

Mettre en œuvre des pratiques d'évaluation sommative notée au service des apprentissages de tous les élèves

Une étude de cas à l'école secondaire

*Implementing graded summative assessment practices to
support learning for all students
A case study from a secondary school*

Fernando Morales Villabona – fernando.morales-villabona@hepl.ch –
<https://orcid.org/0009-0003-7644-8186>

Raphaël Pasquini – raphael.pasquini@hepl.ch – <https://orcid.org/0000-0002-0230-2083>

Bastyan Strehmel – bastyan.strehmel@hepl.ch – <https://orcid.org/0009-0008-8595-3172>

Haute École Pédagogique du canton de Vaud (Suisse)

Pour citer cet article : Morales Villabona, F., Pasquini, R., & Strehmel, B. (2023). Mettre en œuvre des pratiques d'évaluation sommative notée au service des apprentissages de tous les élèves. Une étude de cas à l'école secondaire *Évaluer. Journal international de recherche en éducation et formation*, 9(2), 1-26. <https://doi.org/10.48782/e-jiref-9-2-5>

Résumé

Les développements récents de l'Évaluation-soutien d'Apprentissage (E-sA – Allal & Laveault, 2009) proposent que toute démarche évaluative, y compris dans la certification, devrait être en mesure de fournir des éléments susceptibles de faire progresser les élèves. Les principes de l'E-sA suscitent toutefois nombre de questions chez les enseignants car ils ont régulièrement de la difficulté à les concrétiser dans leurs pratiques évaluatives. Ces constats se confirment davantage lorsque les démarches d'évaluation sommative s'inscrivent dans des classes hétérogènes, où les élèves sont, par exemple, regroupés en plusieurs niveaux.

Nous présentons une étude de cas (Roy, 2009) à l'école secondaire, à propos d'une séquence sur la compréhension d'une œuvre écrite en français, menée dans une classe de 10^e année (13-14 ans) composée de dix-neuf élèves répartis en deux niveaux d'exigences certificatives distincts. Deux types de données ont été récoltés : des artefacts (Lock et al., 2021) produits par l'enseignant (planification de la séquence, outils d'évaluation formative et sommative) et des productions d'élèves. Nous interprétons ces données suivant une méthode d'analyse en mode écriture (Paillé & Muchielli, 2013).

Nos résultats soulignent la complexité des pratiques d'évaluation sommative notée lorsqu'elles visent à soutenir les apprentissages de tous les élèves. Tout en soulevant quelques tensions, nous observons qu'une réflexion sur les apprentissages poursuivis et les attentes pressenties, concrétisées dans des critères d'évaluation, est fondamentale : ce sont les critères qui permettent à l'enseignant de travailler les mêmes contenus avec tous les élèves de manière différenciée et ainsi d'adapter les seuils de réussite pour chaque niveau.

Mots-clés

Évaluation sommative, Évaluation-soutien d'apprentissage, Évaluation différenciée, Hétérogénéité des élèves, Différenciation pédagogique.

Abstract

Recent developments in Assessment for Learning (AfL – Allal & Laveault, 2009) suggest that any assessment practice, including those leading to certification, must be able to provide elements likely to help students progress. However, the AfL principles raise several questions among teachers because they often encounter difficulties implementing them in their assessment practices. These findings are further confirmed when the summative assessment practices take place in heterogeneous classes, where the students are grouped into different levels, for example. We present a case study (Roy, 2009) at secondary school, about a sequence on written comprehension in French, carried out in a 10th year class (13-14 years old) made up of nineteen students divided into two distinct levels of certification requirements. Two types of data were collected: artifacts (Lock et al., 2021) produced by the teacher (sequence planning, formative and summative assessment tools) and student productions. We interpret these data using a method of analysis in writing mode (Paillé & Muchielli, 2013). Our results highlight the complexity of grading and summative assessment practices when they aim to support the learning of all students. While raising some tensions, we observe that a reflection on the learning goals and desired achievements, concretized in assessment criteria, is fundamental: these criteria allow the teacher to work on the same contents with all the students in a differentiated way and thus adapt the success thresholds for each level.

Keywords

Summative assessment, Assessment for Learning, Differentiated assessment, Student heterogeneity, Differentiated instruction.

1. Problématique

Depuis une vingtaine d'années, confrontés à une hétérogénéité grandissante des élèves et à une évolution continue des exigences du monde professionnel et académique, la majorité des systèmes éducatifs de l'Organisation de Coopération et de Développement Économique (OCDE) travaillent à la mise en œuvre d'une évaluation au service des apprentissages de tous les élèves, de manière à lutter contre l'échec scolaire et favoriser les progrès de chacun (OCDE, 2013). Black (1986) est le premier à parler d'*Assessment for Learning* pour appréhender cette nouvelle acception de l'évaluation, en renforçant la conceptualisation de l'évaluation formative. Puis, l'ouvrage de Gipps (1994), *Beyond Testing*, marque un changement radical de paradigme entre la vision psychométrique et celle de l'évaluation pédagogique, en promulguant pour la première fois que l'évaluation sommative devrait également être au service des apprentissages des élèves.

La mise en œuvre de cette Évaluation-soutien d'Apprentissage (E-sA – Allal & Laveault, 2009) s'intensifie depuis peu en Suisse romande, où les prescriptions des sept systèmes éducatifs cantonaux y font référence : l'idée-force est que toute démarche d'évaluation, y compris dans la certification, doit être en mesure de fournir des éléments susceptibles de faire progresser les élèves. Mais comment mettre en œuvre ce principe, notamment dans des classes hétérogènes ? Cette question vive de la pratique fait écho à certains résultats marquants relevés par la recherche. Bien que de nombreux travaux reconnaissent les effets positifs de l'évaluation formative sur les apprentissages des élèves (Lui & Andrade, 2022 ; Mottier Lopez, 2015 ; Wiliam, 2011), certaines études soulignent la difficulté du corps enseignant à la mettre en œuvre (Kingston & Nash, 2011 ; Moss, 2013) : c'est souvent une compréhension superficielle des principes de l'E-sA qui prédomine chez le corps enseignant, et qui est la cause de leur mise en œuvre inefficace (Klenowski, 2009).

Au niveau des pratiques d'évaluation sommative dont il sera question ici, des problèmes importants sont également relevés. Tout d'abord, c'est cette forme d'évaluation qui est la plus courante à l'école (Braxmeyer et al., 2005 ; OCDE, 2013) et l'on y observe plutôt des pratiques traditionnelles (Mottier Lopez, 2013 ; Ohlsen, 2007), c'est-à-dire centrées sur l'établissement d'une note et le calcul de moyennes, sans réelles opportunités de rétroaction. Plus spécifiquement, les tâches des épreuves sommatives n'ont régulièrement qu'un lien ténu avec les situations d'apprentissage menées en classe (Biggs, 2003) et les critères y sont fréquemment absents (Nordrum et al., 2013). Quant aux pratiques de notation, des études montrent que les notes portent souvent sur des éléments qui n'ont rien à voir avec la certification des apprentissages (Brookhart et al., 2016) et leur élaboration relève davantage d'un « fatras » (Cross & Frary, 1999 ; Pasquini, 2021) que de pratiques modélisées. En effet, la construction de la note est peu objectivée (Hadjji, 2016), obéit le plus souvent à l'utilisation d'algorithmes mathématiques (Guskey, 2013, 2022) et sert parfois davantage à classer les élèves qu'à certifier leurs apprentissages (Butera et al., 2012).

Ces constats se confirment encore davantage lorsque les démarches d'évaluation sommative notée s'inscrivent dans des classes où les élèves sont, par exemple, regroupés en plusieurs niveaux, avec des exigences différentes en matière de certification, comme dans notre cas. Toutefois, la recherche est assez discrète lorsqu'il s'agit d'appréhender cette problématique (Noman & Kaur, 2014), se bornant souvent à souligner que, même dans des situations d'enseignement différencié, les démarches d'évaluation sommative restent les mêmes pour tous les élèves (Harris Stefanakis, 2010). Dans ce sens, notons cependant que

les quelques travaux dont nous disposons sur ce sujet sont issus de la pédagogie spécialisée (Brina & Dechamboux, 2021 ; Jung & Guskey, 2010 ; Wakeman et al., 2022), car les différences entre élèves y sont institutionnellement reconnues et appréhendées de manière spécifique.

Au regard de ces éléments de problématisation, notre question de recherche est la suivante : à quelles conditions mettre en œuvre une démarche d'évaluation sommative notée pouvant soutenir les décisions de l'enseignant et les apprentissages des élèves dans une classe à double niveau ? Pour la traiter, nous commençons par dresser notre cadre conceptuel en y définissant les éléments théoriques clés de notre étude. La méthodologie présente ensuite le contexte dans lequel nous avons travaillé, les caractéristiques de notre étude de cas, ainsi que nos données et outils d'analyse. Les résultats présentent les outils d'évaluation construits et analysent leur cohérence. Une discussion s'ensuit, qui relève plusieurs enjeux présents dans la mise en œuvre de cette démarche. Enfin, nous terminons avec une conclusion mettant notamment en évidence les atouts et limites du présent travail, ainsi que des pistes de recherches futures.

2. Cadre conceptuel

2.1. Évaluer pour soutenir l'apprentissage

Les développements récents de l'E-sA (Allal & Laveault, 2009) amènent à une conception plus large de l'évaluation, sous différentes formes, qui, indépendamment de leur fonction première (formative ou sommative), permettent *in fine* de soutenir la progression des élèves. Dans ce cadre, Laveault et Allal (2016) proposent une définition opérationnelle de l'E-sA, comme un processus de « collecte et interprétation d'informations dont l'utilisation intentionnelle permet aux enseignants et aux élèves, agissant de manière individuelle ou interactive, de prendre des décisions ayant un impact positif sur l'enseignement et l'apprentissage » (p. 7, trad. libre). De ce fait, ces auteurs soulignent que l'utilisation faite des informations récoltées – à savoir, soutenir les apprentissages des élèves – est plus importante que le moyen utilisé pour les recueillir. Autrement dit, ce qui caractérise la perspective de l'E-sA « n'est pas à rechercher dans le format, ni dans les circonstances de la prise d'information sur l'apprentissage de l'élève, mais plutôt dans l'effet positif qu'elle engendre pour l'apprenant » (Allal & Laveault, 2009, p. 102).

Notre compréhension de l'évaluation sommative (que nous explicitons ci-après) s'inscrit pleinement dans cette perspective, rejoignant ainsi les dix principes édictés par l'Assessment Reform Group (Broadfoot et al., 2002). Nous exploiterons particulièrement quatre d'entre eux dans cet article, à savoir que l'E-sA :

- s'intègre à la planification de l'enseignement ;
- favorise une compréhension partagée (entre l'enseignant et les élèves) des objectifs d'apprentissage et des critères de réussite ;
- permet de fournir aux élèves des rétroactions constructives et des opportunités pour progresser ;
- vise à améliorer les opportunités d'apprentissage et de réussite de tous les élèves.

2.2. L'évaluation sommative au service des apprentissages des élèves

L'évaluation sommative permet d'établir un bilan qui certifie les apprentissages clés réalisés par l'élève (Brookhart, 2017) à la fin d'une étape dans un cursus de formation (Allal, 2008). Se référer à une évaluation sommative qui se met au service des apprentissages des élèves signifie que celle-ci doit comporter une haute valeur informative, autant pour l'enseignant que pour les élèves, sur les apprentissages évalués (Hadji, 2017) et peut ainsi aboutir à des régulations (Brookhart, 2022). Dans cette optique, les démarches sommatives demandent à être construites dans une perspective qualitative et en respectant une cohérence curriculaire (Pasquini, 2019).

2.2.1. Une évaluation sommative cohérente : l'alignement curriculaire élargi

Le modèle de l'alignement curriculaire d'Anderson (2002) insiste sur le fort lien de cohérence qui doit exister entre les objectifs d'apprentissage du curriculum prescrit, les activités d'apprentissage menées en classe et les démarches d'évaluation. D'un point de vue évaluatif, il s'agit pour l'enseignant de s'assurer que les démarches : 1) portent sur des contenus définis par un curriculum et 2) correspondent à ce qui a été réellement travaillé en classe, en matière de contenus, mais également d'habiletés cognitives sollicitées et entraînées au travers des tâches d'apprentissage proposées aux élèves.

Sur cette base et dans le cadre d'une évaluation sommative notée, le modèle théorique de l'alignement curriculaire élargi (Pasquini, 2019) permet de soutenir la construction de démarches évaluatives cohérentes et dont les résultats finaux (les notes) sont explicitement référés à des apprentissages. Ce modèle, développé dans une perspective systémique et contextuelle, met en relation quatre composantes de l'évaluation sommative :

- la *référenciation restreinte*, qui renvoie aux objectifs d'apprentissage sélectionnés pour être évalués de manière sommative (parmi les objectifs travaillés en classe) ;
- le *design*, qui concerne les tâches évaluatives choisies ou construites pour récolter des informations pertinentes à propos de la maîtrise des objectifs retenus ;
- la *pondération*, qui permet de mettre en évidence le poids attribué aux apprentissages évalués, à travers d'outils comme les points ou, dans une perspective plus qualitative, les critères d'évaluation ;
- la *notation*, qui fait référence aux démarches relatives à la construction de la note, notamment dans la définition des seuils de chaque note relativement à des points ou à des critères (création d'un barème), fidèlement à une échelle donnée¹.

Les rapports entre ces différentes composantes se concrétisent au travers d'un ensemble de questions à destination des enseignants (cf. figure 1). Le modèle de l'alignement curriculaire élargi permet alors d'objectiver et comprendre la manière dont les contenus et habiletés cognitives qui composent les objectifs d'apprentissage évalués peuvent guider le processus de construction de la note (grâce notamment à la définition du seuil de réussite de l'épreuve – Dwyer, 1996), en cohérence avec les critères d'évaluation définis et avec les tâches d'évaluation retenues (Pasquini, 2019, 2021).

¹ Dans notre contexte (canton de Vaud, en Suisse romande), par exemple, dès la 5^e année, les résultats des épreuves sommatives sont exprimés sous la forme de notes allant de 1 à 6, avec demi-points, la note 4 correspondant au seuil de réussite.

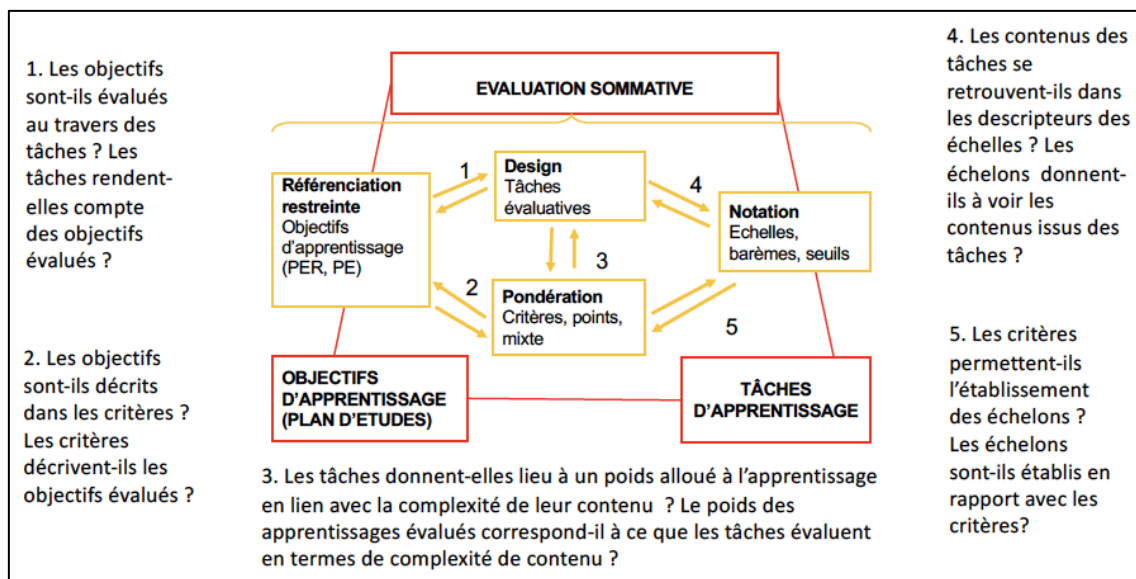


Figure 1. Modèle de l'alignement curriculaire élargi (Pasquini, 2019)

Toujours dans une perspective d'alignement curriculaire, les articulations entre les démarches d'évaluation sommative et les démarches d'évaluation formative, entreprises tout au long de l'enseignement à des fins de régulation, apparaissent essentielles. Pouvant prendre différentes formes (Laveault & Allal, 2016 ; Mrazek & Morales Villabona, 2022), ces synergies favorisent la construction d'une cohérence tout au long du processus d'enseignement-apprentissage, guidée par des attentes précises en matière d'apprentissages (Harlen, 2005). Ces dernières renvoient aux critères d'évaluation, des éléments fondamentaux auxquels nous nous intéressons ci-après.

2.2.2. Une évaluation sommative référée à des contenus essentiels : le rôle des critères

Lorsque l'appréciation des apprentissages des élèves se réalise sur la base des niveaux d'atteinte d'un objectif ou d'un critère, l'évaluation est dite *critériée* ou à *référence critériée* (Allal, 2008 ; Balan & Jönsson, 2018). Celle-ci est donc profondément liée à la définition de l'apprentissage en jeu et, par ce biais, à la mise en valeur de certaines dimensions de cet apprentissage qui se traduisent dans des critères d'évaluation (Mottier Lopez, 2015).

Le rôle clé des critères est constamment mis en avant par la recherche (Andrade & Valtcheva, 2009 ; Braund & Deluca, 2018 ; Brookhart, 2022 ; Fletcher, 2016), que ce soit dans une visée formative ou sommative. A titre d'exemple, Balan et Jönsson (2018) indiquent que le partage des critères d'évaluation avec les élèves peut faciliter leur introduction progressive dans des démarches d'autoévaluation ou d'évaluation par les pairs. Il apparaît également primordial de définir et de communiquer les critères d'évaluation aux élèves le plus tôt possible (McMillan & Hearn, 2008 ; Pasquini, 2021). A ce titre, l'évaluation critériée favorise la transmission de feedbacks clairs aux élèves, référés explicitement à des objectifs d'apprentissage et focalisés sur les moyens que l'élève pourrait utiliser pour progresser.

La définition des critères permet de rendre explicites les apprentissages évalués et, par ce biais, d'établir des liens avec le travail réalisé en classe. Pour ce faire, le modèle de la *planification à rebours* (Tomlinson & McTighe, 2010 ; Wiggins & McTighe, 2005) est porteur. Il consiste, dans une perspective dynamique et contextuelle, à déterminer les objectifs d'apprentissage visés et à établir les critères de réussite pour caractériser l'atteinte de ces objectifs. Ces éléments guident ensuite le choix des tâches les plus propices pour récolter des informations sur les apprentissages des élèves tout au long de la séquence et en lien avec les attentes relatives à l'évaluation sommative finale.

Lorsque la note est construite sur la base des critères définis, il devient possible de transmettre une rétroaction qui accompagne le résultat, donnant ainsi à l'élève l'occasion de se situer dans la maîtrise de l'objectif évalué. A ce niveau, les rapports entre les processus de pondération et de notation apparaissent autant essentiels que complexes. En effet, en plus de définir les critères d'évaluation, il s'agit pour l'enseignant de déterminer le poids qui leur sera assigné en fonction de leur importance relativement aux apprentissages menés (Marzano, 2002 ; Pasquini, 2021). Concrètement, c'est sur cette base que le barème et le seuil de réussite de l'épreuve peuvent être décidés.

Cette recherche d'alignement curriculaire élargi devient d'autant plus complexe lorsque l'hétérogénéité des élèves s'avère particulièrement saillante. C'est pour cette raison que nous nous attardons, dans les lignes qui suivent, à décrire la manière dont cette hétérogénéité peut être saisie dans le cadre des évaluations sommatives, en nous appuyant sur des travaux qui portent sur l'idée d'une évaluation différenciée.

2.3. L'évaluation différenciée

Dans cet article, nous parlerons de différenciation pédagogique, créée par l'enseignant, voire par des établissements, pour tenir compte des différences entre les élèves (Allal, 2017). A l'instar de Moon (2005), Tomlinson (2005) ou Lysaght et ses collègues (2019), nous concevons la différenciation au sein d'un alignement curriculaire. Dans une telle approche, il s'agit pour les enseignants de veiller notamment à ce que : 1) les objectifs d'apprentissage soient clairement définis, pour assurer une centration sur les dimensions essentielles de l'objet disciplinaire concerné, et 2) les démarches d'évaluation, pensées au service des apprentissages des élèves, fournissent également des informations utiles à la planification et à la régulation de l'enseignement.

En cohérence avec les éléments théoriques présentés plus haut, la différenciation pédagogique s'inscrit dans une conception intégrée et systémique des processus d'enseignement, d'apprentissage et d'évaluation. A ce titre, penser l'évaluation du point de vue de la différenciation implique une reconnaissance de la différence en classe pour éviter l'accroissement des écarts entre les élèves. Dès lors, l'évaluation différenciée est conçue comme « un moyen pour l'enseignant de prendre en compte la diversité des élèves dans la logique d'une école inclusive » (Dubois, 2016, p. 80). De plus, reconnaître et exploiter cette diversité au sein de la classe passe par la conception de situations suffisamment flexibles pour permettre à tous les élèves de progresser (Prud'homme & Bergeron, 2012). Ainsi, dans une logique d'évaluation différenciée, les enseignants sont amenés à valoriser des prises d'information formelles et informelles et à varier les modalités et outils d'évaluation (Ali, 2015).

Mais la mise en œuvre d'une évaluation différenciée, notamment si elle est sommative, est loin d'être simple, notamment lorsqu'il s'agit de noter les épreuves des élèves (Lang, 2019).

Un des enjeux pour l'enseignant est de concevoir cette articulation entre différenciation et pratiques d'évaluation sommative notée, au regard des spécificités et contraintes du contexte. Sachant qu'une centration sur l'évaluation sommative en classe a tendance à augmenter les différences entre élèves (Harlen, 2005), la question des liens entre évaluation sommative et différenciation pédagogique est primordiale (Allal, 2017) et complexe. L'étude que nous présentons dans ce texte a pour but justement d'y apporter des éléments de réponse. Pour avancer dans la réflexion, nous nous appuyons sur des propositions comme celles de Tomlinson (2005), qui insiste sur deux conditions en particulier :

- a) la notation doit se référer à des critères précis en matière d'apprentissage (puisque la plupart des problèmes dans la pratique de la notation viennent souvent d'un manque de clarté de ce sur quoi la note porte) ;
- b) éviter la « pollution » des notes, résultant de plusieurs facteurs comme des consignes provoquant des malentendus, des outils d'évaluation standardisés ou encore des ruptures de l'alignement curriculaire (par exemple, entre les objectifs visés et l'outil d'évaluation construit).

A la lumière de ces apports et en cohérence avec nos ancrages théoriques, nous considérons que l'évaluation sommative différenciée se concrétise dans des pratiques que les enseignants ajustent en fonction des caractéristiques de leurs élèves (Jung & Guskey, 2010) et que ces ajustements se concrétisent dans l'exploitation de démarches critériées. A ce titre, l'étude des prises de décision des enseignants en matière d'évaluation est fondamentale (McMillan & Nash, 2000).

3. Cadre méthodologique

3.1. Contexte

Notre recherche s'intéresse à une pratique d'évaluation sommative conçue et mise en place dans la classe de Bastyan, enseignant de français à l'école secondaire.

3.1.1. L'enseignant

Bastyan enseigne le français depuis 2011 dans un établissement secondaire urbain, dans le Canton de Vaud. Depuis plusieurs années, il s'interroge sur la pertinence des épreuves sommatives qu'il propose aux élèves et, de manière plus globale, sur l'impact qu'évaluer peut avoir sur leurs apprentissages. Depuis qu'il a suivi un module de formation initiale dans lequel lui a été présenté le modèle de l'alignement curriculaire élargi (Pasquini, 2019), Bastyan s'efforce de développer des démarches d'évaluation sommative critériées. Par ailleurs, l'enseignant s'intéresse de plus en plus à l'évaluation formative et à son articulation avec l'évaluation sommative dans la planification de ses séquences.

Dans le cadre du projet de recherche dont il est question ici, il a participé à une communauté d'apprentissage professionnelle avec cinq autres enseignants, encadrés par deux chercheurs, qui ont échangé sur leurs pratiques évaluatives. Bastyan souhaitait travailler sur les conditions nécessaires à la mise en œuvre de tâches complexes, lors de démarches évaluatives cohérentes et critériées, afin de rendre compte qualitativement des apprentissages des élèves dans la compréhension de l'écrit en français. Son but était notamment d'éviter une épreuve mobilisant des habiletés cognitives de basse complexité (comme la restitution – Anderson & Krathwohl, 2001). C'est en partant de ce constat que

Bastyan a voulu aborder l'évaluation de la compréhension de l'écrit d'une œuvre littéraire, à l'aide de tâches complexes et de manière différenciée.

3.1.2. La classe

Depuis 2013, à la suite de la mise en place de la dernière loi scolaire de l'école obligatoire vaudoise (LEO, 2011), les élèves du secondaire 1 sont répartis dans deux voies aux exigences distinctes : la voie Prégymnasiale (VP) pour les élèves ayant obtenu les meilleurs résultats à la fin du primaire et la voie Générale (VG) pour les autres. Au sein de la VG, les élèves bénéficient d'un enseignement différencié en allemand, mathématiques et français où ils sont répartis en deux niveaux : le niveau 1 comporte les exigences dites de base et le niveau 2 les exigences supérieures.

Lors de l'année scolaire 2020-2021, Bastyan enseignait le français à une classe de dix-neuf élèves de 10^e année VG (13-14 ans), répartis en deux niveaux d'exigence certificative distincts : neuf de niveau 1 et dix de niveau 2. Les élèves ont été répartis dans ces niveaux sur la base de leur moyenne annuelle obtenue l'année scolaire précédente. Par ailleurs, dans cette classe « ordinaire », aucun élève ne bénéficiait d'aménagements ni d'adaptations², les conditions définies par les directives officielles n'étant pas réunies.

3.2. Données récoltées et méthode d'analyse

Nous présentons ici une étude de cas (Albarello, 2011 ; Roy, 2009) à propos d'une séquence sur la compréhension d'une œuvre écrite en français à l'école secondaire. Le contexte est celui de la classe de 10^e année de Bastyan, présenté ci-dessus. Nous avons choisi l'étude de cas comme méthode de recherche car elle permet une compréhension profonde des phénomènes dans leur contexte, assurant une forte validité interne, sans prétention de généralisation (Gagnon, 2005).

Deux types de données ont été récoltés par l'enseignant lui-même (qui est aussi l'un des auteurs de cet article) dans une logique de participation observante (Van der Maren, 1999), soutenue par des échanges avec des chercheurs et d'autres enseignants dans le cadre d'un projet de recherche et de développement professionnel en évaluation : des artefacts (Lock et al., 2021) produits par l'enseignant (planification de la séquence, outils d'évaluation formative, épreuve sommative, feedbacks écrits) et des productions d'élèves (en cours d'apprentissage et lors de l'évaluation sommative). Dans une approche heuristique (de Saint-André et al., 2010), nous avons adhéré au postulat de faire de la recherche « avec » plutôt que « sur » les praticiens, visant une interfécondation entre les communautés scientifique et professionnelle pour la construction de savoirs au service d'une pratique éclairée (Desgagné & Bednarz, 2005).

Nous avons interprété les données récoltées par la méthode d'analyse en mode écriture, telle que proposée par Paillé et Muchielli (2013), dans laquelle un travail délibéré d'écriture

² Dans notre contexte, les prescriptions institutionnelles en matière d'évaluation sommative (https://www.vd.ch/fileadmin/user_upload/organisation/dfj/dgeo/fichiers_pdf/CGE/CGE.pdf) distinguent les *aménagements* (modification des modalités d'évaluation mais non des objectifs ni des barèmes) des *adaptations* (ajustement des objectifs, pour les faire correspondre à un programme personnalisé élaboré pour l'élève).

et de réécriture (ici, partagé par les trois auteurs du texte) permet la reformulation, l'explicitation et l'interprétation du matériau à l'étude (sans recours à d'autres moyens de réduction ou de codage des données). Selon ces auteurs, cette méthode d'analyse dynamique (dépassant le statisme des catégorisations isolées) permet, à la fois, d'exposer et de développer la pensée : au départ, l'accent est mis sur l'examen attentif des données initiales pour, progressivement, entrer dans une analyse plus dense pouvant traduire plus fidèlement la complexité de la réalité étudiée. Suivant une logique évolutive qui va de la production de notes analytiques et de constats jusqu'à la mise en évidence de récurrences significatives, le texte final produit rend compte d'une proposition cohérente d'interprétation et de compréhension du phénomène étudié (ici, la pratique évaluative d'un enseignant, notamment dans ses prises de décision) et permet de répondre ainsi à notre question de recherche.

4. Résultats

En lien avec notre question de recherche, les figures 2 et 3 ci-après présentent l'épreuve sommative créée par Bastyan. Chaque figure concerne un des deux niveaux réunis dans sa classe. Par la suite, nous décrivons en détail sa démarche évaluative, pour rendre explicites ses décisions et analyser la cohérence de la démarche.

Français 10VG Niveau 1
Lecture suivie – My life Matters
Test significatif de compréhension écrite
2 périodes

Situation de départ
Tu es Marvin le personnage principal du roman
 Tu prépares une marche en mémoire de ton frère et de toutes les victimes des violences policières. Une radio locale te propose d'enregistrer un podcast dans lequel tu dois présenter les derniers événements qui t'ont conduit à contacter Albert Sharp et organiser cette marche. Tu parles également de ton entourage et de certains faits que tu estimes être racistes dans ton quotidien.

Français 10VG Niveau 1
Lecture suivie – My life Matters
Test significatif de compréhension écrite
Critères d'évaluation

PER - L1 31 – Lire et analyser des textes de genres différents et en dégager les multiples sens...
Les critères suivants ont été élaborés à partir des différents objectifs d'apprentissage que nous avons travaillés en classe.
Pour réussir ton évaluation, tu dois satisfaire les critères suivants :

Critères obligatoires à satisfaire pour obtenir la note de 4

1. L'élève a choisi des informations essentielles du récit pour expliquer un problème
2. L'élève explique les sentiments du personnage principal
3. L'élève respecte la chronologie des événements.

Critères secondaires pour viser le 6

- a) L'élève justifie les sentiments en se basant sur le récit
- b) L'élève reformule les événements de manière correcte
- c) L'élève présente les personnages secondaires
- d) L'élève identifie dans le récit des faits ou des comportements qui relèvent du racisme

Consignes de travail
Tu dois produire un enregistrement audio à l'aide de l'ipad ou de ton téléphone. Ton enregistrement devrait durer entre 4 et 6 mn.

Dans ton audio, tu dois faire figurer les éléments suivants :

1. Les événements qui ont fait basculer ta vie ces dernières semaines.
2. Tu expliques ce qui s'est passé pour que les auditeurs comprennent bien ta vie actuelle.
3. Tu fais attention à respecter la chronologie du récit.
4. Tu expliques dans quel état tu es au moment de préparer cette marche en hommage à ton frère.
5. Tu justifies tes sentiments en citant des petits passages du livre (tu peux reformuler).
6. Tu présentes les personnes qui te sont chers dans ton entourage et tu expliques pourquoi.
7. Tu présentes des faits qui te sont arrivés et que tu penses être liés au racisme (attention il s'agit de faits arrivés à Marvin).

Pour effectuer le TS, tu disposes de 2 périodes.
 35-40mn en classe pour préparer ton audio sur feuille. Note des mots clés. Prends le temps de bien lire les consignes et les points importants.
 Enfin la seconde période tu pourras t'enregistrer.
Ton travail doit m'être envoyé à la fin de la seconde période au plus tard.

Travail non effectué	0	1 critère	2 critères	3 critères	3+1	3+2	3+3	3+4
1	2.5	3	3.5	4	4.5	5	5.5	6

Figure 2. Épreuve sommative niveau 1

Français 10VG Niveau 2
Lecture suivie – My life Matters
Test significatif de compréhension écrite
2 périodes

Situation de départ
Tu es Marvin le personnage principal du roman

Tu prépares une marche en mémoire de ton frère et de toutes les victimes du racisme. Une radio locale te propose d'enregistrer un podcast dans lequel tu dois présenter les derniers événements qui t'ont conduit à contacter Albert Sharp et organiser cette marche. Tu parles également de ton entourage et de certains faits que tu estimes être racistes dans ton quotidien.

Consignes de travail

Tu dois produire un enregistrement audio à l'aide de l'ipad ou de ton téléphone Ton enregistrement devrait durer entre 4 et 6 mn.

Dans ton audio, tu dois faire figurer les éléments suivants :

1. Les événements qui ont fait basculer ta vie ces dernières semaines.
2. Tu expliques ce qui s'est passé pour que les auditeurs comprennent bien ta vie actuelle.
3. Tu fais attention à respecter la chronologie du récit.
4. Tu expliques dans quel état tu es au moment de préparer cette marche en hommage à ton frère.
5. Tu justifies tes sentiments en citant des petits passages du livres (tu peux reformuler).
6. Tu présentes les personnes qui te sont chers dans ton entourage et tu expliques pourquoi.
7. Tu justifies leur sentiment en te basant sur le texte.
8. Tu présentes des faits qui te sont arrivés et que tu penses être liés au racisme (attention il s'agit de faits arrivés à Marvin).
9. Tu justifies pourquoi tu penses que cela est raciste.

Pour effectuer le TS, tu disposes de 2 périodes.
 35-40mn en classe pour préparer ton audio sur feuille de brouillon . Note des mots clés. Prends le temps de bien lire les consignes et les points importants.
 Enfin la seconde période tu pourras t'enregistrer.
Ton travail doit m'être envoyé à la fin de la seconde période au plus tard.

Français 10VG Niveau 2
Lecture suivie – My life Matters
Test significatif de compréhension écrite
Critères d'évaluation

PER - L1 31 – Lire et analyser des textes de genres différents et en dégager les multiples sens...

Les critères suivants ont été élaborés à partir des différents objectifs d'apprentissage que nous avons travaillés en classe.

Pour réussir ton évaluation, tu dois satisfaire les critères suivants :

Critères obligatoires à satisfaire pour obtenir la note de 4

1. L'élève a choisi des informations essentielles du récit pour expliquer un problème.
2. L'élève explique les sentiments du personnage principal.
3. L'élève justifie les sentiments en se basant sur le récit.
4. L'élève reformule les événements de manière correcte.
5. L'élève respecte la chronologie des événements.

Critères secondaires pour viser le 6

- a) L'élève présente les personnages secondaires.
- b) L'élève justifie ses propos sur les personnages secondaires en citant le texte.
- c) L'élève identifie dans le récit des faits ou des comportements qui relèvent du racisme.
- d) L'élève justifie son propos sur la question du racisme.

Travail non effectuée	0 critère	1 critère	2 critères	3 critères	4 critères	5 critères	5+1	5+2	5+3	5+4
1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6

Figure 3. Épreuve sommative niveau 2

4.1. Choix de la thématique et définition des objectifs

La démarche de l'enseignant a été conçue autour de la lecture d'une œuvre littéraire complète : *My life Matters* de Jay Coles (2018). Cet ouvrage a été choisi conjointement par l'enseignant et les élèves, très sensibles aux thématiques liées au racisme et à la communauté afro-américaine.

Les objectifs principaux de la séquence, basés sur le Plan d'études Romand (PER), portaient sur la lecture complète d'une œuvre³. L'enseignant a privilégié certaines *progressions d'apprentissage*⁴ issues du PER, telles que :

- l'identification des caractéristiques du genre de texte ;
- la réalisation d'activités résumantes ;
- l'identification des personnages, de leurs relations et de leurs rôles ;
- l'analyse du portrait moral des personnages ;
- le repérage du schéma narratif ;
- la mise en relation d'éléments du récit avec son vécu.

Si certaines de ses progressions sont de l'ordre de l'identification, l'enseignant a choisi de mettre l'accent sur des habiletés plus complexes, telles que la reformulation des événements d'un récit (activités résumantes) et l'analyse des sensations et des émotions des personnages (portrait moral). Par ailleurs, dans sa classe, Bastyan a dû composer avec un rapport à la lecture très hétérogène : par exemple, des élèves expliquaient n'avoir jamais réussi à lire un livre dans son intégralité, considérant que c'était une activité trop difficile. Face à ce contexte, l'enseignant a orienté les apprentissages travaillés dans la séquence autour de la compréhension globale de l'œuvre et de la mise en relation du contenu avec le vécu des élèves.

4.2. Construction de la tâche évaluative

Observant que, pour nombre d'élèves, organiser des idées à l'écrit demeurait très difficile, Bastyan a décidé (en partie, grâce aux discussions menées dans la communauté d'apprentissage professionnelle) de recourir à l'oral pour certifier la compréhension de l'œuvre. La tâche que les élèves avaient à effectuer pour l'épreuve sommative (figures 1 et 2) consistait à la réalisation d'un enregistrement audio, basé sur une situation contextualisée amenant les élèves à développer différents éléments de compréhension et d'interprétation. Cette tâche était similaire pour les deux niveaux, mais avec des exigences différentes, explicitées dans les critères définis pour chaque niveau. Ici, les choix de l'enseignant se sont fondés sur le fait que les apprentissages et la lecture de l'œuvre avaient été réalisés en commun, avec le but de mettre tous les élèves de la classe face à un défi exigeant.

4.3. Système de pondération et construction de la note

Pour proposer une pondération reflétant des exigences adaptées à chaque niveau, Bastyan a construit deux grilles de critères pour l'épreuve sommative. Tous les critères portaient sur

³ https://www.plandetudes.ch/web/guest/L1_31/

⁴ Éléments qui décrivent les connaissances et compétences spécifiques à travailler en classe, précisant ainsi les objectifs d'apprentissage plus généraux.

des objectifs finaux similaires et des éléments travaillés pendant la séquence avec toute la classe. Plus précisément, pour les élèves de niveau 1 (exigences de base), sept critères ont été définis, dont trois considérés comme *obligatoires*, c'est-à-dire, représentant les attentes minimales pour estimer que l'élève maîtrisait l'objectif évalué et, par conséquent, pour atteindre le seuil de réussite (donné par la note de 4). Les élèves de niveau 2 (exigences supérieures), quant à eux, devaient valider cinq critères obligatoires : les attentes y étaient alors plus élevées et les élèves devaient démontrer une plus grande maîtrise des contenus travaillés pour réussir l'épreuve.

Le nombre de critères *secondaires* ou de perfectionnement permettant de qualifier l'excellence et, donc, d'obtenir des notes plus élevées (entre 4,5 et 6), était le même dans les deux cas : quatre critères. Or, deux critères obligatoires pour les élèves de niveau 2 (en lien avec la reformulation d'événements du récit et la justification des sentiments des personnages) étaient considérés comme des critères de perfectionnement pour les élèves de niveau 1. Aussi, pour les élèves de niveau 2, deux critères (justification des propos sur les personnages et sur la question du racisme) s'ajoutent aux exigences de perfectionnement, alors qu'ils étaient complètement absents de la grille d'évaluation du niveau 1. Comme illustré ci-après (§ 4.4), c'est à partir de ses observations pendant la séquence que Bastyan a déterminé cette hiérarchisation des contenus évalués, afin de traduire les exigences pour les élèves de chaque niveau à cet instant précis de leurs apprentissages.

Il convient de préciser que, depuis le début de l'année scolaire, les élèves étaient habitués aux évaluations critériées qui suivent la logique de pondération présentée ici puisque Bastyan, qui leur enseignait également l'histoire et la géographie, utilisait cette manière de pondérer pour toutes les évaluations sommatives proposées. L'enseignant avait également l'habitude de discuter en collectif les critères d'évaluation pour s'assurer de leur compréhension par les élèves.

4.4. Rôle de l'évaluation formative

Dans ce travail de hiérarchisation des contenus, visant à pondérer en différenciant les exigences pour chaque niveau, des informations récoltées par l'enseignant à travers des démarches formatives ont joué un rôle central. C'est le cas d'une démarche où les élèves devaient produire des enregistrements audios (cf. figure 4), modalité choisie par Bastyan pour permettre aux élèves de se familiariser avec le type de tâche prévu pour le travail final. Mais, surtout, des critères d'évaluation qui ont été repris ensuite dans l'épreuve sommative, apparaissaient déjà formalisés ici. En effet, Bastyan a élaboré les deux grilles de l'épreuve finale à partir des informations récoltées dans une visée formative. Par exemple, l'enseignant a repéré que de nombreux élèves de niveau 1 éprouvaient encore des difficultés à respecter la chronologie des événements. Après réflexion, il a déterminé que ce dernier élément était très important pour la compréhension d'un récit, justifiant qu'il devienne un critère fondamental.

My life Matters.

Objectifs : Reformuler par oral des événements d'un récit
Sélectionner des informations pertinentes d'un récit
Reformuler les sentiments d'un personnage

Critère de réussite de la tâche

1. **L'élève a choisi des informations essentielles du récit pour expliquer un problème**
2. **L'élève explique les sentiments du personnage**
3. **L'élève justifie les sentiments en se basant sur le récit**
4. **L'élève reformule les événements de manière correcte**

Tâche d'expression compréhension

Suite à l'événement dans l'épicerie, Marvin souhaite contacter sa tante Nicola, policière, et lui expliquer la tournure des événements.
Il choisit de lui envoyer un message vocal dans lequel il lui explique tout ce qui se passe en ce moment dans sa vie au sujet de Tyler notamment et son écoeurément face à ce qu'il vient de vivre dans l'épicerie.
Enfin, il tente de convaincre sa tante d'intervenir.

Tu te mets à la place de Marvin et décide donc d'envoyer un message audio à Nicola.
Ton message vocal doit comporter les points suivants :

Développe les points suivants :

- **Explique la situation à laquelle il fait face à ce moment-là.**
- **Explique ce qui s'est passé à l'épicerie et les sentiments qu'il peut avoir.**
- **Tu demandes de l'aide.**

Comme tu es à la place de Marvin, tu dois parler en utilisant la première personne du singulier (l'audio doit durer moins de 5 mn).

Figure 4. Démarche formative formelle

Les critères ont été présentés et discutés avec les élèves avant la réalisation de la tâche, en veillant à expliciter leurs liens avec les objectifs d'apprentissage. Ayant clarifié au préalable les consignes et l'organisation de l'activité, l'enseignant a donné du temps en classe aux élèves pour préparer la structure de leur message vocal, puis pour s'enregistrer. L'utilisation des critères a permis de fournir des informations explicites, liées aux apprentissages visés, sur les éléments qui apparaissaient acquis et ceux qui restaient à travailler. Par exemple, l'évaluation formative a mis en évidence que certains élèves ne maîtrisaient pas encore le critère portant sur le choix des informations essentielles du récit pour expliquer un problème. C'est le cas, comme l'illustre la grille ci-après (figure 5), d'une élève, Irina⁵, qui

⁵ Prénom d'emprunt.

citait des événements sans les développer.

My life Matters

Ton message est intéressant mais soit plus précise lorsque tu parles des événements auxquels tu fais référence. Pense que la personne à qui tu les expliques ne connaît peut-être pas tout le déroulement. Illustre ton propos en expliquant clairement les faits.

Objectifs : Reformuler par oral des événements d'un récit
Sélectionner des informations pertinentes d'un récit
Reformuler les sentiments d'un personnage

Critère de réussite de la tâche

- 1. L'élève a choisi des informations essentielles du récit pour expliquer un problème → Les événements sont pertinents mais pense à mieux les expliquer : qui est ce Johntae auquel tu fais référence ?**
- 2. L'élève explique les sentiments du personnage → Les sentiments sont clairement explicités, c'est bien.**
- 3. L'élève justifie les sentiments en se basant sur le récit → Justifie les sentiments, en expliquant comment tu t'es sentie au moment où le policier était sur toi.**
- 4. L'élève reformule les événements de manière correcte → Ok mais un peu rapide. Sois plus précise. Imagine-toi que ton interlocutrice ne connaît pas les événements auxquels tu fais référence.**

Figure 5. Exemple de feedback lors de la démarche formative

4.5. Attribution des notes et feedback dans l'épreuve sommative

La note attribuée par Bastyan à chaque production d'élève reflète l'atteinte des critères définis en fonction du niveau. Dans ce cadre, le recours aux critères vise à soutenir les apprentissages des élèves qui peuvent s'appuyer sur les commentaires transmis pour continuer à travailler lors des prochaines séquences, sachant que la compréhension de textes écrits est une compétence abordée jusqu'à la fin de la scolarité obligatoire. La grille ci-après (figure 5) d'Irina (qui obtient la note de 5,5) illustre sa progression notamment sur l'un des critères fondamentaux, à savoir le choix des informations essentielles du récit pour expliquer un problème, tout en pointant le travail qui reste à faire concernant la reformulation des événements du récit.

Français 10VG Niveau 1									
Lecture suivie – My life Matters									
Test significatif de compréhension écrite									
Critères d'évaluation									
Lors de l'EF précédente, tu avais encore de la difficulté pour choisir et présenter de manière étayée les événements en lien avec le problème. Cette fois-ci, tu as bien pensé à te décentrer et à imaginer que ton interlocuteur ne connaissait pas les événements auxquels tu fais référence pour les clarifier.									
PER - L1 31 – Lire et analyser des textes de genres différents et en dégager les multiples sens...									
Les critères suivants ont été élaborés à partir des différents objectifs d'apprentissage que nous avons travaillés en classe.									
Pour réussir ton évaluation, tu dois satisfaire les critères suivants :									
Critères obligatoires à satisfaire pour obtenir la note de 4									
1. L'élève a choisi des informations essentielles du récit pour expliquer un problème → ok									
2. L'élève explique les sentiments du personnage principal → ok									
3. L'élève respecte la chronologie des événements → ok									
Critères secondaires pour viser le 6									
a) L'élève justifie les sentiments en se basant sur le récit → oui, Justifié de manière cohérente									
b) L'élève reformule les événements de manière correcte → attention ! Le procès n'a pas encore eu lieu. Il y a juste une mise en accusation avec la caution									
c) L'élève présente les personnages secondaires → ok									
d) L'élève identifie dans le récit des faits ou des comportements qui relèvent du racisme → oui, très bien									
	Travail non effectué	0	1 critère	2 critères	3 critères	3+1	3+2	3+3	3+4
Note	1	2.5	3	3.5	4	4.5	5	5.5	6

Figure 6. Exemple de grille d'évaluation sommative avec feedback

5. Discussion

Toujours en lien avec notre question de recherche, nous mettons en discussion ci-après les enjeux et conditions pour le développement de pratiques d'évaluation sommative notée qui visent à soutenir les apprentissages des élèves dans une classe avec des niveaux d'exigence certificatifs différents.

5.1. Comment définir les attentes en matière d'apprentissage ?

Bastyan a construit et mis en œuvre sa séquence en suivant une logique de planification à rebours. Ce modèle, étroitement lié à la différenciation pédagogique (Tomlinson & McTighe, 2010), lui a permis de déterminer les objectifs et les contenus qu'il considérait essentiels à ce moment de l'année pour développer la compétence de la compréhension de textes écrits, parmi tous ceux prescrits par le plan d'études. Ceci représente un élément fort de la pratique évaluative de Bastyan : en effet, les enseignants, même les plus expérimentés, considèrent souvent qu'ils doivent aborder la totalité du programme de formation⁶ (Reeves, 2002) et, par conséquent, que l'évaluation sommative doit aussi aborder tous les contenus travaillés en classe (Pasquini, 2021).

⁶ C'est le cas aussi dans notre contexte, même si les prescriptions institutionnelles prônent plutôt une sélection des objectifs à aborder.

Le choix des objectifs à travailler, puis à évaluer, implique ainsi une connaissance, voire une analyse de l'objet de savoir en jeu, permettant d'en identifier les apprentissages clés à maîtriser. C'est sur cette base que l'enseignant a déterminé les critères de réussite qui définissent la maîtrise de l'objectif visé. La manière dont Bastyan a approché ce travail de définition et de formalisation des attentes fait écho à des constats de la littérature scientifique qui insistent sur l'effet positif que les critères ont sur les apprentissages des élèves, notamment lorsqu'ils sont définis suffisamment tôt, puis communiqués de manière explicite à la classe (Balan & Jönsson, 2018 ; McMillan & Hearn, 2008 ; Jönsson, 2022). En effet, les objectifs de la séquence ont été présentés aux élèves en début de séquence et les critères définis par l'enseignant ont été introduits de manière progressive et flexible, notamment à travers des démarches d'évaluation formelle et informelle (Mottier Lopez, 2015 ; Ruiz Primo, 2017). Dans ce cadre, l'enseignant a travaillé en collectif, sur la base d'exemples, à la modélisation et à l'usage de critères, afin que les élèves puissent les comprendre et se les approprier (Andrade & Valtcheva, 2009 ; Ross, 2006). La grille finale des critères a été distribuée et discutée avec les élèves deux semaines avant la passation de l'évaluation sommative, pour leur permettre de comprendre la pondération de l'épreuve et les nuances entre les deux niveaux.

Soulignons encore que le travail de définition des critères d'évaluation n'est pas une tâche facile et que les enseignants manquent parfois de formation pour ce faire (Moss, 2013 ; Mottier Lopez & Pasquini, 2017 ; Pasquini, 2020 ; Pasquini & Morales Villabona, 2022). Si chaque enseignant connaît implicitement ses critères, toute la difficulté réside dans leur explicitation, pour soi et pour les autres, afin qu'ils deviennent des outils dans la construction des apprentissages (Jönsson, 2022). La participation de Bastyan à une communauté d'apprentissage professionnelle, qui fait suite à d'autres expériences en formation initiale et continue, a certainement contribué au développement de compétences spécifiques en matière d'évaluation lui permettant de réaliser ce travail. En ce sens, l'appropriation par l'enseignant de modèles théoriques (comme celui de la planification à rebours) pouvant être mobilisés en situation nous semble une condition importante – mais pas unique – pour le développement de pratiques d'évaluation sommative différenciées pouvant se mettre au service des apprentissages des élèves.

5.2. Comment différencier au travers des critères d'évaluation ?

Bastyan a mené une réflexion profonde autour de la pondération qui l'a conduit à hiérarchiser les critères définis et, en fonction de leur importance ainsi que des informations récoltées en classe, à construire une évaluation sommative qui tient compte de l'hétérogénéité de la classe (les deux niveaux d'exigence distincts). Ce travail de différenciation possible via des critères d'évaluation s'avère être une piste novatrice concernant la différenciation dans l'évaluation sommative (Baribeau et al., 2022). En effet, le peu de travaux qui existent sur la question insistent essentiellement sur le rôle des tâches d'évaluation, en faisant état d'expériences dans lesquelles les élèves peuvent choisir les tâches à réaliser dans l'épreuve sommative (Dubois, 2016 ; Varsavsky & Rayner, 2013). Bien que cette piste s'avère intéressante du point de vue pédagogique, elle n'est pas facilement envisageable dans tous les contextes. Dans le nôtre, au regard des prescriptions sur l'évaluation sommative, ce type de pratique pourrait générer potentiellement des tensions au sein de la classe. C'est la raison pour laquelle Bastyan a décidé de ne pas proposer deux tâches totalement différentes, pour éviter un sentiment d'inégalité de traitement entre les élèves. Dans cette perspective, il a pensé la différenciation en lien avec la définition des critères, en conservant une même tâche, exigeante pour tous les élèves,

tout en distinguant les attentes pour tenir compte de l'hétérogénéité de la classe.

Rappelons ici, en nous appuyant sur les propos d'Allal (2017), que la démarche de Bastyan répond à une contrainte importante : celle d'une *différenciation structurelle* imposée, donnée par la présence dans sa classe d'élèves enclassés dans deux niveaux. Très vite, cette condition devient l'occasion de mettre en œuvre des stratégies de *différenciation pédagogique* pour tenir compte de l'hétérogénéité dans une perspective de soutien à l'apprentissage et, si possible, pour réduire le risque de stigmatisation des élèves. L'usage des critères comme outil de pondération apparaît comme un levier fort, permettant de soutenir des pratiques d'évaluation sommative différenciée, via la hiérarchisation des apprentissages travaillés. Celle-ci se fait alors au regard de l'objet disciplinaire en jeu (ici, la compréhension de textes écrits) ainsi que des contraintes du contexte. Par ailleurs, le recours aux critères permet ici de mettre en synergie les démarches d'évaluation formative et sommative (Harlen, 2005), rendant ainsi visible pour les élèves que les attentes restent les mêmes tout au long de la séquence quel que soit le niveau.

Néanmoins, il est nécessaire de considérer les possibles limites de cette différenciation par la pondération. Une première concerne la possibilité qu'un élève de niveau 1 puisse réussir largement son épreuve (comme Irina, qui a validé tous les critères de l'épreuve sauf un), se rapprochant ainsi davantage des exigences destinées au niveau 2. Dans ce type de cas, les tensions, voire les ruptures entre les formes structurelle et pédagogique de la différenciation s'accroissent car le message pour l'élève n'est finalement pas si clair. D'un côté, sa réussite peut être motivante, montrant un développement important de ses compétences de compréhension. De l'autre, et en fonction du système scolaire, un tel résultat ne peut potentiellement rien changer pour l'élève dans l'immédiat car il restera enclassé dans le niveau 1. Dès lors, à court terme, la situation peut s'avérer frustrante pour l'élève (et pour l'enseignant). Seulement à long terme et en fonction de l'ensemble des résultats de l'élève⁷, il serait possible pour l'enseignant d'éventuellement soutenir une réorientation, un changement de niveau. Dans ce cas de figure, la fonction pronostique de l'évaluation (Mottier Lopez, 2015) reste limitée, malgré le caractère hautement informatif de la démarche d'évaluation proposée.

Une deuxième limite est donnée par la présence de deux niveaux dans la classe de Bastyan qui, en l'occurrence, apparaît à la fois comme une contrainte et comme une possibilité qui ouvre la porte à une évaluation sommative différenciée. Si cette condition n'était pas présente, proposer deux grilles de certification distinctes pourrait difficilement être envisagée, malgré le fait que l'hétérogénéité des élèves reste une réalité de plus en plus présente dans les classes dites « ordinaires » (Galand, 2009 ; Zakhartchouk, 2016). Autrement dit, différencier dans l'évaluation sommative alors que l'hétérogénéité des élèves n'est pas formellement reconnue par l'institution semble plus difficile que lorsque la différenciation structurelle est présente.

5.3. Comment différencier (aussi) les attentes dans une tâche complexe ?

En s'appuyant sur les modèles de la planification à rebours (Tomlinson & McTighe, 2010) et de l'alignement curriculaire élargi (Pasquini, 2019), la réflexion de Bastyan s'est prolongée

⁷ Comme précisé dans les directives officielles.

dans le choix des tâches nécessaires pour récolter des informations pertinentes sur les apprentissages des élèves, balisées par les attentes définies, en cours d'enseignement et jusqu'à l'évaluation sommative. Le choix d'utiliser une tâche de production orale pour certifier la compréhension d'un texte écrit a notamment été motivé par un questionnement sur des manières d'évaluer permettant de vérifier des apprentissages complexes inhérents à la compréhension d'un ouvrage et, par la même occasion, de contourner les limitations associées à l'usage des tâches plus traditionnelles et simples (des items – Dubus, 2006). Nous observons ici aussi une perspective de différenciation dans l'évaluation sommative, par la prise en compte des obstacles pouvant se présenter aux élèves et l'anticipation de moyens pour les surmonter.

Le fait de proposer à tous les élèves une tâche complexe, adaptée aux deux niveaux d'exigence, constitue un message puissant qui repose sur le principe d'éducabilité. Il s'agit, là encore, d'un choix fort assumé par Bastyan, au regard des particularités de sa classe, des besoins perçus en matière d'apprentissage (y compris en termes de motivation) et de ses intentions pédagogiques. Toutefois, il faut admettre qu'aucune tâche (ou type de tâche) ne pourrait convenir sans exception à tous les élèves (Moon, 2005). Les observations réalisées par l'enseignant pendant sa séquence, y compris lors de l'épreuve sommative, montrent que les élèves les plus en difficulté à l'écrit ont été soulagés de pouvoir se concentrer sur la verbalisation des contenus de leur production. L'analyse des résultats et des productions orales comparée à des travaux précédents d'élèves (comme illustré par les figures 5 et 6) a tendance à corroborer ce propos, à l'exception de deux élèves de niveau 2 qui, peu à l'aise avec cet exercice, y ont vu une source de stress supplémentaire.

5.4. Noter et différencier : quelles tensions ?

La démarche mise en œuvre par Bastyan témoigne de la recherche d'une cohérence maximale dans son épreuve et cela jusqu'aux pratiques de notation qui en résultent et qu'il souhaite constructives – c'est-à-dire, porteuses d'informations utiles et pertinentes pour les élèves et pour lui-même (Hadjji, 2016 ; Pasquini, 2021). Des contraintes institutionnelles se révèlent toutefois être des obstacles potentiels à la mise en œuvre de telles pratiques. C'est le cas de l'existence d'une échelle imposée par les directives cantonales pour toutes les épreuves sommatives (de 1 à 6, avec des demi-points) qui comporte dès lors 11 échelons (11 notes possibles). De ce fait, les pratiques de pondération à l'aide de critères font face à un obstacle technique : au-delà de la cohérence des critères, c'est leur nombre qui joue un rôle, afin de permettre de qualifier chaque note. En effet, bien que guidé par la logique de l'alignement curriculaire élargi (Pasquini, 2019), l'enseignant a réfléchi avant tout en termes d'apprentissages essentiels à maîtriser et, sur cette base, il a défini le seuil de réussite (donné par la note de 4) qui correspond à la validation des critères obligatoires définis. Une conséquence de cette pratique est que l'échelle de notation ne va pas au-delà de la note 2,5 pour les élèves de niveau 1. Pour Bastyan, ce choix s'est avéré cohérent d'un point de vue pédagogique : quel est l'intérêt de qualifier l'insuffisance avec une note, alors que les critères permettent de pointer de manière plus pertinente les contenus que l'élève doit encore maîtriser ? Or, ce type de décision peut s'avérer délicat à assumer face à des collègues qui privilégient des pratiques plus traditionnelles et provoquer des dilemmes importants (Pasquini & DeLuca, 2021).

D'ailleurs, pour les enseignants, la question peut se poser différemment en fonction de l'impact de la note sur la suite du parcours des élèves, en matière de promotion ou, éventuellement, de (ré)orientation. Ainsi, par exemple, la prise de décisions sur les critères

qui définissent les notes entre 3 et 4 peut s'avérer particulièrement complexe et délicate. L'expérience et l'expertise professionnelles (ou, plus largement, le développement d'un jugement professionnel en évaluation – Mottier Lopez & Tessaro, 2016) apparaissent alors primordiales pour affronter ce type de situations et pour éviter le risque de vouloir à tout prix compléter l'échelle dans son entier, même si cela signifie créer artificiellement des critères, source potentielle d'un désalignement curriculaire (Airasian & Miranda, 2002).

6. Conclusion et perspectives

Nos analyses soulignent la complexité des pratiques d'évaluation sommative notées dans une classe hétérogène lorsqu'elles visent également à soutenir les apprentissages des élèves. Nous observons que leur mise en œuvre peut se réaliser au travers d'une réflexion première sur les apprentissages poursuivis et les attentes pressenties. Alors que des travaux abordent la question de l'hétérogénéité en étudiant des variations dans les tâches d'évaluation, l'attention et le choix de l'enseignant ont porté ici sur les critères d'évaluation – sans nier l'importance de la tâche.

Malgré des tensions liées à une pondération différenciée à partir des critères (au regard des contraintes contextuelles sur la communication des résultats, par exemple), ces derniers permettent à l'enseignant de travailler les mêmes contenus et objectifs avec tous les élèves et, en même temps, de distinguer les seuils de réussite pour chaque niveau. Dans ce cadre, l'appropriation de contenus théoriques pouvant être mobilisés en situation (planification à rebours, alignement curriculaire élargi) apparaît comme une condition nécessaire pour : 1) soutenir la construction de démarches évaluatives cohérentes et à haute valeur informative, et 2) faire face aux enjeux propres à une évaluation sommative notée qui tient compte des spécificités et contraintes du contexte, dont l'hétérogénéité des élèves.

Notre étude de cas rend compte d'une articulation rendue nécessaire entre une différenciation structurelle imposée (distinction de niveaux) et l'émergence de stratégies de différenciation pédagogique (en l'occurrence, en matière d'évaluation) pour tenir compte de l'hétérogénéité dans une perspective de régulation de l'enseignement et de l'apprentissage. Or, des questions se posent quant à la possibilité de proposer une évaluation sommative dans d'autres contextes de classe hétérogènes. D'autres travaux pourraient dès lors s'intéresser à l'articulation entre démarches évaluatives et différenciation, notamment lorsque l'hétérogénéité dans les classes ne s'inscrit pas dans des structures spécifiques, reconnues institutionnellement.

Deux limites de notre travail sont enfin à signaler. Premièrement, l'étude de cas porte sur la pratique d'un enseignant en questionnement et dans une posture d'apprenant, au sein d'une communauté d'apprentissage professionnel, induisant possiblement un effet de désirabilité sociale (Leclerc et al., 2011). Deuxièmement, nous avons choisi de nous focaliser sur les décisions et actions de l'enseignant plus que sur des données relatives aux apprentissages des élèves. De ce fait et compte tenu du caractère exploratoire (et donc, non généralisable) de notre étude, d'autres projets de recherche pourraient examiner la transposition de nos résultats à d'autres contextes – en s'intéressant, par exemple, à d'autres disciplines ou objets, à d'autres degrés, aux pratiques d'enseignants ayant des rapports différents à l'évaluation en classe – et poser un regard plus approfondi à propos des apprentissages des élèves et leur soutien au travers des pratiques évaluatives examinées.

7. Références

- Airasian, P. W., & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory Into Practice*, 41(4), 249-254. https://doi.org/10.1207/s15430421tip4104_8
- Albarello, L. (2011). *Choisir l'étude de cas comme méthode de recherche*. De Boeck.
- Ali, H. I. H. (2015). Toward differentiated assessment in a public college in Oman. *English Language Teaching*, 8(12), 27-36. <https://doi.org/10.5539/elt.v8n12p27>
- Allal, L. (2008). Évaluation des apprentissages. Dans A. van Zanten (Dir.), *Dictionnaire de l'éducation* (pp. 311-314). Presses Universitaires de France.
- Allal, L. (2017). *Comment former les enseignants à l'évaluation au service de la différenciation pédagogique ?* Cnesco-Cnam. https://www.cnesco.fr/wp-content/uploads/2017/03/170316_17_Allal.pdf
- Allal, L., & Laveault, D. (2009). Assessment for Learning : évaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation*, 32(2), 99-106. <https://doi.org/10.7202/1024956ar>
- Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory Into Practice*, 41(4), 255-260. https://doi.org/10.1207/s15430421tip4104_9
- Anderson, L. W., & Krathwohl, D. R. (dir.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48(1), 12-19. <https://doi.org/10.1080/00405840802577544>
- Balan, A., & Jönsson, A. (2018). Increased explicitness of assessment criteria: Effects on student motivation and performance. *Frontiers in Education*, 3(81), 1-13. <https://doi.org/10.3389/educ.2018.00081>
- Baribeau, A., Laveault, D., Nizet, I., & Maunier, S. (2022). Les enjeux de l'évaluation scolaire au primaire et au secondaire. Dans S.-G. Chartrand, S. Dancause, K. Dufault, G. Lapierre, C. Lessard, & J. Teasdale (dir.), *Une autre école est possible et nécessaire* (pp. 292-307). Delbusso Editeur.
- Biggs, J. (2003). Aligning teaching for constructing learning. *Higher Education Academy*, 1(4), 1-4. http://egusdsecondaryed.pbworks.com/f/aligning_teaching_for_constructing_learning.pdf
- Black, H. D. (1986). Assessment for learning. Dans D. L. Nuttall (Dir.), *Assessing educational achievement* (pp. 7-18). Falmer.
- Braund, H., & Deluca, C. (2018). Elementary students as active agents in their learning: An empirical study of the connections between assessment practices and student metacognition. *Australian Journal of Educational Research*, 45, 65-85. <https://doi.org/10.1007/s13384-018-0265-z>
- Braxmeyer, N., Guillaume, J. C., & Levy, J. F. (2004). Les pratiques d'évaluation des enseignants au collège. *Note évaluation*, 13, 1-5.
- Brina, M., & Dechamboux, L. (2021). Usage de l'alignement curriculaire élargi en enseignement spécialisé : intérêts, ajustements et questionnements. *Revue suisse de pédagogie spécialisée*, 4, 17-22. <https://www.szh-csps.ch/r2021-12-02/>
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. University of Cambridge School of Education. http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf
- Brookhart, S. M. (2017). *How to use grading to improve learning*. ASCD.
- Brookhart, S. (2022). Criteria connect grading with other assessment practices. *La Revue LEEe*, 6. <https://doi.org/10.48325/rleee.006.01>

- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- Butera, F., Buchs, C., & Darnon, C. (2011). *L'évaluation, une menace ?* Presses Universitaires de France.
- Coles, J. (2018). *My life Matters*. Hachette Romans.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12* (1), 53-72. https://doi.org/10.1207/s15324818ame1201_4
- De Saint-André, M. D., Montésinos-Gelet, I., & Morin, M.-F. (2010). Avantages et limites des approches méthodologiques utilisées pour étudier les pratiques enseignantes. *Nouveaux cahiers de la recherche en éducation, 13*(2), 159-176. <https://doi.org/10.7202/1017288ar>
- Desgagné, S., & Bednarz, N. (2005). Médiation entre recherche et pratique en éducation : faire de la recherche « avec » plutôt que « sur » les praticiens. *Revue des sciences de l'éducation, 31*(2), 245-258. <https://doi.org/10.7202/012754ar>
- Dubois, B. (2016). La flexibilité d'une évaluation : une réponse à la prise en compte de la diversité des élèves ? *La nouvelle revue de l'adaptation et de la scolarisation, 74*, 79-87. <https://doi.org/10.3917/nras.074.0079>
- Dubus, A. (2006). *La notation des élèves. Comment utiliser la docimologie pour une évaluation raisonnée*. Armand Colin.
- 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>
- Fletcher, A. K. (2016). Exceeding expectations: Scaffolding agentic engagement through assessment as learning. *Educational Research, 58*(4), 400-419. <https://doi.org/10.1080/00131881.2016.1235909>
- Galand, B. (2009). Hétérogénéité des élèves et apprentissage : quelle place pour les pratiques d'enseignement ? *Les cahiers de recherche en éducation et formation, 71*. https://shs.hal.science/file/index/docid/561564/filename/cahier_71_galand_bis_1_.pdf
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. Routledge.
- Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership, 71*(1), 68. https://uknowledge.uky.edu/edp_facpub/22
- Guskey, T. R. (2022). Can grades be an effective form of feedback? *Phi Delta Kappan, 104*(3), 36-41. <https://kappanonline.org/grades-feedback-guskey/>
- 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 ?* (pp. 113-135). Presses Universitaires du Septentrion.
- Hadji, C. (2017). Savoir mettre en œuvre une évaluation constructive. Dans A. Bentolita (Dir.), *L'essentiel de la pédagogie* (pp. 229-250). Nathan.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16*(2), 207-223. <https://doi.org/10.1080/09585170500136093>
- Harris Stefanakis, E. (2010). *Differentiated assessment: How to assess the learning potential of every student*. Jossey-Bass.
- Hayward, L., & Spencer, E. 2010. The complexities of change: Formative assessment in Scotland. *The Curriculum Journal, 21*(2), 161-177.

- Jönsson, A. (2022). Young students' use of explicit grading criteria for self-regulation. *La Revue LEEe*, 6. <https://doi.org/10.48325/rleee.006.02>
- Jung, L. A., & Guskey, T. R. (2010). Grading exceptional learners. *Educational Leadership*, 16(5), 31-35. https://uknowledge.uky.edu/edp_facpub/4?utm_source=uknowledge.uky.edu%2Fedp_facpub%2F4&utm_medium=PDF&utm_campaign=PDFCoverPages
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), 263-268. <https://doi.org/10.1080/09695940903319646>
- Lang, M. L. (2019). Planning for differentiated instruction: Instructional leadership practices perceived by administrators and teachers in middle schools. *Educational Planning*, 26(2), 29-45. <https://files.eric.ed.gov/fulltext/EJ1217447.pdf>
- Laveault, D., & Allal, L. (2016). Implementing assessment for learning: Theoretical and practical issues. Dans D. Laveault et L. Allal (dir.), *Assessment for learning in the classroom: Meeting the challenge of implementation* (pp. 1-18). Springer. https://doi.org/10.1007/978-3-319-39211-0_1
- Leclerc, C., Bourassa, B., Picard, F., & Courcy, F. (2011). Du groupe focalisé à la recherche collaborative : avantages, défis et stratégies. *Recherches qualitatives*, 29(3), 145-167. <https://doi.org/10.7202/1085877ar>
- Loi sur l'enseignement obligatoire – LEO (2011, 7 juin). Le grand conseil du canton de Vaud. https://www.vd.ch/fileadmin/user_upload/themes/formation/sante_scolaire/fichiers_pdf/Loi_sur_l_enseignement_obligatoire__LEO_.pdf
- Lysaght, Z., Scully, D., Murchan, D., O'Leary, M., & Shiel, G. (2019). *Aligning assessment, learning and teaching in curricular reform and implementation*. National Council for Curriculum and Assessment. <http://www.tara.tcd.ie/handle/2262/91608>
- Lock, J., Becker, S., & Redmond, P. (2021). Teachers conceptualizing and developing assessment for skill development: Trialing a maker assessment framework. *Research Evaluation*, 30(4), 540-551. <https://doi.org/10.1093/reseval/rvab029>
- Lui, A. M., & Andrade, H. L. (2022). Inside the next black box: Examining students' responses to teacher feedback in a formative assessment context. *Frontiers in Education*, 7, article 751549. <https://doi.org/10.3389/feduc.2022.751548>
- Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessments. *Applied Measurement in Education*, 15(3), 249-267. https://doi.org/10.1207/S15324818AME1503_2
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49. <https://www.jstor.org/stable/42923742>
- McMillan, J. H., & Nash, S. (2000, 27 avril). Teacher classroom assessment and grading practices decision making. Annual meeting of the National Council on Measurement in Education, New Orleans, LA, USA. <https://files.eric.ed.gov/fulltext/ED447195.pdf>
- Moon, T. R. (2005). The role of assessment in differentiation. *Theory Into Practice*, 44(3), 226-233. https://doi.org/10.1207/s15430421tip4403_7
- Moss, C. M. (2013). Research on classroom summative assessment. Dans J. H. McMillan (dir.), *SAGE Handbook of Research on Classroom Assessment* (pp. 235-255). SAGE Publications. <https://dx.doi.org/10.4135/9781452218649.n14>
- Mottier Lopez, L. (2013). Pour une culture partagée de l'évaluation. Du retour à des pratiques

- d'évaluation traditionnelles vers la construction d'un nouveau rapport entre la recherche, le terrain et le politique. *Revue Éducation Canada*. <https://www.edcan.ca/articles/pour-une-culture-partagee-de-levaluation/?lang=fr>
- Mottier Lopez, L. (2015). *Évaluations formative et certificative des apprentissages. Enjeux pour l'enseignement*. De Boeck.
- Mottier Lopez, L., & Pasquini, R. (2017). Professional controversies between teachers about their summative assessment practices: A tool for building assessment capacity. *Assessment in Education: Principles, Policy & Practice*, 24(2), 228-249. <https://doi.org/10.1080/0969594X.2017.1293001>
- Mottier Lopez, L., & Tessaro, W. (dir.) (2016). *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages*. Peter Lang
- Mrazek, J., & Morales Villabona, F. (2022). Évaluation critériée et résolution de problèmes arithmétiques : un récit de pratique accompagné de pistes de réflexion sur l'évaluation constructive. *La Revue LEE*, 6. <https://doi.org/10.48325/rleee.006.05>
- Nordrum, L., Evans, K., & Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: Strategies for formative assessment. *Assessment & Evaluation in Higher Education*, 38(8), 919-940.
- Noman, M., & Kaur, A. (2014). Differentiated assessment: A new paradigm in assessment practices for diverse learners. *International Journal of Education and Applied Sciences*, 1(4), 167-174. <https://hdl.handle.net/20.500.12540/648>
- Organisation de Coopération et de Développement Économiques (2013). Les grandes espérances : Comment les notes et les politiques éducatives façonnent-elles les aspirations des élèves ? *PISA à la loupe*, 26. <https://doi.org/10.1787/5k49czl17mr5-fr>
- Ohlsen, M. T. (2007). Classroom assessment practices of secondary school members of NCTM. *American Secondary Education*, 36(1), 4-14. <https://www.jstor.org/stable/41406094>
- Paillé, P., & Mucchielli, A. (2013). *L'analyse qualitative en sciences humaines et sociales* (3^e éd.). Armand Colin.
- Pasquini, R. (2019). Élargir conceptuellement le mode de l'alignement curriculaire pour comprendre la cohérence des pratiques évaluatives sommatives notées des enseignants : enjeux et perspectives. *Mesure et évaluation en éducation*, 42(1), 63-92. <https://doi.org/10.7202/1066598ar>
- Pasquini, R. (2020). Et si la recherche collaborative participait au développement de pratiques évaluatives sommatives cohérentes ? *Mesure et évaluation en éducation*, 43(1), 33-66. <https://doi.org/10.7202/1076965ar>
- Pasquini, R. (2021). *Quand la note devient constructive. Évaluer pour certifier et soutenir les apprentissages*. Presses de l'Université Laval.
- Pasquini, R., & 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., & Morales Villabona, F. (2022). Comment appréhender une littérature en évaluation pour documenter et comprendre le développement des compétences des enseignants ? *Revue canadienne de l'éducation*, 45(2), 512-554. <https://doi.org/10.53967/cje-rce.v45i2.5171>
- Prud'homme, L., & Bergeron, G. (2012). Au-delà de la communication des contenus : une vision plus flexible de l'enseignement. *Prismes*, 17, 12-13.
- Reeves, D. B. (2002). *Making standards work: How to implement standards-based assessments in the classroom*,

- school, and district* (3^e éd.). Advanced Learning Press.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, 11(10), 1-13. <https://doi.org/10.7275/9wph-vv65>
- Roy, S. N. (2009). L'étude de cas. Dans B. Gauthier (Dir.), *Recherche sociale : de la problématique à la collecte des données* (5^e éd.) (pp. 199-225). Presses de l'Université du Québec.
- Ruiz Primo, M. A. (2017). Informal assessment. Dans M. A. Peters (dir.), *Encyclopedia of Educational Philosophy and Theory* (pp. 1142-1147). Springer. <https://www.daneshnamehicsa.ir/userfiles/files/1/Encyclopedia%20of%20Educational%20Phi.pdf>
- Tomlinson, C. A. (2005). Grading and Differentiation: Paradox or Good Practice? *Theory Into Practice*, 44(3), 262-269. https://doi.org/10.1207/s15430421tip4403_11
- Tomlinson, C. A., & McTighe, J. (2010). *Intégrer la différenciation pédagogique et la planification à rebours*. Chenelière Éducation.
- Van der Maren, J.-M. (1999). *La recherche appliquée en pédagogie : des modèles pour l'enseignement*. De Boeck.
- Varsavsky, C., & Rayner, G. (2013). Strategies that challenge: exploring the use of differentiated assessment to challenge high-achieving students in large enrolment undergraduate cohorts. *Assessment & Evaluation in Higher Education*, 38(7), 789-802. <https://doi.org/10.1080/02602938.2012.714739>
- Wakeman, S. Y., Thurlow, M., Reyes, E., & Kearns, J. (2022). Fair and equitable grading for ALL students in inclusive settings. *Inclusive Practices*, 1(4), 127-131. <https://doi.org/10.1177/27324745211055398>
- Wiliam, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Wiggins, G. (1998). *Educative assessment. Designing assessments to inform and improve student performance*. Jossey-Bass.
- Wiggins, G., & McTighe, J. M. (2005). *Understanding by design*. ASCD.
- Zakhartchouk, J.-M. (2016). *Enseigner en classes hétérogènes* (3^e éd.). ESF éditeur.