Comparative and International Education / Éducation comparée et internationale

Volume 49 | Issue 2

July 2021

Grading in a Dilemmatic Space : An Exploratory Cross-cultural Analysis of Mathematics and Language Secondary Teachers

Raphaël Pasquini
University of Teacher Education, State of Vaud, Lausanne, Switzerland
raphael.pasquini@hepl.ch
Christopher DeLuca
Queen's University,
cdeluca@queensu.ca

Recommended Citation
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

Grading in a Dilemmatic Space: An Exploratory Cross-cultural Analysis of Mathematics and Language Secondary Teachers

Noter dans un espace dilemmatique : une analyse exploratoire interculturelle d'enseignants de mathématiques et de langues du secondaire

Raphaël Pasquini, University of Teacher Education, State of Vaud, Lausanne, Switzerland
Christopher DeLuca, Queen's University

Abstract

Numerous studies have shown that grading is a complex process that involves negotiating technical, social, and ethical factors. While previous research has primarily focused on the reliability, composition, and validity of teachers' grades, few studies have examined grading practices across cultural contexts and teaching subjects. The purpose of this exploratory study was to analyze how *culture* and *teaching subject* influence teachers' grading dilemmas. Based on individual and group interviews with 11 Canadian and eight Swiss teachers, and using a "dilemmatic space" conceptual framework, this article inductively identified five dilemmatic spaces across cultural contexts and teaching subjects. The paper concludes with a discussion of the cross-cutting dimensions across these five dilemmatic spaces and articulates implications for future research and practice.

Résumé

De nombreuses recherches ont montré que la notation est un processus complexe qui implique la négociation de facteurs techniques, sociaux et éthiques. Alors que la plupart d'entre elles se sont principalement centrées sur la fiabilité, la composition et la validité des notes, peu d'études ont examiné les pratiques de notation en fonction des contextes culturels et des sujets d'enseignement. L'objectif de cette étude exploratoire était d'analyser comment la culture et la matière d'enseignement influencent les dilemmes de notation des enseignants. Sur la base d'entretiens individuels et collectifs avec 11 enseignants canadiens et huit enseignants suisses, et en utilisant un cadre conceptuel d'« espace dilemmatique », cet article a identifié de manière inductive cinq espaces dilemmatiques dans les contextes culturels et les matières d'enseignement. L'article se termine par une discussion des dimensions transversales de ces cinq espaces dilemmatiques et articule les implications pour la recherche et la pratique futures.

Keywords: grading, dilemmatic space, summative assessment, secondary teachers, qualitative research

Mots-clés : noter, espace dilemmatique, évaluation sommative, enseignants du secondaire, recherche qualitative

Introduction

Grading is a longstanding tradition across systems of education (Brookhart et al., 2016). Drawing on McMillan's (2019) definition, "grading practices refer to the ways teachers use information from assessments and other sources of information to determine and report student grades, whether on papers, unit tests, or semester reports" (p. 85). In this paper, we define grading as the scoring and evaluation of students' summative assignments or tests in relation to evaluation criteria and disciplinary curriculum expectations. In addition, grades are used to communicate student achievement of assessed curriculum expectations and differ from other non-summative forms of assessment such as diagnostic and formative assessment. One of the most consistent findings across grading literature is that teachers have variable grading practices, differently weighing evidence, student expectations, and grade consequences (Brookhart et al., 2016; McMillan & Nash, 2000; Randall & Engelhard, 2010). As teachers continually consider diverse pieces of evidence to render their grade decisions (Chen & Bonner, 2017; Cross & Frary, 1999), the summative assessment practices and the grading practices are

articulated, as Brookhart (2017) noted "we should not expect graded achievement and tested achievement to be the same thing" (p. 19).

For many teachers, grading represents an area of challenge in their practice (Alm & Colnerud, 2015) and can be well characterized as a dilemmatic space (Honig, 1996). In generating final grades for students, teachers need to often balance competing policies, tools, consequences, evidence, contextual and social conditions, subject traditions, and assessment theories when making grading decisions. Accordingly, teachers "find themselves in situations in which there is often no right way of acting, but only a way of 'acting for the best'" (Fransson & Grannäs, 2013, p. 5).

In this research, we draw on the notion of "dilemmatic space" (Fransson & Grannäs, 2013; Honig, 1996; Singh et al., 2015) to provide a novel theoretical framework for interpreting teachers' grading practices. While *dilemmas* are traditionally defined as problems without clear solutions (Fransson & Grannäs, 2013), adding *space* to the notion of *dilemma*, brings forward the relational environment and makes visible the contextual factors that shape dilemmas. Hence, by using *dilemmatic space* as a framework to interpret teachers' grading work, we view grading dilemmas not as "specific events or situations, but things that are ever-present" and as events that always tethered to a complex and incongruent context (Fransson & Grannäs, 2013, p. 7). Theorizing grading as dilemmatic space pushes beyond notions of context-dependency by characterizing grading as an always negotiable practice.

Walvoord and Johnson Anderson (2009, p. 2) argued that grading is a "complex context-dependent process that serves multiple roles." While previous research in grading has focused primarily on the reliability and composition of teachers' grades (e.g., Brookhart, 2013; McMillan, 2008) and the predictive and concurrent validity of grades in relation to other achievement measures (Thorsen & Cliffordson 2012), few studies have yet examined teachers' grading practices across cultural contexts and across teaching subjects. Moreover, little research has investigated how teachers across assessment cultural context and subjects negotiate grading dilemmas (Cheng et al., 2020). This research is important as an extension of previous studies which have highlighted that assessment cultures shape teacher practice. In particular, macro-cultures which includes education, social, and cultural contexts, traditions, and policies factor into teachers' beliefs, practices, and approaches as bounded by geographic space (Allal, 2016). In addition, micro-cultures—local dynamics at school, classroom, and person levels—shape how macro-factors are interpreted, negotiated, and implemented to ultimately yield diverse experiences of teaching, learning, and assessment (Mottier Lopez, 2016). Hence, culture, and specifically assessment culture, may shape teachers' grading practices and their engagement with grading dilemmas.

The purpose of this exploratory study was, first, to analyze how teachers experienced grading as a dilemmatic space. Second, to consider how *teaching subject* and *culture* factored into teachers' grading practices within a dilemmatic framework. In particular, we were interested in exploring teachers' experiences of grading as a dilemmatic space across teaching subjects—mathematics and language—and across cultural contexts— Canada and Switzerland. Through this study, we aimed to provide more nuanced evidence on how grading dilemmas are shared across these two contextual features. The primary research question guiding this study is: How is grading characterized as a dilemmatic space in two cultures and two teaching subjects? Secondary research questions are:

1. How do mathematics and language teachers in Canada and Switzerland experience grading as a dilemmatic space?
2. What are the dimensions of the dilemmatic space for mathematics and language teachers' grading work?

Research on Grading

Grades are used in educational systems around the world, at almost all levels and in almost all disciplines (OECD, 2012). For instance, in Canada, 86% of teachers report that they assess their students by administering a teacher-created test (CMEC, 2010). While grades can take various forms (within and between educational systems and levels of schooling), regardless of context, classroom assessment information is most often summarized and communicated through a grade (OECD, 2012). The widespread use and impact of grading has also led a large number of education systems to establish criteria to clarify the aims, requirements, and purposes of grading through system-level grading policies and sometimes through common grading tools (Guskey, 2013; Klapp, 2015).

A considerable amount of research shows that grading is a process that shapes learning progress and quality (Biggs, 2003; Shepard et al., 2018) and reflects social concerns (Allal, 2012; Alm & Colnerud, 2015). Grades typically serve three core purposes:
(a) provide feedback to students and parents about classroom learning and student achievement, (b) provide teachers with feedback for future instructional planning, and (c) certify that students have achieved a sufficient level of mastery in relation to curriculum standards and expectations (Carey & Carifio, 2012).

Despite the widespread and longstanding use of grades across educational systems, research has consistently shown a high degree of variability across teachers' grading practices. Brookhart et al. (2016) noted that "one hundred years of grading research have generally confirmed large variation among teachers in the validity and reliability of grades, both in the meaning of grades and in the accuracy of reporting" (p. 835). One key area that contributes to variability in teachers' grading practices is the assessment and inclusion of academic and nonacademic achievement factors in one grade (Brookhart et al., 2016; Isnawati & Saukah, 2017; OECD, 2012; Resh, 2009). Another source of variability is the use of normative grading systems in contexts where criterion-based practices would better serve the primary purpose of grading. Norm-referenced grading remains an active practice despite being criticized by educational researchers as not effectively supporting or reporting on student learning at classroom levels (Antibi, 2003; Crahay, 2007; Knight & Yorke, 2003; OECD, 2012; Shepard et al., 2018).

Fundamentally, contributing to the complexity of teachers' grading practices is the diversity of evidence used to make grading decisions. Teachers engage in various approaches to summative assessment and testing (design, administration, and scoring) (Dubus, 2006; Guskey, 2013) with some focused on simple or lower-level learning (Biggs, 2003; Shepard et al., 2018) and others measuring across ranges of cognitive complexity (Braxmeyer et al., 2005). Various assessment practices render a diversity of evidence for grading decisions, posing considerations for reliability and validity of classroom grades. The use of criteria and practices of moderated marking help to reduce variability and bolster consistency across teachers' grades (Brookhart et al., 2016; Docan, 2006; Sadler, 2009). However, often the weighting or interpretation of criteria are rarely stable across teachers or contexts (Andrade, 2005; Balan & Jönsson, 2019; Jönsson, 2014).

More precisely, research has shown that grading practices vary in relation to three contextual factors, which directly relate to our topic. First, there appears to be a difference between elementary and secondary teachers' grading practice (Link, 2018), with the former tending to use more open-ended and formative assessments (Braxmeyer et al., 2005) while secondary teachers are more reliant on paper-pencil assessments, however not exclusively (Zhang & Burry-Stock, 2003). There is also acknowledgment that teaching subject plays a role in grading practices (Cheng et al., 2018; Pasquini, 2019).

Second, mandatory policies matter in grading practices, shaping the educational culture in which grading occurs. While some research shows that teachers may neglect to implement

policies, especially if they don't consider the policy useful for supporting student learning (Tierney et al., 2011), policies point towards the cultural values, assessment and grading in a given educational context. This factor is further optimized in relation to teachers' assessment literacy and knowledge of assessment theory (DeLuca et al., 2018; Timperley & Parr, 2009), which can vary by educational system and culture (Cheng et al., 2018).

Third, grading practices differ considerably from one discipline to another, as teachers have a strong relationship to their subjects' epistemology, which often includes how a subject could be or should be assessed (Duncan & Noonan, 2007; McMillan, 2001; Meier et al., 2006; Pasquini, 2019; Resh, 2009). Many researchers have contrasted language and mathematics teachers' grading practices (e.g., Pasquini, 2019; Thorsen & Cliffordson, 2012), underlining that each discipline has its own grading style: mainly performance oriented for the former and more effort centered for the latter (Biberman-Shalev et al., 2011). Regarding mathematics, teachers tend to consider grading to be easier than in other subjects (Prøitz, 2013), and they seem to pay particular attention to their students' progress and achievement (Braxmeyer et al., 2005). However, holistic grading practices remains more challenging for mathematics teachers compared to language teachers, in part because of the characteristics of disciplinary subjects (Meier et al., 2006; Pasquini, 2019). Research on language teachers' grading practices suggests that they tend to be more open to continuous negotiation with students (Prøitz, 2013), use qualitative proficiency grading scales (Marzano & Heflebower, 2011), and consider student effort more in the grading process (Resh, 2009).

Given these elements, many researchers agree that grading is a complex practice that is potentially shaped by teaching context, subject, and education culture (Cheng et al., 2018; Guskey & Link, 2019; McMillan, 2001; Smith, 2003). Therefore, it is unsurprising that, on one hand, grading is described as a teacher's toughest task because it is unavoidable, complex, multifaceted, and laborious (Alm & Colnerud, 2015; Crahay, 2007; Ulvik et al., 2009), while on the other hand, it represents one of the most important levers teachers use to guide students' learning (Brookhart, 2017). Consequently, we assert that teachers are confronted with a dilemmatic context when grading students, which is common across education cultures and school contexts.

Conceptual Framework

Grading in this study is conceptualized through the emerging theory of dilemmatic space (Fransson & Grannäs, 2013; Turner, 2016). The concept of dilemmatic space is rooted in conceptualizing dilemmas within their sociocultural and political spaces. Dilemmas are understood as contradictions or challenges without an apparent solution and often result in balancing competing demands (Havnes & McDowell, 2008; Honig, 1996). As both Cuban (1992) and Perrenoud (2004) recognized, dilemmas are inherent in education in part due to the multiple value systems and ideologies facing educators and educational stakeholders, and the diverse moral scripts that govern decision-making. Research on dilemmas across contexts has identified different types of dilemmas including intrapersonal, interpersonal, and ethical. Intrapersonal dilemmas involve conflicts within oneself, whereas interpersonal dilemmas involve conflict in collaborations (i.e., between individuals) (Pareja Roblin & Margalef, 2013). Intrapersonal dilemmas among teachers most often involve curriculum matters, pedagogical decisions, assessment, and classroom management (Enyedy et al., 2006; Lampert, 1985; Lyons, 1990). Pareja Roblin and Margalef (2013) argued that the increasingly collaborative nature of teaching and teacher learning (e.g., professional learning communities) brings out intrapersonal dilemmas, but that both types of dilemmas can have positive influence on teacher development and change (Achinstein, 2002). Ethical dilemmas are more widely researched, both across sectors and in teaching, including in assessment contexts (Pope et al., , 2009). In education, as teaching is value-laden and as "honesty, integrity and professionalism are deemed

characteristics of ethical behaviour," ethical dilemmas are commonplace (Ehrich et al., 2011, p. 174). As recognized by Shapira-Lishchinsky (2011), there is a wide variety of dilemmas that fall under the banner of ethics given the multiple roles teachers are expected to fulfill. She also noted that the majority of studies in the field assert that teachers feel underprepared and lacking the tools to adequately address ethical dilemmas.

A key characteristic for understanding dilemmatic space is contextualizing these various types of dilemmas within their sociocultural, epistemological, and political space. As noted by Fransson and Grannäs (2013), *space* adds the relational aspect in which educators are forced to both consider the various actors and contexts as they work through dilemmas but also the consequential outcomes that result to others based on their decisions. Dilemmatic space, as a concept, was first introduced by Honig (1996) within the field of politics but has since been applied across sectors of public service, social work, and more recently, education (Singh et al., 2015; Turner, 2016). In education, dilemmatic space considers the "social constructions resulting from structural conditions and relational aspects in everyday practices," as teachers face dilemmas throughout their professional practice (Fransson & Grannäs, 2013, p. 7). As Turner's (2016) research identified, conceptualizing teacher practice through dilemmatic space draws focus to the power relationships that shape teacher decisions, noting specifically that engaging in these spaces shapes professional identity construction and communities of practice (Looney et al., 2017).

A central area in which teachers face dilemmas is classroom assessment (Jorro, 2013), and specifically, grading (Green et al., 2007; Pope et al., 2009; Sun & Cheng, 2014). Cogently, as Farias et al. (2010, p. 338) stated, "the most challenging dilemma for teachers: does the teacher play the role of a judge or serve as a partner in students' learning?" Studies in teachers' responses to grading and assessment dilemmas are timely given the current widespread accountability mandate across North America and other countries, which places pressure on teachers to integrate assessment throughout instruction and report on learning through summative and large-scale assessments (OECD, 2012).

Additional studies across contexts using diverse methodologies are needed and recommended in part because understanding teachers' approaches to and responses to assessment dilemmas allows for the emergences of new assessment practices that address persistent issues in current and traditional approaches (Havnes & McDowell, 2008). Further, new dilemmas emerge from these practices, and working through these dilemmas has the potential to spur teachers' learning in assessment (Havnes & McDowell, 2008; Pareja Roblin & Margalef, 2013).

Methods

Contexts

In order to understand grading as a dilemmatic space across cultural and subject contexts, data for this study were derived from two case contexts: one from teachers in Ontario, Canada, and one from teachers in the State of Vaud, Switzerland. In both cases, selected focal teachers reflected the perspectives of mathematics and language educators. Teachers' grading practices in each of these cases were shaped by local education and assessment policies as well as the social and cultural values of assessment within each context.

In the case of Ontario, assessment and grading practices are guided by the province's *Growing Success: Assessment and Reporting in Ontario Schools* (2010) policy document. Teachers are directed to engage a balanced approach to assessment that involves assessment *for* (formative), *of* (summative), and *as* learning practices. Reporting occurs through graded achievement in relation to provincial standards and curriculum expectations as based on varied forms of assessment (e.g., performance tasks and tests). Grading scales for reporting differ based on level of education (letters in elementary years and percentage in secondary); however,

the province does provide an overarching 4-level achievement chart framework to guide assessment activities, which is criterion-based.

In Switzerland, educational policies are decentralized. The federal government sets the general direction for education, but each canton (state) enacts its own laws and standards. For example, each language region has its own curriculum. In the case of the State of Vaud, the Education Law (LEO) specifies that any assessment process must support learning. However, the policies for teachers focus on summative assessment: an average number of tests per semester and per discipline is imposed with guidelines specifying how to develop grading scales (from 1 to 6, with half grades, and grade 4 as sufficient) and calculating subject averages. These requirements have led a significant number of teachers to focus on grading issues from an administrative rather than a pedagogical perspective (Pasquini, 2017).

Participants

Qualitative data on secondary teachers' grading work were collected from 11 Ontario and eight Swiss mathematics and language teachers (respectively English and French). Ontario teachers were invited to participate via email recruitment. They ranged in years of teaching experience from 2 to 27. The teachers primarily taught mathematics or English language courses, among other subjects, at schools in rural (n = 4) and suburban (n = 7) contexts, representing four schools in total.

Swiss teachers were recruited through their affiliation with a training program. They ranged in years of experience from 3 to 21, and they taught in suburban contexts in the last 3 years of school (students' age: 13–16). Among them, four taught mathematics (three of whom worked at the same school), and the remainder taught French at the same school. Teachers represented three schools. They were all involved in a training session on summative assessment. Table 1 provides demographic details on the teacher participants.

Table 1: Demographic Profiles of Teacher Participants

Country	Teacher ID	Subject	Teaching Experience (Years)	Teaching Context
Canada	Teacher 1	Mathematics	27	Suburban
	Teacher 2	Mathematics	5	Suburban
	Teacher 3	Mathematics	8	Suburban
	Teacher 4	Mathematics	12	Suburban
	Teacher 5	Mathematics	3	Rural
	Teacher 6	Language	2	Rural
	Teacher 7	Language	4	Suburban
	Teacher 8	Language	13	Suburban
	Teacher 9	Language	12	Suburban
	Teacher 10	Language	7	Rural
	Teacher 11	Language	17	Rural
Switzerland	Teacher 1	Mathematics	15	Suburban
	Teacher 2	Mathematics	13	Suburban
	Teacher 3	Mathematics	3	Suburban
	Teacher 4	Mathematics	9	Suburban
	Teacher 5	Language	14	Suburban
	Teacher 6	Language	16	Suburban
	Teacher 7	Language	8	Suburban
	Teacher 8	Language	21	Suburban

Data Collection

With reference to our research questions, we examined the grading experiences of mathematics and language teachers within two cases (i.e., Ontario and Vaud). Our goal was to describe a complex human process, without suggesting the generalization of results (Albarello, 2003; Miles & Huberman, 2010). The sample was formed in order to interview different, non-exhaustive profiles of secondary teachers from two different contexts, underlying the cross-cultural perspective of grading (Cheng et al., 2018; 2020).

All Ontario teachers participated in group-based interviews with one teacher participating in an individual interview (Patton, 2015). These interviews lasted approximately 60 minutes, were audio-recorded, and transcribed verbatim (see also Cheng et al., 2020). Before and after the training session, the Swiss teachers were each involved in individual interviews (Kaufmann, 2011) which lasted nearly 2 hours and were recorded and transcribed verbatim. All interviews were conducted by the researchers, and in the Canadian context, by trained research assistants. All interviews used a semi-structured protocol, and the questions focused on teachers' approaches to grading, challenges, and factors that influenced their grading practices, actions, and decisions.

Data Analysis

In order to analyze teachers' grading work, we used an inductive approach (Anadon & Guillemette, 2007). More specifically, we followed Willis et al.'s (2017) analytic approach. This approach was predicated on Archer's (2003, 2007, 2012) three-dimensional framework for teacher reflexivity that considers teachers' work in relation to their personal and contextual conditions. We selected this framework because it enabled us to consider the conditions that shaped teacher's grading dilemmas (i.e., research question 1) and characterize the dimensions of the dilemmatic space of teachers' grading work (i.e., research question 2).

Specifically, Archer's framework includes an analysis of teachers' discernments, deliberations, and dedications. *Discernments* involves teachers' identification of positive and negative concerns related to their practice (i.e., identification of dilemmas; Willis et al., 2017). An analysis of *deliberations* refers to the contextual factors teachers negotiate in response to dilemmas (i.e., discernments), which often involves weighing perspectives and evidence. Finally, *dedications* relates to teachers' response to the dilemma based on their deliberations, resulting in a commitment to an understanding or action. Dedications may also involve inaction. In this study, these 3-Ds served to analyze the dilemmatic space of grading by exploring not only dilemmas of practice (i.e., discernments) but also teachers' deliberations and dedications towards grading dilemmas.

Data across Swiss and Canadian teachers were analyzed jointly following Willis et al.'s (2017) approach in order to, first, identify the spaces teachers experience while grading. Subsequently, we examined how the data were represented by cultural affiliation and teaching subject. In addition, the Canadian data used in this study was part of a larger research study that explored broader themes in relation to grading practices in Canada and China (see also Cheng et al., 2020). In this paper, we used a subsection of the Canada data (i.e., English and mathematics teachers who completed the interviews) to explore the specific phenomenon of grading as a dilemmatic space. As part of our effort to explore the cross-cultural dimensions of grading, in this study we compared Canadian and Swiss teachers' grading practices. We inductively examined all interview and focus group data from Canadian and Swiss teachers through a three-step process. First, we coded transcripts for teachers' articulated grading discernments (i.e., dilemmas of practice). We then re-coded data associated with each discernment by teachers' deliberations (i.e., the factors and processes of deliberations that shaped each dilemma). Finally, we coded teacher dedications associated with identified discernments (i.e., responses to each dilemma), including any dedications of inaction. Each

grouping of codes (discernment-deliberation-dedication) were then clustered via logical association or co-occurrence to characterize broader dilemmatic spaces of grading faced by teachers across contexts (Canada and Switzerland). This analytic approach resulted in five dilemmatic spaces that characterized clusters of discernments, deliberations, and dedications associated with teachers' grading work (Raab, 2015). Our inductive analyses were conducted by two raters who applied the codes across the verbatim transcripts and discussed any disagreements in applying the codes until consensus was reached. Our results pointed to five spaces teachers across cultures and teaching subjects encounter while grading.

Results

Our analysis is presented in relation to five overarching spaces, which present areas of discernment for mathematics and language teachers' grading work and the related deliberations and dedications that were considered by these teachers: (a) justifying criteria and scale sufficiency, (b) inconsistencies across teaching contexts, (c) commitment to student learning, (d) grading in a multicultural and multilingual context, and (e) grading in relation to future goals, destinations, and consequences. Combined, these dilemmatic spaces offer evidence of the complex context of grading, a context characterized by multiple traditions, epistemologies, stakeholders, and power relationships. Embedded within our articulation of these spaces, we identified data across cultural contexts and teaching subjects, demonstrating that these five spaces show applicability (to varying degrees) within and across these contexts.

Justifying Criteria and Scale Sufficiency

This space is characterized by teachers' capacity to justify and use criterion-based arithmetical scales to quantify learning into differentiated levels of performance. At the heart of this space was the discernment of insufficiency of scales to effectively represent the diversity of learning within and between subjects. For example, in some cases, mathematics and language Swiss teachers were challenged in using an 11-point scale even in contexts where student learning could be considered "right" or "wrong" (i.e., dichotomous). In other instances, they were challenged to differentiate more complex learning across several criterion intervals, yet with little guidance on how to establish and justify thresholds between scale levels.

In deliberating about this discernment, some teachers of both disciplines recognized that despite the widespread policy endorsement of criterion-based grading scales, such scales were difficult to implement for all types of learning in their classroom (i.e., lower to higher order learning). One Canadian mathematics teacher commented, "when we moved towards the achievement chart system [grading using a criterion-based 4-level rubric] it pushed us to really think about the difference between one piece of work and other. But sometimes it's hard to use all four levels for an assignment. The student either gets it or they don't" (Teacher 3). As a way to address this issue, which was also prevalent in Switzerland, two Swiss French teachers decided to reduce the scales. Teacher 5 stated: "I wondered why it was necessary to go down so low. I mean, I don't do it. I say to myself, if it's not enough, it's 3, or 3.5, but it's not enough, that's it. I don't need to grade any lower." In relation to grammar assessment, Teacher 6 stated, "for the moment ... I was thinking, for example, to assess things more grammatically, or like that, I haven't yet switched to qualitative criteria. So, for now, I'm staying on points." Similarly, Swiss math Teacher 2 noted that "in setting the criteria ... I'm not convinced to ignore the points ... [points] allow me to be better in my coherence."

For higher order learning objectives, teachers were equally challenged in using criterion-based grading scales, particularly in establishing justifiable thresholds between scale levels; rather, the criteria across levels were viewed as subjectively established and at times difficult to differentiate. Accordingly, teachers manipulated the criteria and scales to their own approach to grading and to their own students' learning. Canadian math Teacher 2 captured this

challenge well, “I mean levels are confusing,” they further went on to discuss how they “play with” level criteria to adjust grades because they are so subjective:

> So, I may give a Level 4 overall for an assignment, really unfortunately that doesn’t mean a whole lot in the grand scheme of things. Or fortunately it allows me to play with it because for the overall expectation in their [the students’] math course that I’m offering, they may have gotten a Level 2.

In the Swiss context, teachers manipulated the scale in various ways to establish a justifiable approach to grading. In both disciplines, some teachers reduced the lower end of the scale (i.e., levels associated with failure) as they felt that there was little value in “grading failure.” Other teachers created grading bands, clustering levels to reduce the 11 to 6 or 8 differentiable levels, recognizing that their assessments were not precise enough to differentiate student learning across 11 levels of performance. Swiss math Teacher 3 noted: “what I would like to do, it might be to keep the grades from 4 to 6 and then just report when it's not successful, and then what's not successful,” and another, Teacher 1, followed: “I find it difficult, after a while, to be sharp enough on this, and then ... indeed, it's a pity that we couldn't keep only the 4, 5, and 6 grades. Because we would be fairer.” When teachers manipulated the scales in these ways (i.e., dedication), they were better able to justify their grading decisions to their students, parents, and others, avoiding the common experience of bargaining with students over points on assessments.

Inconsistencies Across Teaching Contexts

A core discernment observed in both Swiss and Canadian teachers’ data was the persistence of grading inconsistency within their own subject and across secondary grade levels, other teachers and disciplines, and in relation to previous grading traditions. What was driving this discernment was the expectation from students and parents for consistent grading practices across contexts.

As teachers deliberated this discernment, they discussed several factors that influenced their grading practice. When grading, teachers considered the convention of their discipline and their orientation towards grading (e.g., rather criterion-based for language teachers and points-based for mathematics teachers) in relation to system priorities, historical orientations, and other teachers’ practices. Specifically, it first attempts to describe why teachers grade differently in their discipline. As an example, Swiss math Teacher 2 compared his criterion-based assessment practice to his French colleagues’:

> In a French test, there had to be certain elements present to get a grade of 4. And then, little by little, the student, depending on the elements present, could reach a grade of 6. I found it very interesting and ... I would have liked to do something a bit similar. But I can't really see, in math, how to do it. Because there may be too few elements. In a test, I don't know, there might be five learning objectives assessed ... I don't know how, I couldn't do five levels, it's clearly not possible.

Evident in this discernment is the variable adoption of newer grading approaches, particularly the use of criterion-based systems, across teachers. Canadian Teacher 7 noted:

> I do think there are teachers out there that still have the very traditional marking scheme right. Like I know the majority of us use rubrics, right? And I know there are some teachers you know who I think will use point systems and evaluate tests that way and it’s unclear for the students, like ‘how do I get 3 marks on this? I don’t know what 3 marks looks like.’

Similarly, in the Swiss context, math Teacher 3 expressed, “transforming criteria into grades is not easy, or it means putting as many criteria as there are grades and then we find ourselves in the system of grades. It doesn't change anything, let's say.” As a result, students experience

different approaches to grading, rooted in different epistemological orientations, depending on their teacher and, to some extent, the discipline. While there may be ways to enhance consistency in grading practices across a school (e.g., professional development), teachers in our study generally did not articulate a dedication to this discernment other than providing their own students with a well-articulated justification for their grading practice.

Commitment to Student Learning

This space characterizes the difficult discernment teachers faced in promoting student-centered teaching and feedback despite the pervasive grading culture in their classroom and school. Central to this discernment was teachers' desire to redirect students' interest and attention towards feedback that promote their learning rather than focusing on grades. One English language teacher in Canada effectively captured this discernment:

> I spend so much time writing very specific, very beautiful rubrics that nobody reads; they will never read them. And even when you hand it back, they never read them. They just look at the mark. They look at the percentage. That's all they want. And then they say, "Why did I get this?" "Well, I wrote a comment, but all they want is the grade" (Teacher 8).

In this discernment, teachers fundamentally asked, "how do you de-emphasize the importance of grades?" (Canadian Teacher 9). In deliberating this discernment, three factors were discussed: (a) the "crushing" (Swiss mathematics Teacher 1) effects on students caused by a single grade when it reflects a low level of performance, (b) the linkage of criteria to grades as a strategy to bring meaning to grades, and (c) the quality of student feedback and opportunity to use feedback to support learning.

Teachers' dedications to this discernment varied. Despite the fact that for some teachers, grades do not, to themselves, provide enough information to direct learning, they did recognize that grades could stimulate learning through purposeful linking to criteria and through having conversations with students about their grades. Swiss math Teacher 1 observed the following in relation to student learning:

> All this work had a qualitative influence on the preparation of the assessment. And afterwards, the student will not have, through this grade, a qualitative answer. On the other hand, it allows me to invite him/her to a qualitative work. I think I've changed the structure of my teaching a little bit … to make the students more aware of what they were going to be assessed on, all the time. To say, "Can you tell me what that refers to in the objectives I gave you?"

In some other situations, teachers went further. They acknowledged that criteria played a key role within the construction of grades and was the basis for effective feedback. Swiss math Teacher 3 described the following criterion-based approach to grading and feedback:

> I tell myself ... I've defined thresholds of sufficiency. And then the idea was that all the sufficiency thresholds would be reached. Now I imagine a student who has reached almost all the thresholds, except on one criteria. I can't see myself telling him, "Here it is, your evaluation it's not worth 4, because there's an objective or a task, well, an objective that you didn't achieve." So, telling the student that it's a failure just for that, no. On the other hand, it's good to tell him, "Listen, there, there, I mean, you have to get better at this place, you don't know how to do it, so you have to work on that specific criteria." In that way, it's good. But I don't think you can do that and then put a grade. At last, I can't see myself doing it. We need to discuss the criteria and grade through feedback.

In a similar way, a Canadian mathematics teacher's dedication to this discernment was to not provide students with grades initially but rather only criterion-based feedback. She noted:

> If they're hearing that their parents want that number, 'I want you to get that 90,' then it's hard for us to then say, 'well I didn't actually put a number on it but if you were measuring that against, you know exactly where you stand, and here's the next step that I've given you feedback

> wise to move you forward.' So, I now start with the feedback and then invite them to work on it. Then I give them their grade (Teacher 1).

In this way, some teachers are trying to change the culture—both for students and parents—from one that focused on grades to one that focused on criteria and feedback for learning. Across both contexts and disciplines, however, teachers articulated this cultural shift to be challenging and that many teachers continued to confront this discernment despite their progressive dedications.

Grading in a Multicultural and Multilingual Context

Teachers in our study recognized that grading was increasingly referenced to the diversity, specifically cultural and linguistic, of students in their classes, which further challenged notions for grade fairness, access to assessments, and cultural understandings about the value and role of grades. Swiss math teachers and Canadian teachers (both disciplines) articulated that a high degree of cultural diversity amongst students pushes them to consider the importance of consistent grading across students versus accommodating assessments and grading practices for diverse learners. Further complicating this dilemmatic space was the involvement of parents from diverse cultures in grading: navigating and negotiating their expectations for student grades and communicating local grading practices.

With respect to altering grading practices for linguistically diverse students, teachers noted that in contrast to students with identified exceptionalities, there were fewer formal provisions for accommodations and guidelines with respect to grading. One Canadian math teacher clearly articulated:

> I have students that are extremely strong in mathematics, but their English isn't there yet. I can't give them full points because they can't communicate a thinking problem or something more complex. However, I'm sure if it was in their first language, they would nail it" (Teacher 3).

This quotation points to a common sentiment across teachers in which they wrestled with maintaining a consistent approach to assessing linguistically diverse learners alongside native-language learners despite recognizing the limitations of assessment to fully measure their learning. In the same way, Swiss math Teacher 4 noticed: "foreign students with special needs, when I assess the fractions, I let them use the calculator for the booklets because I don't want to penalize them on that, that's not the point either." A Canadian mathematics teacher recognized that "it is frustrating and very difficult because I can get from them … the deeper stuff that I'd like to investigate with them. So, it's one of our biggest challenges" (Teacher 4).

To complicate grading in a multicultural and multilingual setting was to acknowledge that grades are understood and represented differently across cultures. Hence, students approach summative assessment differently based on their cultural background and hold different expectations about the outcome from such assessments. A Canadian mathematics teacher noticed that assessment "is where you see quite a difference in the education systems between the different cultures" (Teacher 5). Teachers recognized that, in fact, assessment itself was the key to negotiating understanding about grading. Through increased levels of feedback, not only about student performance but also about the value and role of the assessment, some teachers were able to bridge cultural understanding of assessment.

Returning to the mathematics example used above, the teacher resolved cultural differences in grade understandings in the following way:

> I don't focus so much on what I assess and how I assess. It's more the different types of feedback I give. A good example for this would be my Grade 10 class last semester. I had one student who had all of the math concepts down. He knew more than what he needed to know. But he couldn't communicate it—he couldn't clearly communicate his thinking. So, with him, all of the feedback I was giving him was

> related to "okay, this is the correct answer, but you need to make sure you define your variables, you need to make sure you show your steps, you need to make sure you write that final statement, and that relates back to the problem. The purpose of this assessment isn't just to get the right answer but also for you to communicate your thinking.

In this example, we see that students from a diverse cultural background might not fully understand the intentions of an assessment in a new country and that the role of feedback and grading can support their acculturation into a new grading culture.

Negotiating understandings about grading with parents from diverse cultures was also noted by teachers, commenting that parents often raised expectations that their child received a certain grade. One Canadian mathematics teacher stated, "sometimes parents have different understandings of how we grade here and part of our jobs is helping them to understand that" (Teacher 4). Another Canadian mathematics teacher supported this notion, offering the following anecdote:

> I had one newcomer family who wanted their son to become a doctor and go to a top school. So, they would always come back to me after every test and ask why the grade wasn't higher. They kept suggesting that I wasn't doing my job enough if their son wasn't getting high enough grades (Teacher 2).

In that case, the teacher went on to say that they tried to communicate the province's standards-based approach to grading and make the grading criteria transparent but there was little change in the parent's approach over the term.

Grading in Relation to Future Goals, Destinations, and Consequences

This space discerns the consequential aspects that influence grading practices and decisions. In particular, teachers' deliberations showed that they were challenged by pressures provoked by postsecondary consequences (e.g., acceptance to university, college and scholarships, or professional apprenticeship) with these pressures optimized through (a) parental expectations to inflate grades, and (b) a personal desire to support students' success in their postsecondary school ambitions. Interestingly, while teachers strived to support students in their future goals, they also wanted to be "honest" with students about the realities of assessment in university, college, and "real life" (Swiss math Teacher 2) and ensure that they were prepared to meet those challenges. In a recursive move, rather than inflating grades, for some teachers from both contexts and disciplines, supporting students for the future involved increasing the difficulty of assessment and grading in their secondary school context.

Parental expectation for inflated grades was a commonly expressed pressure, particularly in relation to supporting student in gaining admission to university or college. A Canadian math teacher provided additional rationale for parental pressure and assertion linking it to a broader shift in parenting and societal expectations. He observed that parents used to accept and respect the students' assessment of their child and when the child was not successful the parent would more often than not ask the child why they didn't succeed rather than ask teachers to raise the grade. Now, he noted:

> The idea of the helicopter parents, the I'm gonna come in, I'm gonna support my son as much as I can, I'm gonna show him that I love him and that I will go to the wall for him. And so, parents want their kids to succeed, and again I think this issue isn't a school issue, it's more a societal issue and how society has really really changed. (Teacher 4)

While teachers tended not to alter grades due to parent requests, they did tend to consider students' future goals in their own grading. For example, in relation to one struggling student who wanted to attend university, a Canadian mathematics teacher stated:

> Facing those students who try really hard is difficult because you want to help them to succeed. But I'm not really sure, you can't adjust their grades … but you want to go out of your way to provide them with maybe an extra assignment that you know is going to assess their strength but still going to meet curriculum expectation. So, you know, providing them with the way to show that they, they can do everything. (Teacher 1)

Adjusting assessment and grading practices for students who "try hard" or for those who need a higher grade to achieve a future goal was a suggested practice by several teachers.

In contrast, there were other teachers who wanted to prepare students for future assessment practices at university, college, or the workplace, even though doing so resulted in a more difficult assessment experience at the secondary level. However, what's important here is that these teachers modified their assessment and grading practices based on students' future goals and destinations. Swiss math Teacher 2 said:

> I'm really keen to bring them into the professional world and have them ... well-armed for it. I already find that when we go to see the managers, they tell us, "There's a huge gap between school and afterwards." So, I say to myself, "the sooner students are ready for what comes next, the better it is for them." And, in the professional world, when they go to do tests, everything is assessed in the same way ... my goal for these students who have difficulties in math is to really bring them towards, as close as possible to what they will have afterwards.

Similarly, one Canadian math teacher recounted an anecdote of a graduate from their school who returned to speak to the students about the realities of college life and assessment. The teacher described his speech:

> I told my calculus class, really be careful, next year's gonna be hard. It's gonna be a slap in the face. I then show them all these statistics about the failure rates of colleges and showing the success rates; only about 60 % of the people move on after first year … So, when we did this kinda survey, this one girl said, are you, are you trying to scare me? 'Cause I really feel scared about going to university. No, I'm trying to be honest with you. So, when we did the grad panel today, everything kind of, it felt better coming from a peer than from a teacher. (Teacher 1)

Practically, the same teacher also described modifying his grading practice to mirror more common practices used at universities and colleges: "I tend to use more tests, multiple choice, and challenging questions because this is what it's gonna be like for them" (Teacher 1).

Discussion

This study aimed to explore more deeply the dilemmatic space that characterizes mathematics and language teachers' grading practices and decisions. Through a framework of dilemmatic space, we acknowledge that grading is shaped by a multiplicity of factors that are situated in a relational environment that involves negotiation and balancing of the ever-present social, cultural, political, epistemological, and consequential factors (Fransson & Grannäs, 2013). Therefore, in characterizing the dilemmatic space of grading, we interviewed teachers in Canada and Switzerland to analyze their grading work and the factors that shaped their grading practice. In doing so, we arrived at five cross-cultural spaces that articulated central grading discernments, deliberations, and dedications (Archer, 2003, 2007, 2012). Each of these spaces represented a complexity of negotiation, suggesting that grading is indeed emblematic of a dilemmatic space (Fransson & Grannäs, 2013; Singh et al., 2015). Previous literature would

suggest that culture shapes assessment practice (Allal, 2016; Looney et al., 2017), yet findings from this research showed that there were consistencies (to varying extents) in the dilemmatic space in relation to grading experienced by the teachers in this study, who were teaching across two cultural contexts. Hence, we assert that while culture does lead to unique practices, there may be underlying connections and similarities in complex assessment processes when cultures operate from similar education frameworks and principles.

In this discussion, we read across the spaces to consider the essential dimensions of the dilemmatic space in which teachers grade. At the onset of this discussion, we acknowledge that these dimensions result solely from our exploratory study, which should be extended with research into other contexts. Our intention in articulating these dimensions is to begin mapping the broader sociocultural condition that shapes grading practices and decisions in schools—an activity we believe to have merit for the development of more nuanced grading policies, as a foundation for future research, and as validation for teachers' grading realities. Derived from each space, and reflective of teachers in both Canada and Switzerland, the following four dimensions begin to characterize the dilemmatic space of grading based on our sample:

1. ***Temporal Dimension:*** The temporal dimension recognizes that grading is influenced by conceptions, practices, and consequences of grading that range from a historical (i.e., subjects' traditional grading practices) to a future view. On the one end of this dimension, the regressive end, mathematics and language teachers typically consider the tradition of assessment and grading that influence their practice in relation to discipline-specific approaches or entrenched school practices and policies. On the other end, the progressive end, almost all teachers consider the influence and consequences of their grading actions based on the validity of grade decisions and on the outcomes of students' futures (e.g., entrance to university). Put together, this dimension challenges teachers to consider the value of historically rooted grading conventions (i.e., a regressive interrogation) as serving grading actions that account for current learning and which yield valid consequences for the future (i.e., a progressive interrogation).
2. ***Epistemological Dimension:*** At the heart of this dimension is the negotiation of what constitutes learning in relation to how it is assessed and graded. Teachers in both disciplines struggled to negotiate between competing epistemological views of learning—one where learning could be effectively summarized in a grade and which spurred a "grading culture" of learning, as in mathematics—and one where learning was best represented through ongoing feedback and considered along a trajectory of continuous growth, as in language (i.e., "feedback culture") and in some cases, mathematics. For all teachers, a big challenge was negotiating their aims to cultivate a feedback culture in their classroom despite a common grading culture endorsed by students and parents.
3. ***Sociopolitical Dimension:*** The sociopolitical dimension considers the influence of others on teachers' grading practice including parents, students, and colleagues. Through these relationships, a network of competing demands is placed on teachers through their grading work including the need to (a) ensure communication, justification, and understanding of grades across stakeholders, (b) demonstrate accountability of teaching and learning, and (c) negotiate and establish expectation for grades. Importantly and regardless of subject matters, teachers' response to these demands are further complicated as they operate within a political context, shaped by a power hierarchy between parents and system pressures. Hence, what should be primarily focused on constructing accurate representations of students' learning becomes, in part, about negotiating a complex sociopolitical sphere.
4. ***Adaptive dimension****:* This final dimension considers grading as a responsive act in which all teachers adapt to an ever-changing context. Both mathematics and language

teachers noted how their grading work has been altered or challenged by changes in grading policies, classroom demographics and diversity, parental expectations, and orientations towards learning (by students and parents). Embedded in this dimension is teachers' adaptive capacity to embrace new grading requirements (e.g., grading scales) and to merge these with newer notions of assessment (e.g., the rise of assessment for learning mandates) while supporting parents' and students' understandings of these shifts. Central to this dimension was then the teachers' recognition that they were required to accept and negotiate different teachers', parents', and students' endorsement of newer approaches, which often created tensions in practice and different expectations amongst stakeholders. Also embedded in this dimension is the influence of changing classroom demographics—classroom with increased diversity— which contributes to multiple perspectives and expectations of grading by parents and students from diverse backgrounds.

In earlier work, McMillan and Nash (2000) noted that a confluence of factors (i.e., teachers' beliefs and values, classroom realities, and external factors) led to tensions in teachers' grading work. Our research suggests that grading is always, by its nature, in a dilemmatic space. The central tension that inevitably arises from such a space is the necessity to arrive at grading decisions—dedications—embedded in specific disciplinary epistemologies, and despite the uneasy and unsettled ground in which these dedications are situated. Hence, we argue that despite the requirement for grading decisions and actions, there is an inherent difficulty of dedication because grading is situated in a dilemmatic space that involves continuous negotiation of temporal, epistemological, sociopolitical, and adaptive dimensions.

In a reflective way, this argument counters traditional conceptions of grading as a seemingly objective process; instead, it is one that heeds the complexity of teaching, learning, and assessment in today's diverse classrooms (see also Guskey, 2013). Through this argument, we emphasize that grading in a dilemmatic space, in contrast to a strictly arithmetical view, teachers actively consider the role of students' learning in decision making. It therefore pushes them to consider students' diverse forms of learnings rather than just measuring them (Kohn, 2011; Shepard et al., 2018; Wiliam, 2008).

However, as seen in our data, the ways in which this finding plays out in practice differs greatly by teachers and sometimes by subject. Accordingly, we assert that additional research should explore how grading in dilemmatic space—where students are positioned at the center of grading decisions and in consideration of the four fundamental dimensions articulated above—translates into teachers' grading dedications. Pursuing this line of research may not only deepen understandings about the complexity of grading practices but also open new spaces of teacher inquiry. By inviting teachers to interpret and analyze their grading practice from a dilemmatic perspective, we hope that it may yield additional possibilities for grading and assessment practice (Havnes & McDowell, 2008; Pareja Roblin & Margalef, 2013).

While findings from this study have enabled a deeper understanding of the dilemmatic space situating teachers' grading work, results should be interpreted within the scope of data collected. Even though using a qualitative interview methodology was consistent with our small sample, specific findings from this study can't be generalized and applied to other teachers or contexts. Further, data were collected in slightly different ways in each context due to local opportunities to lead such research. Nonetheless, data collected in this research provided a valuable foundation for theorizing and interrogating the dilemmatic space of grading for more extensive research.

To this end, we assert that future studies on grading should recognize the complex context in which grading actions and decisions are rendered and continue to map this dilemmatic space. Our research was reflective of two learning cultures, hence it examined dilemmatic space as a

cross-system phenomenon (Cheng et al., 2018, 2020). In so doing, we were able to find consistent dimensions that characterized this space for both Canadian and Swiss mathematics and language teachers. Additional scholarship is needed, however, to explore the relevance of these dimensions across other teachers and contexts, in the same disciplines or in others with specific epistemologies (e.g., arts, history, geography, philosophy), and to continue to refine and expand our characterization of grading as dilemmatic space. In this pursuit, a complementary necessary line of research is to explore how such an expanded and complex view of grading aligns with classroom assessment theories related to validity, reliability, and fairness, and to consider what affordances are required to these theories to accommodate such a view. From our perspective, driving this research agenda should be the construction of assessment theory, policy, and practice that adequately recognize and support the difficulty of grading deliberation and dedication within dilemmatic spaces.

References

Achinstein, B. (2002). Conflict amid community: The micropolitics of teacher collaboration. *Teachers College Record, 104*(3), 421–455.

Albarello, l. (2003). *Apprendre à chercher : l'acteur social et la recherche scientifique*. Bruxelles : De Boeck.

Allal, L. (2012). Les exigences inconciliables des activités évaluatives et leurs paradigmes de référence. In L. Mottier Lopez & G. Figari (dir.), *Modélisations de l'évaluation en éducation* (p. 181–194). De Boeck.

Allal, L. (2016). The co-regulation of student learning in an assessment for learning culture. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 259–273). New York, NY: Springer.

Alm, F., & Colnerud, G. (2015). Teachers' experiences of unfair grading. *Educational Assessment, 20*(2), 132–150. https://doi.org/10.1080/10627197.2015.1028620

Anadon, M., & Guillemette, F. (2007). La recherche qualitative est-elle nécessairement inductive? *Recherches qualitatives, 5*, 26–37.

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.* Toulouse: Math'Adore.

Archer, M. (2003). *Structure, agency and the internal conversation*. Cambridge: Cambridge University Press.

Archer, M. (2007). *Making our way through the world*. Cambridge: Cambridge University Press.

Archer, M. (2012). *The reflexive imperative in late modernity*. Cambridge: Cambridge University Press.

Balan, A., & Jönsson, A. (2019). Increased explicitness of assessment criteria: effects on student motivation and performance. *Frontiers in Education, 3*(81), 56–68.

Biberman–Shalev, L., Sabbagh, C., Resh, N., & Kramarski, B. (2011). Grading styles and disciplinary expertise: The mediating role of the teacher's perception of the subject matter. *Teaching and Teacher Education, 27*, 831–840. https://doi.org/10.1016/j.tate.2011.01.007

Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Proceedings of Teaching and Learning in Higher Education: New Trends and Innovations. University of Aveiro: Portugal.* https://pdfs.semanticscholar.org/edd0/9f550a0cef684a4ea0bfeefdcc25cc43d90d.pdf

Braxmeyer, N., Guillaume, J.-C., & Levy, J.-F. (2005). Les pratiques d'évaluation des enseignants au collège. *Les dossiers du Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, Direction de l'évaluation et de la prospective, 160*, 5–117.

Brookhart, S. M. (2013). Grading. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257–271). Thousand Oaks, CA: SAGE Publications.

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

Brookhart, S., Guskey, T. R., Bowers, A. J. McMillan, J. H., Smith, J. K., & Smith … 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.

Carey, T., & 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.

Chen, P. P., & Bonner, S. M. (2017). Teachers' beliefs about grading practices and a constructivist approach to teaching, *Educational Assessment, 22*(1), 18–34.

Cheng, L., DeLuca, C., Braund, H., & Yan, W. (2018). Grading in Canada and China: A comparative study. *Comparative and International Education, 47*(1), 1–24.

Cheng, L., DeLuca, C., Braund, H., Yan, W., & Rasooli, A. (2020). Teachers' grading decisions and practices across cultures: Exploring the value, consistency, and construction of grades across Canadian and Chinese secondary schools. *Studies in Educational Evaluation, 67*, https://doi.org/10.1016/j.stueduc.2020.100928.

Conseil des Ministres de l'Éducation (Canada) (CMEC). (2010). *Programme Pancanadien d'Évaluation de 2010. Rapport de l'évaluation pancanadienne en mathématiques, en sciences et en lecture* (ISBN 978-0-88987-213-4). https://www.cmec.ca/412/PPCE_2010.html

Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire?* Bruxelles: De Boeck.

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.

Cuban, L. (1992). Managing dilemmas while building professional communities. *Educational Researcher, 21*(1), 4–11. https://doi.org/10.3102/0013189X021001004

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 355–375.

Docan, T. N. (2006). Positive and negative incentives in the classroom: An analysis of grading systems and student motivation. *Journal of Scholarship of Teaching and Learning, 6*(2), 21–40.

Dubus, A. (2006). *La notation des élèves. Comment utiliser la docimologie pour une évaluation raisonnée*. Paris: Armand Colin.
Duncan, R. C., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *The Alberta Journal of Educational Research, 53*(1), 1–21.
Ehrich, L., Kimber, M., Millwater, J., & Cranston, N. (2011). Ethical dilemmas: A model to understand teacher practice. *Teachers and Teaching: Theory and Practice, 17*(2), 173–185.
Enyedy, N., Goldberg, J., & Welsh, K. (2006). Complex dilemmas of identity and practice. *Science Education, 90*(1), 68–93.
Farias, G., Farias, C. M., & Fairfield, K. D. (2010). Teacher as judge or partner: The dilemma of grades versus learning. *Journal of Education for Business, 85*, 336–342.
Fransson, G., & Grannäs, J. (2013). Dilemmatic spaces in educational contexts—towards a conceptual framework for dilemmas in teachers work. *Teachers and Teaching, 19* (1), 4–17. https://doi.org/10.1080/13540602.2013.744195
Green, S. K., Johnson, R. L., Kim, D.-H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education, 23*, 999–1011. https://doi.org/10.1016/j.tate.2006.04.042
Guskey, T. R. (2013). The case against percentage grades. *Educational, School, and Counseling Psychology Faculty Publications. 22*. https://uknowledge.uky.edu/edp_facpub/22
Guskey, T. R., & 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
Havnes, A., & McDowell, L. (2008). Introduction: Assessment dilemmas in contemporary learning cultures. In A. Havnes & L. McDowell (Ed.), *Balancing dilemmas in assessment and learning in contemporary education* (pp. 3–15). New York : Routledge.
Honig, B. (1996). Difference, dilemmas, and the politics of home. In S. Benhabib (Ed.), *Democracy and difference. Contesting the boundaries of the political.* (pp. 257–277). New Jersey: Princeton University Press.
Isnawati, I., & Saukah, A. (2017). Teachers' grading decision making. *TEFLIN Journal, 28*, 155–169.
Jönsson, A. (2014). Rubrics as a way of providing transparency in assessment, *Assessment and Evaluation in Higher Education, 39*(7), 840–852. https://doi.org/10.1080/02602938.2013.875117
Jorro, A. (2013). L'accompagnement des enseignants dans l'activité évaluative face à des situations de production écrite. *Revue française de linguistique appliquée, (XVIII),* 107–116. https://doi.org/10.3917/rfla.181.0107
Kaufmann, J.-C. (2011). *L'entretien compréhensif*. Paris: Armand Colin.
Klapp, A. (2015) Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy & Practice, 22*(3), 302–323. https://doi.org/10.1080/0969594X.2014.988121
Knight, P., & Yorke, M. (2003). *Assessment, learning and employability*. Maidenhead: Open University Press.
Kohn, A. (2011). The case against grades. *Effective Grading Practices, 69*(3), 28–33.
Lampert, M. (1985). How do teachers manage to teach? Perspectives on problems in practice. *Harvard Educational Review, 55*(2), 178–194.
Link, L. J. (2018). Teachers' grading perceptions: How pre-service training makes a difference. *Journal of Research in Education*, *28*(1), 62–91.
Looney, A., Cumming, J., van Der Kleij, F., & 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
Lyons, N. (1990). Dilemmas of knowing: Ethical and epistemological dimensions of teachers' work and development. *Harvard Educational Review, 60*(2), 159–180.
Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Effective Grading Practices, 69*(3), 34–39.
McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x
McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd ed.). Thousand Oaks, CA: Sage.
McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading* (p. 84–112). Alexandria: ASCD.
McMillan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making*. https://files.eric.ed.gov/fulltext/ED447195.pdf
Meier, S. L., Rich, B. S., & Cady, J. (2006). Teacher's use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring. *Assessment in Education, 13*(1), 69–95. https://doi.org/10.1080/09695940600563512

Miles, M., & Huberman, A. M. (2010). *Analyse des données qualitatives*. Bruxelles : de Boeck Université.
Mottier Lopez, L. (2016). La microculture de classe : un cadre d'analyse et d'interprétation de la régulation située des apprentissages des élèves. In B. Noël & S. C. Cartier (dir.) *De la métacognition à l'apprentissage autorégulé* (p. 67–78). Bruxelles: De Boeck.
Organisation for economic co-operation and development (OECD). (2012). *Grade expectations: How marks and education policies shape students' ambitions*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264187528-en
Pareja Roblin, N., & Margalef, L. (2013). Learning from dilemmas: Teacher professional development through collaborative action and reflection. *Teachers and Teaching, 19*, 18–32. https://doi.org/10.1080/13540602.2013.744196
Pasquini, R. (2017). Évaluation. Première question vive de la pratique. *Éducateur, 4*, 38–39.
Pasquini, R. (2019). Élargir conceptuellement le modèle de l'alignement curriculaire pour comprendre la cohérence des pratiques évaluatives sommatives notées des enseignants: enjeux et perspectives. *Mesure et évaluation en éducation*, *42*(1), 63–92. https://doi.org/10.7202/1066598ar
Patton, M. Q. (2015). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
Perrenoud, Ph. (2004). Évaluer des compétences. *Éducateur*, 8–11.
Pope, N., Green, S. K., Johnsson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education, 25*, 778–782.
Prøitz, T. S. (2013). Variations in grading practice—subjects matter. *Education Inquiry*, *4*(3), 555–575. https://doi.org/10.3402/edui.v4i3.22629
Raab, R. (2015). Quelles opérations de recherche dans une démarche inductive inspirée de l'analyse par catégories conceptualisantes? *Recherches qualitatives, 34* (1), 122–142.
Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26* (7), 1372–1380. https://doi.org/10.1016/j.tate.2010.03.008
Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal, 12*(3), 315–325. https://doi.org/10.1007/s11218-008-9073-z
Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (p. 1–15). London: Springer.
Shapira-Lishchinsky, O. (2011). Teachers' critical incidents: Ethical dilemmas in teaching practice. *Teaching and Teacher Education, 27,* 648–656.
Shepard, L. A., Penuel, W. R., & 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
Singh, P., Märtsin, M., & Glasswell, K. (2015). Dilemmatic spaces: high-stakes testing and the possibilities of collaborative knowledge work to generate learning innovations. *Teachers and Teaching, 2*(4), 379–399.
Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice, 22*(4), 26–33. https://doi.org/10.1111/j.1745-3992.2003.tb00141.x
Sun, Y., & 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
Thorsen, C., & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation*, *18*(2), 153–172. https://doi.org/10.1080/13803611.2012.659929
Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum, 75*(3), 210–227.
Timperley, H. S., & Parr, J. M. (2009). Chain of influence from policy to practice in the New Zealand literacy strategy. *Research Papers in Education, 24*(2), 135–154. https://doi.org/10.1080/02671520902867077.
Turner, M. (2016). Emergent identity and dilemmatic spaces: Pre-service teachers' engagement with EAL collaboration. *Teachers and Teaching, 22*(5), 570–585. https://doi.org/10.1080/13540602.2016.1158466
Ulvik, M., Smith, K., & Helleve, I. (2009). Novice in secondary school—The coin has two sides. *Teaching and Teacher Education, 25*, 835–842.
Walvoord, B. E., & Johnson Anderson, V. (2009). *Effective grading: A tool for learning and assessment*. San Francisco: Jossey Bass.
Wiliam, D. (2008). Balancing dilemmas: Traditional theories and new applications. In A. Havnes & L. McDowell (Ed.), *Balancing dilemmas in assessment and learning in contemporary education* (pp. 267–282). New York : Routledge.
Willis, J., Crosswell, L., Morrison, C., Gibson, A., & Ryan, M. (2017). Looking for leadership: The potential of dialogic reflexivity with rural early-career teachers. *Teachers and Teaching: Theory and Practice, 23*(7), 794–809. https://doi.org/10.1080/13540602.2017.1287695

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16* (4), 323–342. https://doi.org/10.1207/S15324818AME1604_4.

Raphaël Pasquini is an associate professor of the Training and Research Unit in Teaching, Learning and Assessment at the University of Teacher Education of the State of Vaud, Switzerland. His research documents the complexity of teachers' assessment and grading practices and conceptualizes under what conditions they can become more valid. Raphaël is involved in several assessment networks at the regional, national, and international levels.

Christopher DeLuca is an associate dean at the School of Graduate Studies and associate professor in educational assessment at the Faculty of Education, Queen's University. Chris leads the classroom assessment research team and is a director of the Queen's Assessment and Evaluation Group. His research examines the complex intersection of curriculum, pedagogy, and assessment as operating within the current context of school accountability and standards-based education.