

A SECONDARY SCHOOL TEACHER'S DESCRIPTION OF THE PROCESS OF DETERMINING REPORT CARD GRADES¹

MARIELLE SIMON, ROBIN D. TIERNEY, RENÉE FORGETTE-GIROUX, &
JULIE CHARLAND *University of Ottawa*
BRIAN NOONAN & RANDY DUNCAN *University of Saskatchewan*

ABSTRACT. This paper presents a grade 10 mathematics teacher's in-depth description of the process she uses to determine the final grade for the report card within a standards-based context. Her case was part of a three-year comparative study of grading practices of teachers from two Canadian provinces that differ in their level of standardization of education. Survey, interview, and document data provided by the teacher revealed clear tensions between current grading policy and principles particularly within mathematics.

DESCRIPTION PAR UNE ENSEIGNANTE DU SECONDAIRE DU PROCESSUS DE FIXATION DES RÉSULTATS AU BULLETIN

RÉSUMÉ. Dans cet article, une enseignante en mathématiques de secondaire 4 décrit en détails le processus par lequel elle détermine la note finale figurant au bulletin dans un contexte d'évaluation critériée. Son analyse fait partie d'une étude comparative de trois ans examinant les pratiques d'évaluation d'enseignants œuvrant dans deux provinces canadiennes dont les niveaux de normalisation de l'éducation diffèrent. Les données exposées par l'auteure – émanant de sondages, d'entrevues et de documents – mettent en évidence des tensions évidentes entre les politiques actuelles d'évaluation et les principes, particulièrement en ce qui a trait aux mathématiques.

Educational systems around the world have undergone major accountability and assessment reforms in recent years. The occurrence of reform is not in itself remarkable, as reforms wash repeatedly over the educational landscape (Cuban, 1990; Hargreaves, 2002), but the intensity and pervasiveness of this particular wave is notable (Linn, 2000; O'Connor, 2002; Young & Levin, 2002). It began with conservative concerns in the 1980s, and swelled into the accountability movement of the 1990s (Horn, 2004; Young & Levin, 2002), which ultimately resulted in the widespread adoption of standards-based systems

across North America (O'Connor, 2002). Standards-based systems typically have centrally-developed curriculum, common reporting mechanisms, and large-scale assessment programs.

An assumption that underlies standardization is that variations in teachers' practices will be reduced, allowing assessment results to be interpreted consistently by stakeholders and ultimately improving education. Evaluation of the impact of standards-based reform in the United States has shown some gains in student achievement (Hamilton, McCaffrey, Stecher, Klein, Robyn, & Bugliari, 2003; Hanushek & Raymond, 2003; Supovitz & Taylor, 2003), but there is currently little empirical evidence of the extent to which standards-based systems have actually improved the quality or consistency of teachers' grading practices. This paper is thus concerned with examining a secondary school mathematics teacher's narrative account of her grading practice and reactions within an educational system that values accountability and consistency in teaching and reporting grades.

Grading is a process within the practice of classroom assessment. It focuses specifically on the point where students' final grades are being determined by the teacher for report cards. It differs from assessment in that grading doesn't necessarily include all the assessment results collected over a time period (Senk, Beckman, & Thompson, 1997), but it should be based on multiple and varied sources of evidence (Joint Committee on Standards for Educational Evaluation [JCSEE], 2003; National Council for Teacher of Mathematics [NCTM], 2000). While the use of multiple and varied sources of evidence may be "more likely to yield an accurate picture of what each student knows and is able to do," it makes the grading process "less straightforward" for teachers (NCMT, p. 23).

As the importance of listening to teachers' voice in reform contexts has been strongly advocated (Bailey, 2000; Cheung, 2002), we wished to highlight the personal journey of an individual grade 10 mathematics teacher in order to attempt to better understand the current issues involved in dealing with contemporary and traditional grading methods and ideals within a standards-based educational system. This teacher's account was obtained as part of a larger comparative study where teachers ($n=315$) in two Canadian provinces (Ontario and Saskatchewan) responded to a written survey and where some of them participated in a follow-up interview ($n=12$ in Ontario and $n=5$ in Saskatchewan).²

LITERATURE REVIEW

In what is now considered a pioneering study, Stiggins, Frisbie, and Griswold (1989) looked at the grading practices of 15 experienced secondary school teachers in the United States. They identified 19 recommendations for grading in measurement textbooks, and they found that the teachers followed less

than half of these. Stiggins and colleagues did consider reasons for this, but their emphasis lay more on the need for additional research, and they raised questions relating to a range of issues, including the quality of the data and the use of non-achievement factors in grading (e.g., effort), the relationship between grades and motivation, the place of policy, and the interpretation of grades by stakeholders.

In tracing the history of grading, Brookhart (2004) notes that many of the concerns that were raised in the past about grading, such as the unreliability of percentage systems, the misinterpretations of report cards, and the inclusion of non-achievement factors in grades, continue to resurface in the present. An accumulation of research showed that teachers' grading practices not only diverged from the recommendations of measurement specialists, they also varied considerably across teachers and were sometimes inconsistent even within a teacher's own practice (Anders & Richardson, 1992; Brookhart, 1993; 1994; Cross & Frary, 1996; Friedman & Troug, 1998). McMillan (2001) surveyed teachers in the United States ($n=1,483$) and confirmed that the "hodgepodge" (Brookhart, 1991, p. 36) nature of grades and grading practices had continued into the 21st century.

Duncan and Noonan (2007) built on McMillan's (2001) work and surveyed secondary teachers ($n=513$) within the context of assessment reform in Saskatchewan. Their findings were consistent with previous research in that the teachers' practices varied by subject-area, and non-achievement factors were frequently included in the calculation of students' grades. Other areas of focus in recent research on grading include the relationship between students' characteristics and their grades (Guskey, 2004), the grading process in special education (Guskey & Jung, 2009; Silva, Munk, & Bursuck, 2005), and the fairness of grading practices (Guskey, 2006; Resh & Dalbert, 2007; Zoekler, 2005). Although some of the research discussed here questions the influence of standardization on grading practices, how teachers produce final grades in established standards-based systems has yet to be well explored.

As part of our research, we also examined existing studies on teachers' grading practices in secondary school mathematics. Most of these looked at factors associated with grading (e.g., class size, difficulty level of mathematics course, student ability, role of non-academic variables, semestering, and teacher beliefs) (Bonesronning, 2004; Cicmanec, 1999; Deeter, 2002; Howley, Kusimo, & Parrott, 2000; Resh, 2009). One research team in Ontario looked at the extent to which teachers practice new forms of grading in mathematics when all their students are required to take a mandated large-scale assessment (Suurtamm, 2004; Suurtamm, Koch, & Arden, 2010; Suurtamm, Lawson, & Koch, 2008). Based on the results of a questionnaire administered to grades 7-10 teachers ($n=1,096$), they reported that the teachers had embraced the new ways of approaching mathematical teaching and assessment for learning but not those associated with grading (Suurtam, Koch, & Arden, p. 16).

In many of these studies, grading was examined as a mediating factor (e.g., Trautwein, Lüdtke, March, Köller, & Baumert, 2006), or its concept was relatively loosely defined to include assessment and marking. Resh's (2009) study is perhaps the only one that explicitly looked at what rules or weighted combination of rules teachers applied in determining grade allocation to their students in mathematics, language arts, and sciences. However, it focused particularly on the relative weights teachers attached to performance and to effort, and whether their practice differentiated between weak and strong students. Finally, most reviewed studies adopted initial experimental designs that yielded anecdotal results due to reduced and non-random samples. With its intended qualitative design, this paper adds to the knowledge base developed to date by offering an intimate view of grading as defined earlier. It provides a detailed narrative account of the rules that guided one teacher's allocation of final grades within the reform context and of her thoughts about the issues and tensions that this process entails.

FRAMEWORK

The framework used for this study is three dimensional. It looks at the alignment of grading principles, policies, and practices within a particular educational context.

Grading principles

Of 28 standards for classroom assessment developed by the JCSEE (2003), 20 are listed under the heading of *Grading*. However, their direct relationship to grading, and more specifically to the aggregation process for the purpose of determining the final grade, is not always apparent in the standard statement itself. This is the same for the information found in the *Principles for Fair Assessment in Canada* (Joint Committee, 1993) and in the NCTM (2000) principles document, except for the following: "To the extent possible, achievement, effort, participation, and other behaviours should be graded separately" (Joint Committee, p.11).

Various principles specific to the aggregation process have been promoted by assessment specialists over the past few decades. For example, Brookhart (2004) and O'Connor (2002) recommend calculating the median of a student's summative assessment scores for a reporting period to arrive at a final grade, while Marzano (2000) suggests computing the median across learning expectations. Arter and McTighe (2001) and Arter and Chappuis (2006) present the concept of disjunctive combination to aggregate rubrics derived from ordinal and continuous data. Marzano also proposed the selection of most recent scores or best scores overall (or eyeballing) to arrive at a final grade when dealing with qualitative data. In other words, grades should be reported as levels, medians, or modes as determined by teacher judgement (standards-based approach) rather than as simple means (measurement-based approach).

Grading policies

In addition to the standards-based and measurement-based approaches to grading, Noonan (2002) also refers to policy-based grading practices. The policy-based approach to grading is related to teacher practices that are principally guided by school, district, or ministry policies. The Ontario Ministry of Education (OME) adopted the standardized approach to public education in 1995, at which time it released a number of key policy documents related to classroom assessment and grading (OME, 1999; 2000; 2004). In 2008, OME introduced a draft document titled *Growing Success* which grouped all assessment related policies into a single document. Following a lengthy consultation process, the document was released for implementation in Ontario's elementary and secondary schools in September, 2010 (OME, 2010a).

The various OME documents promote the use of a four-level rubric (i.e., the Achievement Chart) in which level 3 is the provincial standard. The purpose of the Achievement Chart is to "enable teachers to make judgements about student work that are based on clear performance standards" (OME, 2008, p. 5). The Achievement Chart in the grades 9 and 10 Mathematics Curriculum contains four categories of knowledge and skills: a) Knowledge and Understanding, b) Thinking, c) Communication, and d) Application (OME, 2005). Teachers are encouraged to assess student learning using all four categories and levels from the Achievement Chart, but they are required to report a single, numerical grade in the form of a percentage for each course, and for each strand in the case of mathematics, on the secondary provincial report card. Although the provincial report cards were recently revised (also for implementation in September 2010), the new secondary report card contains the same basic elements with percentage grades being accompanied by comments and an evaluation of learning skills (or work habits) on a 4-point scale (Excellent, Good, Satisfactory, Needs Improvement).

Grading practices

The study focuses specifically on the rules and process teachers actually follow and apply to combine assessment results for the final grade that will appear on a high school student's report card. This process takes into account the number (e.g., mark book columns) and types of marks to be combined (alpha, narrative or numerical), their relative weights, and what each mark represents. Although all of these approaches are plausible and recommended by respectable experts or organizations, they present more choice than consensus, which may be contributing to the inconsistency of teachers' grading practices. Furthermore, teachers may be unconsciously or unknowingly struggling between the familiar measurement-based approach, the dictated policy-based approach, and the emergent standards-based approach. This single case narrative study sheds light on the complexity of grading practices.

METHOD

The case presented here is bound by space and time as it is situated in Ontario's accountability-driven system and the participant's account was given between March and June, 2007. It can be described as a "particularistic" case (Merriam, 1998, p. 29) because it concerns a specific aspect of the summative assessment process, and also an "instrumental" case (Stake, 2005, p. 445) because it goes beyond description to consider the results in relation to existing policies and principles. A single case is presented for the insight it provides about a teacher's interpretation of her grading practices and related issues in a standards-based educational system.

The high school mathematics context

The Ontario secondary Mathematics Curriculum was released by the OME in 1999 and revised in 2005. It provides some information on classroom assessment. The Education Quality and Accountability Office (EQAO), which is an arms-length agency created in 1996, has been responsible for yearly provincial testing since 1997, and it assesses all students in mathematics in grades 3, 6, and 9. Because one of EQAO's goals is capacity building in assessment and grading, assessment materials are provided on the website, and teachers are invited to participate in the development and scoring of the yearly assessments. Curriculum documents are revised on a regular cycle in Ontario (OME, 2010b), and at the time of this study the OME was reviewing its assessment policies.

The participant

The teacher, Anne (a pseudonym is used to ensure anonymity), was from an urban, French-language school located within a French-language public school board in the northern region of Ontario. She was one of the first participants who volunteered to be interviewed. At the time of the interview, Anne had been certified to teach in Ontario for five years, and had taught mathematics mainly to grades 9 and 10 students. She was responsible for teaching two grade 10 applied mathematics classes, each with approximately 20 students. This student load was typical of the 12 participating Ontario teachers interviewed for the study. Anne regularly worked and consulted with the Mathematics Department Head in her school. She was also involved in marking the EQAO grade 9 mathematics assessment in 2006.

Although data were available from 12 teachers from the larger comparative study, we purposively selected this teacher for several reasons. First, she captured most of the dilemmas other participants experienced in summarizing different types of assessment data. Second, she clearly articulated key tensions in grading that arise in standardized contexts where accountability pressures may sometimes conflict with the ideals of assessment *for learning*. Third, she also provided extensive accompanying artifacts to support and illustrate her grading practices. This allowed for a full discussion of sufficient key issues

associated with current grading practices in an era of accountability and of success for all. The findings related to the 12 teachers regarding grade allocation and more are the topic of a separate paper.

Data collection

Multiple sources of information were used for the study, including the teacher's survey responses, a semi-structured interview, and relevant related documentation. Following a review process and discussions with members of the research team and colleagues, the planned interview protocol was reduced from eight to four key questions (see Appendix A). The first two questions (What grading process do you go through for one student to calculate the final mathematics grade for the report card? Can you describe that process using a clear example from start to finish?) are key for this paper. Other informal interview questions required the participants to provide information about the current number of grade 10 mathematics courses under their responsibility and about the number of students currently enrolled in those courses. The interview also provided an opportunity to clarify some of the teacher's responses to survey questions. Participants received the questions prior to the interview by email. The interview with Anne took place at her school and lasted 38 minutes. It was audio-recorded and transcribed verbatim. Follow-up e-mails were used for clarification questions and member checking.

All relevant documents supporting the grading process were either provided by the teacher or the school, or obtained from the school board and the OME websites. Anne offered copies of the following: a) a four-page mathematics course syllabus, b) the school's assessment guidelines, c) the OME's achievement chart, d) a checklist with six homework assignments, e) a copy of the electronic log sheet, and f) the OME assessment guidelines and Achievement Chart. Documents a, b, and c were part of a kit that was sent home to parents in Anne's school board.

Data analysis

In order to ensure dependability of the results, the principal researcher and two graduate students independently analyzed Anne's interview transcript, sometimes referring to the audio file to get all relevant subtleties. Analyses were based on combined information provided by the interview and survey questions. They first consisted of a breakdown of the transcripts into responses to each individual interview questions. Then they were subjected to a coding system that relied on key concepts within the questions, such as the number of columns in mark log sheet, the nature of the marks (percentages, levels, letters, etc), the weighting system, the treatment of the four categories, the role of assignment tasks (homework, tests, quizzes, examination, project, etc.), the actual aggregation methods (mean, median, eyeballing, etc.), and the underlying principles and policies. These codes were provided in a spreadsheet with

additional space for emerging themes. Contextual details were drawn from accompanying policy documents and the teacher's marks log. These documents were also used to triangulate some details with the teacher's interview and survey responses.

Limitations and benefits

As with any type of research, the single case study has both limitations and benefits. Case studies have been criticized because they do not generalize to a larger population, they can present a biased perspective, and they can be lengthy to produce and read (Merriam, 1998; Yin, 2008). However, by clearly situating a case in a particular context, the transferability of the results may be revealed (Patton, 2002). Case studies can be used as the building blocks for meta-evaluations, allow for greater focus, and can also provide valuable insight for understanding the results of larger-scale studies. For example, Hamilton and colleagues (2003) noted the important role of case studies in their discussion about the challenges they faced as they evaluated the impact of large-scale, standards-based educational reform in the United States. Some strategies that are recommended in methodological texts for case study designs (Merriam, 1998; Yin, 2008), such as saturation and repeated observations across time, were not used for this case because it was drawn from a larger comparative study. However, to ensure the quality of the results, the collection and thorough analysis of supporting documents compensated for the relatively brief interviewing process. The research team also engaged in triangulation, systematic coding procedures, and discussions during the analysis to reduce the possibility of bias and present a balanced and credible interpretation.

RESULTS

The results are presented according to four themes arising from Anne's interview. These are: a) type and aggregation of marks, b) written information given to students on final grades, c) grading principles or guidelines in effect, and d) other specific grading issues. Each is closely examined in the following sections.

Type and aggregation of marks

As mentioned earlier, the first two interview questions invited the participants to explicitly describe the rules and process used to calculate the final grade and to support these with a clear example and artifacts. Anne referred to the grade book in which she had recorded her students' summative assessment results as percentages. Her grade book had a row for each student and five columns. The first two columns reported data from the first half of the semester, which represented 35% of the final grade. The second two columns showed results from the second half of the semester, also worth 35%. Each of the four columns was associated with a mathematics strand (e.g., algebra)

result and all four were given equal weight. These two sets of columns formed 70% of the final grade. A fifth column represented the final examination and was worth 30%. The simple addition of these two percentages provided the final grade for the report card.

Next, Anne described how she used to assign levels to student assignments based on teacher-made rubrics inspired by the provincial Achievement Chart. She would then use her school board's chart to transform the rubric levels to percentages (see Table 1). Eventually, however, Anne was told by both her school's principal and the Mathematics Department Head to revert to the traditional use of percentages throughout the grading process. In that respect, she explained that, "Before I only used the rubric, no percentages, and assigned a level, 4, 4- or 4++. . . Now I give only percentages."³ Further conversation with Anne revealed that she was very much aware of inconsistencies across teachers' grading practices, and that she tried to exercise her own professional judgement when grading.

TABLE 1. Scale for converting from rubric levels to percentages for report cards

0	R-	R	1	1+	2	2+	3	3+	4-	4	4+	4++
0	1.34	35.49	50.54	55.59	60.64	65.69	70.74	75.79	80.84	85.89	90.94	95.100

Written information given to students on final grades

Referring to the grade 10 mathematics course syllabus, Anne explained how the students were informed, in writing, of the grading system. At the beginning of the course, each student received information on late assignments and on the OME requirement to assess all four Achievement Chart categories. In the same document, students received a copy of the scale used to convert levels to percentages (see Table 1) and an explanation of the weighting system adopted by the school (first half of the semester = 35%; second half = 35% and exam = 30%). There was also a note explaining that student progress was taken into consideration when determining the final grade. Finally, students were told that an assessment result was provided on the report card for each of the following work habits: use of oral language (French), works independently, teamwork, organization, work habits/homework, and initiative.⁴ Students and their parents were expected to sign the document.

Although the written information that students received did not explicitly indicate whether homework was considered in the final grade, Anne hinted to her students that it was in order to motivate them. She recalled how she "threatened" the students to do their homework: "If you miss many homework assignments, chances are you will fail.' When I tell them that, they [the students] think that there is a policy on homework."

Grading principles and policies in effect

Throughout the interview, and more specifically in response to the question addressing grading principles, Anne referred to a key provincial policy that prescribed a 70% - 30% split between student products submitted throughout the semester and the final summative assessment. The latter may be a combination of what the OME calls a “culminating task” plus a formal examination, or it may consist of only the examination. A school guideline further dictated that the semester be divided in half, each worth 35%. Here, Anne saw a problem because this provincial policy clashed with another one stating that teachers should take into account the most recent and most consistent work produced by the student (OME, 2000). She explained:

Last year we made an error, ah, I'm not sure if it's an error but we counted the first half on 35, the second also on 35 and then 30 % for the culminating task. In separating the session like this, we don't get a good assessment of the student because the first grade is set as soon as it is out there. So given that the Ministry wants to see progress, the mark should be spread out over the semester and be worth 70 %.

As she understood it, Anne felt she had to assign a level to students' work based on the OME Achievement Chart, collect a number of levels across the term, calculate the final level, and then use the school board's chart to convert the final level into a percentage for the provincial report card. In her teacher education program, Anne was trained to develop and apply adapted rubrics, but she found this problematic within the school system. She described her dilemma as follows:

Really, for me, unfortunately, the rubric means very little for me. I use it and I attach it to the student's work but the parents here don't like it. They hate it because it means absolutely nothing to them. And the problem is when I used it religiously. The problem was that a 4 had such a range that to get a 4 meant between 80 and 100 but a 3 was between 70 and 80. It's the first year that I use the rubric very little. Before, I used only the rubric, no points. But I had too many problems with the parents. Try to explain that; it's incredible! Try to tell them that a 3, sir, is the provincial norm, listen, this is good. It just didn't go through. They have trouble with the fact that a 76 is the norm. So I have lots of trouble with this, the school principal has even more trouble with explaining this process. That's why they asked me to go with percentages.

Finally, Anne also knew that she had to produce a single percentage for each mathematics strand on the report card, according to the OME guidelines, and that this percentage had to reflect the four categories in the provincial Achievement Chart. She felt she was fortunate to have been able to attend a grading session at the EQAO, but she was still confused about her interpretation of the four categories and how to find balance among them when assessing and grading.

Other grading issues

In response to the final interview question, Anne expressed concerns and confusion about two other key issues: borderline cases and marking homework. Considering borderline cases, Anne interpreted her school board policy to mean that a grade below 45 could not be assigned in the first half of the session. She explained:

We can't give a grade between 45 and 50. If the final grade is 47, then you have to decide whether the student fails by assigning a 45% or passes with a 50%. However, to show that the 50% was not granted only to pass the student, then we normally would assign 53%.

She was essentially concerned about higher achieving students not being given the same advantage as students with borderline grades, which she considered to be unfair. Anne also expressed uncertainty about grading homework, and struggled to find a way to incorporate homework in the final grade rather than just filling the separate section on the provincial report card for work habits/homework. She stated that:

I haven't been assessing homework because I don't have a rubric for that.... What I would like to do is to be able to count the number of homework assignments done on the total number of homework assignments (about 50) and integrate this score in the final grade.

She proposed to grade homework based on the amount of homework completed rather than on its quality, but was unsure about proceeding without a rubric.

DISCUSSION AND IMPLICATIONS

This contextually specific and detailed examination of a teacher's grading practices is revealing because it provides an example of how various principles and policies are translated into practice. It also offers interpretive insight into understanding the various pressures, personal views, and administrative constraints that may sometimes clash with or contradict the ideals imposed within a standards-based context.

The concerns Anne voiced regarding grading practices can be grouped into at least four categories for discussion. First, one source of tension resulted in part from conflicting policy guidelines at various system levels. For example, Anne's aggregation of marks was accomplished by combining 70% for the semester assignments and 30% for the final examination, which conformed to the OME policy. However, the school board's directive to further divide the 70% into two set percentages conflicted with another provincial policy stating that the "most consistent level" and "more recent evidence" of student achievement should be used to calculate grades (OME, 2000, p.15). Also, the school administration's directive to use percentage throughout the term proved for

Anne to be inconsistent with the OME guidelines and her sense of integrity. Suurtamm (2004) highlighted the need for administrative support to help teachers develop grading practices that are compatible with newer assessment methods and communicate about grading with parents and students.

Second, some grading principles/policies may not easily apply to the context of mathematics. For example, although “the most recent, most consistent” policy reflects current assessment principles that foster learning (Marzano, 2000; O’Connor, 2002), in practice it may be problematic for subjects such as mathematics that are structured around content rather than skills. The mathematics curriculum presents content and skills under strands such as Trigonometry and Quadratic Relations (OME, 2005), and these are still typically taught independently rather than on an integrated basis, thus making it difficult to consider the most recent evidence. In such circumstances, teachers may react differently. Some may consider the most recent and consistent evidence within each strand, some may consider it for each of the four Achievement Chart categories across the strands, and others may disregard the policy and average all of a student’s assessment results to produce a final grade. Although this policy is based on a relatively recent grading principle, it has been retained in the most current version of *Growing Success* (OME, 2010a). As such, future studies should delve further into this area of practice to better understand how this policy is applied by teachers across levels and subjects.

The third source of tension involves compliance to grading policies that contradict standards-based principles. For example, Anne had to convert rubric levels to report the students’ final grades as percentages as mandated by OME. This type of issue may be reflective of the policy-based approach to grading at play in Ontario. It clearly shows evidence of conflict between the OME’s need to bide by the traditional and popular use of percentages while simultaneously incorporating the recent standard-based reforms in assessment such as the use of rubrics. Suurtamm (2004) observed a similar tension in the case studies she undertook with five secondary mathematics teachers in Ontario, and she explained that:

Several of the teachers discussed the difficulty of matching authentic assessment techniques with a more traditional method of reporting using percentage marks. The data gathered through authentic assessment frequently consists of levels on a rating scale or rubric or is anecdotal. This type of information is unsuited to being directly translated into a percentage mark. (p. 506)

These concerns support Brookhart’s (2004) conclusion regarding the pervasiveness of percentages despite the standards-based movement, and suggest the need for further inquiry into the impact of adopting varying scales on the perceptions, acceptance, and use of rubrics and performance levels by various stakeholders in order to counterbalance the community’s perceived need for the use of percentages.

The fourth and final discussion theme stemming from Anne's experience relates to the role of non-academic variables such as effort, homework, and late assignments in calculating final grades. Anne believed that homework should be incorporated within the overall grade but was unsure of how to proceed. On the secondary provincial report card, students' work habits should be reported in a separate section for learning skills.⁵ The issue of non-achievement factors has long been evident in research on teachers' grading practices, especially in relation to borderline cases (Cicmanec, 1999; Howley et al, 2000; Marzano, 2000; McMillan, 2001, 2008; Resh, 2009; Rich, 2002). This seems to be particularly true when a standards-based approach is adopted. As Duncan and Noonan (2007) have shown, teachers often consider non-achievement factors in adjusting their students' final achievement grades. In theory, some non-achievement factors should not influence grades (e.g., students' characteristics such as gender), but others are considered to be academic enablers (i.e., work habits such as completion of homework, late assignments, etc.). Research suggests that students will complete their homework, whether it is graded or not, if it is closely tied to the learning objectives (Scriffiny, 2008). Given the importance that teachers, parents and students attribute to these enablers, their role within assessment practices should be investigated further in different educational contexts.

In the context of accountability and responsibility, the need to regularly assess on a large-scale basis while advocating reform may undermine smooth transition in teaching and assessment practices at the classroom level. Whereas the OME promotes a mathematics curriculum that values professional judgement, the EQAO which assesses all grade 9 students across the province every year tends to adhere to the traditional, measurement approach to develop its assessments (Suurtamm et al. 2008). As both the OME's and EQAO's given goals, objectives, and purposes are to directly impact teaching and learning, these often unintended, conflicting underlying assessment models become a significant source of dilemma for those teachers who believe in the visions of their educational contexts and who wish to assimilate their respective demands.

CONCLUSION

This paper provides meaningful information regarding the rules and process a secondary teacher experienced while calculating final grades for standardized report cards and the difficulties and tensions that this ensued. Such data would not be captured by a broad survey or large-scale questionnaire. Despite the fact that it is based on a single participant and is specific to Ontario's educational context, Anne's case clearly illustrates some of the tensions that characterize the implementation of measurement principles and policy guidelines in a standardized educational system. In particular, it highlights the difficulty of reporting percentage grades in a system that promotes the use of achievement levels (rubrics) for assessment. It also shows the misalign-

ment that can happen at the various administrative levels (i.e., mathematics department, school, school board, and OME). To some extent, this tension is consistent with observations made by Brookhart (2004), McMillan (2001), Suurtamm (2004), and Suurtamm and colleagues (2008, 2010). The teacher in this study was also very much aware of inequalities and inconsistencies of grading practices among teachers, despite standardization, and she tried to exercise her professional judgement in adopting the provincial guidelines. She also felt torn between ideals and their potential articulation in practice, such as how to treat non-achievement factors, including homework, in the calculation of the final grade. Finally, Anne struggled with fairness issues resulting from the distortion of policies, such as the imposition of a minimum grade (40%) on all students, which favours students with borderline assessment results.

The issues emanating from this teacher's account form the basis for meaningful areas of research, particularly with respect to the grading process within a standards-based approach where higher order skills are targeted, where assessments hold the dual goal of accountability and assessment for learning, and where the grading process incorporates both qualitative and quantitative assessment data. Although the teacher in this case was not an expert in grading, her account contributes to the knowledge base by revealing specific information about an area of practice that is generally considered "mysterious" (Speck, 1998, p. 17) because it relies on professional judgement. Understanding the issues and dilemmas she experienced while grading lends some transparency to the process, which may ultimately improve the fairness of teachers' grading practices. It also highlights the need for ongoing support for teachers to develop the knowledge and understanding of grading principles and policies that should inform their professional judgement. This may be especially important in educational systems where assessment policies change through continuous reform, and teachers and school leaders are expected to adapt their practices accordingly. Our work in this area suggests that the extent to which teachers and school leaders are able to implement evolving assessment principles and policies, and the reasons they may face challenges, should be studied regularly as part of the reform process. At present, it should not be assumed that the grades on students' final report cards are calculated with sufficient consistency to allow meaningful interpretation by all stakeholders.

NOTES

1. This paper reports a part of a study that was funded by the Social Sciences and Humanities Research Council standard program, Canada.
2. The results of the interviews with the 17 participants will be published in a separate paper.
3. All quotations from Anne have been translated from French to English.
4. At the time of this study, these were the six learning skills included on the provincial report card for secondary students in French-language school boards in Ontario.

5. The new assessment policies and standardized report cards (September 2010) in Ontario continue to mandate the separation of achievement and learning skills/work habits (see *Growing Success*, Ontario Ministry of Education, 2010a, p.10).

REFERENCES

- Anders, P., & Richardson, V. (1992). Teacher as game-show host, bookkeeper, or judge? Challenges, contradictions, and consequences of accountability. *Teachers College Record*, 94(2), 382-397.
- Arter, J., & Chappuis, J. (2006). *Creating & recognizing quality rubrics*. Toronto: Allyn & Bacon
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press, Inc.
- Bailey, B. (2000). The impact of mandated change on teachers. In N. Bascia & A. Hargreaves (Eds.), *The sharp edge of educational change*. London: RoutledgeFalmer.
- Bonesronning, H. (2004). Do the teachers' grading practices affect student achievement? *Education Economics*, 12(2), 151-167.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, Spring, 35-36.
- Brookhart, S. M. (1993). Teacher' grading: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.
- Brookhart, S. M. (2004). *Grading*. Upper Saddle River, NJ: Pearson Education.
- Cheung, D. (2002). Refining a stage model for studying teacher concerns about educational innovations. *Australia Journal of Education*, 46(3), 305-322.
- Cicmanec, K. B. (1999). *High school mathematics teachers: Grading practice and pupil control ideology*. Unpublished doctoral dissertation, Ohio University. UMI # AAT 9956769.
- Cross, L. H., & Frary, R. B. (1996). *Hodgepodge grading: Endorsed by students and teachers alike*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, April 9-11.
- Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher*, 19(1), 3-13.
- Deeter, B. C. (2002). *A comparison between teachers' assessment, grading and instructional practices on quarter timetable vs. ten-month timetables in British Columbia secondary schools*. Unpublished doctoral dissertation. Simon Fraser University, Canada.
- Duncan, R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *The Alberta Journal of Educational Research*, 53(1), 1-21.
- Friedman, S. J., & Troug, A. J. (1998). Evaluation of high school teachers' written grading policies. *ERS Spectrum*, 17(3), 34-42.
- Guskey, T. R. (2004). *Stability and change in high school grades*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Guskey, T. R. (2006). "It wasn't fair!" *Educators' recollections of their experiences as students with grading*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Guskey, T. R., & Jung, L. A. (2009). Grading and reporting in a standards-based environment: Implications for students with special needs. *Theory into Practice*, 48, 53-62.
- Hanushek, E. A., & Raymond, M. E. (2003). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, 2 (2-3), 406-415.
- Hargreaves, A. (2002). Sustainability of educational change: The role of social geographies. *Journal of Educational Change*, 3, 189-214.

- Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, 25 (1), 1-29. Retrieved from <http://eeppa.aera.net> on March 1, 2010
- Horn, R. A., Jr. (2004). *Standards*. New York: Peter Lang.
- Howley, A., Kusimo, P. S., & Parrott, L. (2000). Grading and the ethos of effort. *Learning Environments Research*, 3(3), 229-246.
- Joint Committee. (1999). *Principles for fair assessment in Canada*. University of Edmonton.
- Joint Committee on Standards for Educational Evaluation [JCSEE]. (2003) *The student evaluation standards: How to improve evaluations of students*. Washington DC: AERA, APA, NCME.
- Linn, R. L. (2000). Assessment and accountability. *Educational Researcher*, 29(2), 4-16.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria: Association for Supervision and Curriculum Development.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practices*, 20(1), 20-32.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd Ed.). Thousand Oaks: Corwin Press, A SAGE company.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education* (Rev. ed.). San Francisco, CA: Jossey-Bass/John Wiley & Sons.
- National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Noonan, B. (2002) Interpretation panels and collaborative research. *Brock University*, 12(1), 89-100.
- O'Connor, K. (2002). *How to grade for learning*. Glenview, IL: Pearson Education.
- Ontario Ministry of Education. (1999). *Guide to the provincial report card, Grades 9-12*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Ministry of Education.(2000). *The Ontario curriculum Grades 9 to 12: Program planning and assessment*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Ministry of Education (2004). *The Ontario curriculum grades 1 to 12: Achievement charts (Draft)*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Ministry of Education. (2005). *The Ontario curriculum grades 9 and 10: Mathematics*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Ministry of Education (2008). *Growing success: Improving student learning*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Ministry of Education (2010a). *Growing success: Assessment, evaluation and reporting in Ontario schools (First edition, covering grades 1 to 12)*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Ministry of Education (2010b). *Ontario curriculum: Frequently asked questions*. Retrieved from <http://www.edu.gov.on.ca/eng/curriculumsecondary/index.html#role>, on 02 July, 2010.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education*, 12, 315-325.
- Resh, N., & Dalbert, C. (2007). Gender differences in sense of justice about grades: A comparative study of high school students in Israel and Germany. *Teachers College Record* 109(2), 322-342.
- Rich, R. H. (2002). *Hidden factors in teachers' secondary grading practices*. Unpublished doctoral dissertation, Seton Hall University. UMI # AAT 3036925
- Senk, S. L., Beckman, C. H., & Thompson, D. R. (1997). Assessment and grading in high school Mathematics classrooms. *Journal for Research in Mathematics Education*, 28(2), 187-215.

- Scriffiny, P. (2008). Seven reasons for standards-based grading, *Educational Leadership*, 66(2), 70-74.
- Silva, M., Munk, D. D., & Bursuck, W. D. (2005). Grading adaptations for students with disabilities. *Intervention in School and Clinic* 2005, 41, 87.
- Speck, B. W. (1998). Unveiling some of the mystery of professional judgment in classroom assessment. In R. S. Anderson & B. W. Speck (Eds.), *Changing the way we grade student performance: Classroom assessment and the new learning paradigm*, pp.89-96. San Francisco, CA: Jossey-Bass.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed.), pp.443-466. Thousand Oaks, CA: Sage.
- Stiggins, R. J., Frisbie, D. A., Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practices*, 1, 5-14.
- Supovitz, J. A., & Taylor, B. S. (2003). *The impact of standards-based reform in Duval County, Florida: 1999-2002*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania. Retrieved from <http://www.cpre.org/> on 16 June 2010.
- Suurtamm, C. (2004). Developing authentic assessments: Case studies of secondary school mathematics teachers' experiences. *Canadian School of Science, Mathematics and Technology Education*, 4(4), 497-513.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' emerging assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy, and Practice*, 17(4), 399-417.
- Suurtamm, C., Lawson, A., & Koch, M. (2008). The challenge of maintaining the integrity of reform mathematics in large-scale assessment. *Studies in Educational Evaluation*, 34, 31-43.
- Trautwein, U., Lüdtke, Q., Köller, O., Marsh, H. W., & Baument, J. (2006). Tracking, grading, and student motivation: Using group composition status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, 98(4), 788-806.
- Yin, R. K. (2008). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA : Sage.
- Young, J., & Levin, B. (2002). *Understanding Canadian schools: An introduction to educational administration* (3rd Ed.). Scarborough, ON: Nelson/Thomson.
- Zoekler, L. G. (2005). *Moral dimensions of grading in high school English*. Unpublished doctoral dissertation, University of Indiana. UMI No. AAT3183500.

APPENDIX A.

TOWARD A BETTER UNDERSTANDING OF GRADING PRACTICES WITHIN ASSESSMENT-DRIVEN CONTEXTS, MAY 2007

SEMI-STRUCTURED INTERVIEW QUESTIONS

The purpose of the interview is threefold: a) to obtain actual examples of the grading processes teachers use to calculate the final grade for the report card, b) to further question teachers on their knowledge and familiarity with grading principles as found in the literature and c) to clarify information in light of the data obtained from the written questionnaire.

Questions of interest:

1. What grading process do you go through for one student to calculate the final Mathematics grade for the report card? Can you describe that process using a clear example, from start to finish? If possible, provide all relevant documents.
2. What instructions do you provide to students regarding the grading of their work in Mathematics?
3. Do you follow any specific grading principles when determining the final grade in Mathematics? If so, which one(s)?
4. Are there grading practices policies or principles that you think should be a) emphasized? b) de-emphasized? Which ones?

MARIELLE SIMON is full professor at the Faculty of Education. Her teaching and research interests include classroom assessment, large-scale assessment, assessment rubrics, secondary data analyses, grading issues and policies.

ROBIN D. TIERNEY was an elementary teacher in Ontario and a research assistant at the Faculty of Education, University of Ottawa. She graduated with a Doctorate in Philosophy (Education) in 2010. Her doctoral dissertation draws on teachers' practical wisdom (phronesis) to understand fairness in classroom assessment. She now lives in California.

RENÉE FORGETTE-GIROUX was full professor at the Faculty of Education at the time of the study and was one of the principal investigators. Her work is mainly in the area of evaluation, assessment, research methods and statistics.

JULIE CHARLAND is currently Principal of a French language secondary school in Eastern Ontario. She was a doctoral student at the time of this study. Her research focuses on the assessment of school principals.

BRIAN NOONAN was associate professor in the Department of Educational Psychology and Special Education at the time of the study. He was also one of the principal investigators of the study.

RANDY DUNCAN studied in the Department of Educational Psychology and Special Education, College of Education and obtained his PhD degree in measurement and education in 2009. He was involved in this project from beginning to the end.

MARIELLE SIMON est professeur titulaire à la Faculté des sciences de l'éducation de l'Université d'Ottawa. Ses intérêts d'enseignement et de recherche englobent l'évaluation en contexte de la salle de classe, l'évaluation à grande échelle, les grilles d'évaluation, l'analyse des données secondaires ainsi que les problématiques et les politiques de notation.

ROBIN D. TIERNEY a enseigné à l'élémentaire en Ontario et travaillé comme assistante de recherche à la Faculté des sciences de l'éducation de l'Université d'Ottawa. Elle a reçu son diplôme de doctorat en philosophie de l'éducation en 2010. Sa thèse de doctorat analyse le discernement et la sagesse pratique des enseignants en classe (phronésis) pour expliquer le phénomène de justice dans l'évaluation en classe. Elle habite maintenant en Californie.

RENÉE FORGETTE-GIROUX était professeur titulaire à la Faculté des sciences de l'éducation de l'Université d'Ottawa au moment de l'étude et en est un des principaux investigateurs. Sa démarche de recherche couvre principalement l'évaluation, l'appréciation, les méthodes de recherche et les statistiques.

JULIE CHARLAND est présentement directrice d'une école secondaire francophone dans l'est de l'Ontario. Elle était une étudiante au doctorat au cours de l'étude. Ses recherches portent sur l'évaluation des directeurs d'école.

BRIAN NOONAN était professeur agrégé au département d'Educational Psychology and Special Education de l'Université de la Saskatchewan lors de l'étude. Il est également un des principaux investigateurs de la recherche.

RANDY DUNCAN a étudié au département d'Educational Psychology and Special Education du College of Education de l'Université de la Saskatchewan. Il a reçu son doctorat en mesure et éducation en 2009. Randy Duncan a été impliqué dans la réalisation du projet du début à la fin.

