un leurre de précision.

- Les échelles en pourcentages ne fonctionnent que sur l'exactitude des calculs des seuils des notes, sans rapport avec la maîtrise avérée des apprentissages des élèves dans les tâches évaluatives. Établir des seuils de manière appropriée doit plutôt combiner le jugement professionnel des enseignants en référence aux contenus abordés et aux habiletés cognitives requises par les tâches évaluatives, qui devraient être alignées avec les compétences *via* des critères (Tierney *et al.*, 2011).

Tous ces éléments sont abordés progressivement en formation, et participent à l'appropriation de la logique d'une évaluation soutien d'apprentissage et d'une note constructive.

## **II. Les enjeux que nous relevons avec les enseignants**

---

Lorsque nous accompagnons des enseignants, nous abordons trois enjeux majeurs relatifs au développement de pratiques de notation constructives.

### **A. L'évaluation sommative et le soutien des apprentissages des élèves**

Nous observons que l'évaluation notée est parfois exploitée à des fins non pédagogiques (Hadji, 2012) : maintenir les élèves dans le travail, les motiver à apprendre, gérer la discipline, asseoir le pouvoir de l'enseignant, ou encore récompenser (un effort par exemple). Elle est dès lors souvent au service d'une entreprise qui la dépasse. Or, en l'intégrant dans une réflexion systémique centrée sur les processus d'enseignement-apprentissage, nous soulignons que l'évaluation sommative notée participe à faire progresser chaque élève (Harlen 2005, 2012 ; Laveault et Allal, 2016 ; Mottier Lopez, 2015).

### **B. Des pratiques de notation en liens avec les objectifs pédagogiques**

De nombreux travaux montrent que plusieurs éléments interviennent dans les pratiques de notation : la culture de l'établissement (Duru-Bellat, 2010 ; Kelly, 2008), le niveau des élèves (Guskey, 2009), leur milieu socio-économique (Hattie, 2008 ; Klapp Lekholm, 2011), le degré de sévérité de l'évaluateur (Kitabgi, 2009), ses valeurs (Sun & Cheng, 2014), ses émotions (Brackett *et al.*, 2013 ; Tierney, 2015), son niveau de connaissances didactiques, pédagogiques et évaluatives (Xu & Brown, 2016) ou encore sa capacité de comprendre et d'interpréter la logique des prescriptions auxquelles il doit se soumettre (Tierney *et al.*, 2011) sont autant de freins potentiels à une construction de notes constructives. Par ailleurs, nous avons mis en évidence que les prises de décisions des enseignants lorsqu'ils notent font l'objet de multiples négociations entre différents facteurs aboutissant régulièrement à des dilemmes (Pasquini et DeLuca, 2021). Les enseignants reconnaissant ces biais, nous soulignons toujours avec eux que les notes ne devraient poursuivre que des objectifs pédagogiques : (1) certifier les apprentissages considérés comme importants dans les programmes et les établissements, (2) offrir des possibilités des rétroactions formatives aux élèves, et (3) donner des informations aux enseignants pour leur permettre de planifier les apprentissages à poursuivre (Carey & Carifio, 2012).

### **C. Les barèmes et les échelles : des outils à prendre au sérieux**

À ce niveau, nous engageons toujours avec les enseignants une démarche compréhensive sur les questions de l'élaboration des notes et des différentes techniques de manipulation, agrégation, conservation qu'ils leur font subir, ainsi que des genres de décisions auxquelles ces techniques conduisent (Dubus, 2006). Plus spécifiquement, nous accompagnons les enseignants pour qu'ils comprennent mieux sur quelles bases ils décident de la valeur d'une note – et plus particulièrement celle de la suffisance, le 10, qui rend compte du seuil de réussite, le plus important (Dwyer, 1996) – et comment se déploie l'ordonnement des notes sur les échelles.

Ces réflexions nous permettent également de remettre en question la construction de notes normatives, où l'élève est classé à partir de ses résultats par comparaison avec ceux de ses camarades. Nous soulignons le caractère discriminant de telles pratiques qui aboutissent à la mise en évidence des différences individuelles (Shepard *et al.*, 2018), poursuivent une visée de sélection et de compétition (par. ex., Brookhart, 2005 ; Mottier Lopez, 2015) et produisent toujours de l'échec (Hadji, 2016). Concrètement, l'évaluation normative se traduit par la création d'une courbe de Gauss ou « en cloche ». Elle est très présente en France selon Antiby (2003), qui l'a thématisée sous le nom de « constante macabre ».

### **III. Que nous apprennent les réflexions du corps enseignant aux prises avec la mise en œuvre d'une note constructive ?**

---

#### **A. Nuancer la qualité est parfois délicat**

Tous degrés et disciplines confondus, de nombreux enseignants nous partagent leur difficulté à qualifier graduellement la certification d'une compétence, ne sachant pas par exemple comment expliciter la différence entre une excellente et une bonne performance. En effet, plus l'échelle de notation est ample, plus la définition de chaque note en lien avec des critères est délicate (Guskey & Brookhart, 2019). Pour cela, nous les encourageons à définir les attentes au travers de critères pour signifier au moins la valeur de la note suffisante symbolisant le seuil entre réussite et échec.

#### **B. Élaborer des critères ne va pas de soi**

Dans de multiples situations d'évaluation, les enseignants que nous rencontrons travaillent et évaluent sans critère, ce qui tendrait à confirmer plusieurs résultats de recherche (Nordrum *et al.*, 2013). Cela est moins dû au fait que les enseignants ne les connaissent pas qu'à leur difficulté à expliciter concrètement leurs attentes. Ce travail de formulation explicite est souvent perçu comme difficile (Wyatt-Smith & Adie, 2021) et chronophage. Mais la démarche induite par la planification à rebours est ici aidante, même si elle est souvent jugée complexe car elle oblige à penser aux apprentissages puis aux tâches en rapport, alors que les habitudes vont souvent dans l'autre sens.

#### **C. Tenir compte des spécificités de sa discipline est fondamental**

Nous observons que le contenu disciplinaire joue un rôle majeur dans la possibilité, ou non, de construire des notes à haute valeur informative basées sur des critères (Duncan & Noonan, 2007 ; Meier *et al.*, 2006 ; Prøitz, 2013). Par exemple, au primaire et au secondaire, les disciplines des langues semblent mieux se prêter à l'exercice, notamment en compréhension et en production de l'écrit. Cela est probablement dû au fait qu'il est plus aisé de faire la différence entre une performance juste

suffisante ou bonne d'une compréhension d'œuvre littéraire en italien par exemple, mais que le même exercice semble plus délicat lorsqu'il s'agit de qualifier la bonne ou très bonne utilisation du cercle chromatique en arts visuels. Cela étant, nos dernières recherches montrent que des enseignants parviennent à critérier tout de même des notes dans des disciplines comme les mathématiques (Mrazek & Morales Villabona, 2022).

## **Conclusion : Deux éléments pour penser la formation**

---

### **A. La réception de la formation par les enseignants**

Les formations continues que nous dispensons sont toutes évaluées selon une démarche spécifique : les enseignants sont questionnés en ligne sur les apports et limites de la formation à la fin du dispositif, puis une seconde fois quelques mois plus tard, afin de voir dans quelle mesure les compétences travaillées sont mises en œuvre. Aussi, tous degrés confondus, nous observons que les enseignants avec lesquels nous collaborons soulignent la pertinence des apports proposés. Ils relèvent toutefois que la réflexion exigée est complexe, demande un temps d'appropriation et d'expérimentation conséquent et nécessite une coordination en équipe. C'est la raison pour laquelle nous privilégions des dispositifs collaboratifs en écoles, centrés sur les questions et problèmes relevés par les enseignants en fonction de leurs contextes. Nous travaillons depuis les pratiques réelles des participants, de manière à tenir compte de leurs élèves, de leurs contraintes et de leurs prescriptions. Cette méthode de travail nous permet d'accompagner chaque année la mise en œuvre de démarches dans plusieurs classes du primaire et du secondaire, qui montrent toutes que l'on peut conjuguer cohérence, transparence et exigence (par exemple, voir Pasquini, 2020).

### **B. Le changement de posture des enseignants**

Notre approche permet le plus souvent aux enseignants de modifier leurs conceptions de l'évaluation et de la notation, de prendre conscience des phénomènes de cohérence à ce niveau. Globalement, ils s'approprient plus profondément le programme, hiérarchisent les compétences à enseigner et à évaluer, élaborent des tâches évaluatives complexes, critérient en référence aux compétences évaluées et concentrent leur enseignement sur les contenus qu'ils jugent importants pour les progrès des élèves (Pasquini, 2018 ; Pasquini & Morales Villabona, 2022a). À ce titre, nous pensons que leur identité d'évaluateur (Looney *et al.* 2017) ainsi que leurs connaissances et compétences en évaluation (Xu & Brown, 2016) évoluent positivement (Pasquini & Morales Villabona, 2022b). Des tensions restent toutefois toujours présentes dans leurs pratiques de notation, notamment dans leur difficulté à mettre en œuvre des démarches qualitatives dans des environnements et des équipes aux pratiques traditionnelles, centrées sur les notes strictes et les moyennes (Mottier Lopez *et al.*, 2012).

## Références

---

- Anderson, L. W. (2002). Curricular alignment : A re-examination. *Theory into Practice*, 41(4), 255-260. [https://doi.org/10.1207/s15430421tip4104\\_9](https://doi.org/10.1207/s15430421tip4104_9)
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-31.
- Antibi, A. (2003). *La constante macabre ou comment a-t-on découragé des générations d'élèves*. Math'Adore.
- Balan, A. et Jönsson, A. (2018). Increased explicitness of assessment criteria: effects on student motivation and performance. *Frontiers in Education*, 3(81), 56-68.
- Brackett, M. A., Floman, J. L., Ashton-James, C., De, L. & Salovey, P. (2013). The influence of teacher emotion on grading practices: a preliminary look at the evaluation of student writing. *Teachers and Teaching*, 19(6), 634-646. <https://doi.org/10.1080/13540602.2013.827453>
- Brookhart, S. M. (2013). Grading. Dans J. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 257-271). SAGE Publications.
- Brookhart, S. M. (2017). *How to use grading to improve learning*. ASCD.
- Brookhart, S. M. (2022). Connecting grading with other assessment practices. *La Revue LEE* (*Assessment in Education Journal*), (6). <https://doi.org/10.48325/rlee.006.01>
- Butera, F., Buchs, C. et Darnon, C. (2011). *L'évaluation, une menace ?* PUF.
- Carey, T. et Carifio, J. (2012). The minimum grading controversy: Results of a quantitative study of seven years of grading from an urban high school. *Educational Researcher*, 41(6), 201-208. <https://doi.org/10.3102/0013189X12453309>
- Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire ?* De Boeck.
- De Ketele, J.-M. et Freres, G. (2009). Les commentaires des enseignants et des élèves : simples jugements ou processus évaluatifs ? *Les Cahiers de Recherche en Éducation et Formation*, 74, 4-32.
- Dubus, A. (2006). *La notation des élèves. Comment utiliser la docimologie pour une évaluation raisonnée*. Armand Colin.
- Duncan, R. C. et Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *The Alberta Journal of Educational Research*, 53(1), 1-21.
- Duru-Bellat, M. (2010). Ce que la mixité fait aux élèves. *Revue de l'OFCE*, 3 (114), 197-212. <https://doi.org/10.3917/reof.114.0197>
- Dwyer, C. A. (1996). Cut scores and testing: statistics, judgment, truth, and error. *Psychological Assessment*, 8(4), 360-362. <https://doi.org/10.1037/1040-3590.8.4.360>
- 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 du Centre de recherche interuniversitaire sur la formation et la profession enseignante). CRIFPE, Université Laval.

- Guskey, T. R. (2009, avril). Bound by tradition: Teachers' views of crucial grading and reporting issues [communication orale]. *American Educational Research Association*, San Francisco, Etats-Unis.
- Guskey, T. R. (2013). The case against percentage grades. *Educational, School, and Counseling Psychology Faculty Publications*, 22. [https://uknowledge.uky.edu/edp\\_facpub/22](https://uknowledge.uky.edu/edp_facpub/22)
- Guskey, T. R. et Brookhart, S. M. (Ed.) (2019). *What we know about grading*. ASCD.
- Guskey, T. R. et Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, 26(3), 303-320. <https://doi.org/10.1080/0969594X.2018.1555515>
- Hadji, C. (2012). *Faut-il avoir peur de l'évaluation ?* De Boeck.
- 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.
- Harlen, W. (2005). Teacher's summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16(2), 207-223. <https://doi.org/10.1080/09585170500136093>
- Harlen, W. (2012). On the relationship between assessment or formative and summative purposes. Dans J. Gardner (dir.), *Assessment and learning* (p. 87–102). Sage.
- Hattie, J. (2008). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Jönsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment et Evaluation in Higher Education*, 39(7), 840-852. <https://doi.org/10.1080/02602938.2013.875117>
- Kelly, S. (2008). What types of students' efforts are rewarded with high marks? *Sociology of Education*, 81(1), 32-52. <https://doi.org/10.1177/003804070808100102>
- Kitabgi, S. (2009). *L'évaluation scolaire est-elle au service de l'orientation ?* Centre d'orientation de CCIP. <https://jmb21470.pagesperso-orange.fr/mysegpa21/documents/Etude-evaluation-scolaire-orientation-2009.pdf>
- Klapp Lekholm, A. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian Journal of Educational Research*, 55 (6), 587–608. <https://doi.org/10.1080/00313831.2011.555923>
- Kohn, A. (2011). The case against grades. *Effective Grading Practices*, 69(3), 28-33.
- Laveault, D. et Allal, L. (Eds.). (2016). *Assessment for learning: Meeting the challenge of implementation*. Springer International.
- Looney, A., Cumming, J., van Der Kleij, F. et Harris, K. (2017): Reconceptualising the role of teachers as assessors : teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467. <https://doi.org/10.1080/0969594X.2016.1268090>
- Martone, A. et Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332-1361. <https://doi.org/10.3102/0034654309341375>
- Marzano, R. J. et Heflebower, T. (2011). Grades that show what students know. *Effective Grading Practices*, 69(3), 34-39.

- McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. Dans T. R. Guskey et S. M. Brookhart (dir.), *What we know about grading* (p. 84-112). ASCD.
- Meier, S. L., Rich, B. S., Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: factors related to discrepancies in scoring. *Assessment in Education: Principles, Policy & Practice*, 13(1), 69-95. <https://doi.org/10.1080/09695940600563512>
- Merle, P. (2012). L'évaluation par les notes : quelle fiabilité et quelles réformes ? *Regards croisés sur l'économie*, 2(12), 218-230. <https://doi.org/10.3917/rce.012.0218>
- Moss, C. (2013). Research on classroom summative assessment. Dans J. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 235-256). SAGE Publications.
- Mottier Lopez, L., Tessaro, W., Dechamboux, L. et 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*, 6(18). <https://doi.org/10.4000/questionsvives.1235>
- Mottier Lopez, L. (2015). *Évaluations formative et certificative des apprentissages : enjeux pour l'enseignement*. De Boeck.
- Mrazek, J. et Morales Villabona, F. (2022, prépublication). Évaluation critériée et résolution de problèmes numériques. Un récit de pratique accompagné de pistes de réflexion sur l'évaluation constructive. *La Revue LEE*, 6. <http://revue.leeonline/index.php/info/article/view/137>
- Nordrum, L., Evans, K. et Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. *Assessment and Evaluation in Higher Education*, 38, 919-940. <https://doi.org/10.1080/02602938.2012.758229>
- Organisation for Economic Cooperation and Development (OECD) (2012). *Grade Expectations: How marks and education policies shape student's ambitions*. PISA, OECD Publishing. <http://dx.doi.org/10.1787/9789264187528-en>
- Pasquini, R. (2019a). 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. (2019b). Construire des évaluations sommatives communes : une mauvaise « bonne idée » ? *Éducateur*, 6, 4-5.
- Pasquini, R. (2021). *Quand la note devient constructive. Évaluer pour certifier et soutenir les apprentissages*. Presses de l'Université Laval
- Pasquini, R. et Deluca, C. (2021). Grading in a dilemmatic space: An exploratory cross-cultural analysis of Mathematics and Language Secondary teachers. *Comparative and International Education/Éducation comparée et internationale*. 49(2), 51-70. <https://doi.org/10.5206/cieeci.v50i1.14133>
- Pasquini, R. et Morales Villabona, F. (2022a). Comment appréhender une littératie en évaluation pour documenter et comprendre le développement des compétences des enseignants ? *Revue canadienne de l'éducation*, 45(2). 512-544. <https://cje-rce.ca/wp-content/uploads/sites/2/2021/03/45-2-9-5171-Pasquini-512-554.pdf>

Pasquini, R. et Morales Villabona, F. (2022b, 22 avril). *When Assessment for Learning Meets Teachers' Assessment Culture: Issues and Prospects* [communication orale dans un symposium]. Colloque de l'AERA 2022, San Diego, Etats-Unis.

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

Quittre, V., Dupont, V., et Lafontaine, D. (2019). Enseigner au quotidien. Talis 2018. <http://www.enseignement.be/index.php?page=28161>

Shepard, L. A., Penuel, W. R. et Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21-34. <https://doi.org/10.1111/emip.12189>

Sun, Y. et Cheng, L. (2014). Teachers' grading practices: meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326-343. <https://doi.org/10.1080/0969594X.2013.768207>

Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, 8, 5-30. <http://dx.doi.org/10.18296/am.0002>

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

Tomlinson, C. A. et McTighe, J. (2010). *Intégrer la différenciation pédagogique et la planification à rebours*. Chenelière Éducation.

Wyatt-Smith, C. et Adie, L. (2021). The development of students' evaluative expertise: enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, 53(4), 399-419. <https://doi.org/10.1080/00220272.2019.1624831>

Xu, Y., et Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.