The Educational Forum, 75: 210–27, 2011

ISSN: 0013-1725 print/1938-8098 online
DOI: 10.1080/00131725.2011.577669

THE EDUCATIONAL Forum

Being Fair: Teachers' Interpretations of Principles for Standards-Based Grading

Robin D. Tierney, Marielle Simon, and Julie Charland
Measurement, Evaluation, and Assessment Research Unit, University of Ottawa, Ottawa, Canada

Abstract
Knowing that grades can have long-term consequences for students, teachers voice concern about being fair in the grading process. However, their interpretations of fairness are varied and sometimes contradictory. This study looked at how teachers in one standards-based educational system determined secondary students' grades, focusing specifically on the extent to which they followed a specific set of principles for grading. The results support previous research, and suggest that a better understanding of essential principles is needed for grades to accurately reflect students' achievement.

Key words: *academic achievement, classroom assessment, educational standards, evaluation, fairness, grades, grading, report cards, secondary education, teachers, testing.*

Teachers' decisions can have long-lasting social, emotional, and academic consequences for students. The imperative that this provides for understanding teachers' assessment practices was repeatedly noted more than one decade ago (Stiggins, Frisbie, and Griswold 1989; Brookhart 1994; Delandshere and Jones 1999; Gipps 1999). It is clear that teachers are concerned about making fair assessment decisions (Brookhart 1993; McMillan, 2003; Eggen 2004; Rex 2005; Yip and Cheung 2005), but their interpretations of fairness vary greatly (Yung 2001; Zoeckler 2005). However, policies relating to classroom assessment have been changing in many educational systems as a result of standards-based reform. As a result, ongoing inquiry is needed to understand classroom practices, and particularly the basis of teachers' assessment decisions in changing educational contexts.

This study looked at how teachers determine students' grades in one standards-based system. Standards-based educational systems usually have centrally developed

Address correspondence to Robin D. Tierney, Measurement, Evaluation, and Assessment Research Unit, 376 D Lamoureux Hall, Faculty of Education, University of Ottawa, 145 Jean Jacques Lussier, Ottawa, Canada K1N 6N5. E-mail: robin.tierney@comcast.net

curriculum and a common report card for communication about student achievement. The purpose of grading in standards-based systems is to "compare student performance to established levels of proficiency in knowledge, understanding, and skills" (McMillan 2009, 108). This is different from norm-referenced grading where the purpose is to rank students, or self-referenced grading where the purpose is to support individual learning (Brookhart 2004). In using the term *grading,* we refer to a process within the practice of classroom assessment, specifically at the point when teachers calculate students' grades for standardized report cards. In theory, standards-based grading is fairer for students' report cards because achievement is more accurately represented in relation to learning expectations or standards. However, Guskey (2009, 2) noted that no challenge has been "thornier or more vexing than grading and reporting" in standards-based reform.

Some grading issues are remarkably persistent in that they can be traced back through educational contexts for decades (Brookhart 2004). At the end of the 20th century, concern about grading was heightened because of the lack of congruence between teachers' practices and measurement theory. Stiggins, Frisbie, and Griswold (1989) found that 15 experienced secondary teachers followed less than one-half of 19 recommendations for grading identified in measurement textbooks. Researchers began referring to the "hodgepodge" (Brookhart 1991, 36) nature of grades, and an accumulation of evidence showed that teachers' grading practices not only diverged from measurement theory, but also varied considerably among teachers, and were sometimes inconsistent even within a teacher's own practice (Anders and Richardson 1992; Brookhart 1993; Cross and Frary 1996; Howley, Kusimo, and Parrott 2000). In a survey of U.S. teachers ($N = 1{,}483$), McMillan (2001) confirmed that hodgepodge grading had continued into the 21st century. Building on McMillan's (2001) work, Duncan and Noonan (2007) 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 (e.g., effort) were frequently included in the calculation of students' grades. Recent research on grading has also focused on the relationship between students' characteristics and their grades (Guskey 2005) and the effect of grading (Bonesronning 2004), the grading process in special education (Silva, Munk, and Bursuck 2005; Guskey and Jung 2009), the fairness of grading practices (Zoeckler 2005; Guskey 2006), and particularly students' perceptions of fair grading (Bursuck, Munk, and Olson 1999; Dalbert, Schneidewind, and Saalbach 2007; Resh and Dalbert 2007).

The researchers built on this existing body of research by focusing specifically on the extent to which teachers follow current principles that have been established for grading in standards-based systems. As the terms *principle* and *policy* are often used interchangeably in educational literature, a distinction was made in this work. A policy is a "plan of action adopted by an individual or social group; a line of argument rationalizing the course of action of a government" (WordNet 2006). In contrast, a principle is a "basic generalization that is accepted as true and that can be used as a basis for reasoning or conduct; a basic truth or law or assumption; an explanation of the fundamental reasons" (WordNet 2006). Policies are considered to be directives given by organizations with the authority to mandate what *must* be done in practice, whereas principles are abstract guidelines about what *should* be done in practice. Principles are generated from theory that has been

generally accepted within a given community and, thus, are often stated by committees. For example, the principle that "summary comments and grades should be based on more than one assessment result" (Joint Advisory Committee [JAC] 1993, 11) is specified in an assessment policy document as "seventy percent of the grade will be based on evaluation conducted throughout the course … thirty percent of the grade will be based on a final evaluation" (Ontario Ministry of Education [OME] 2010a, 41). Although policies usually give specific direction for practice in a particular educational context, ideally they are based on the principles accepted by the broader educational community.

Framework

To develop a framework for this study, the authors looked at guidance for determining students' grades in five sources. Two of these, *The Principles for Fair Student Assessment Practices in Canada* (JAC 1993) and *The Student Evaluation Standards* (Joint Committee on Standards for Educational Evaluation 2003) were produced with representatives from a variety of educational organizations. Although standards for educational and psychological testing were developed much earlier (American Psychological Association 1954), these were the first joint committee documents that offered guidance specifically for classroom assessment. Guidelines offered for grading in three textbooks written for teachers by classroom assessment specialists were also considered (O'Connor 2002; Brookhart 2004; McMillan 2004). A synthesized set of four essential principles for grading in a standards-based system was developed:

1. When the purpose of grading is to report on student achievement, grades should be referenced to the curriculum objectives or learning expectations (criterion referenced).
2. A grade should be an accurate representation of achievement, so non-achievement factors should be reported separately to permit valid interpretation by stakeholders.
3. Results from multiple assessments should be combined carefully, with weighting that reflects the learning expectations, to ensure that the grade accurately summarizes achievement.
4. Information about grading should be clearly communicated so that grades are justified and their meaning is understood by students, parents, and other teachers.

These four principles informed the development of the data collection tools, and they frame the analysis and discussion for this paper.

Methodology

This work is part of a nationally funded comparative study on teachers' grading practices in the provinces of Saskatchewan and Ontario, Canada. This project aimed to determine how teachers (N = 315) calculated students' final report card grades in two educational systems with differing assessment policies.

Data Sources and Analysis

This paper focuses specifically on a sub-sample of teachers (n = 77) who taught Grade-10 mathematics in publicly funded, English-language schools in Ontario between

2006 and 2008. Their teaching experience at that time ranged from less than 1 year to 31 years (M = 12 years). They responded to a survey developed for the larger study, which contained a total of 47 items (Likert-type, checklist, and open-ended) relating to teachers' demographics, their grading practices, and their awareness and use of grading principles and policies. Five teachers in the Ontario English-language sample also volunteered to participate in semi-structured interviews upon completion of the survey.

Two survey items and two interview questions asked teachers explicitly about their awareness and use of grading principles. The extent to which teachers follow the four essential principles was inferred from selected responses on the survey items and through follow-up interview questions. Statistical software was used for quantitative analysis where appropriate. A coding scheme based on the four grading principles was used for the open-ended survey responses and interview transcripts, with the addition of an "other" category to allow for the emergence of issues that might not be captured with a purely deductive approach.

Educational Context

Ontario is a large province with over two million students in 72 publicly funded district school boards (districts). Students in Grade 10 are required to take mathematics at either the academic or applied levels, and these lead to university or workplace preparation courses in senior years. Teachers in Ontario are certified by the Ontario College of Teachers [OCT] (2008), and teaching mathematics at the secondary level requires an undergraduate degree with courses in mathematics and successful completion of a teacher education program. Based on enrolment statistics for the province, more than 5,000 teachers in total teach Grade-10 mathematics per year (OME 2010b).

The OME began the process of standardization in the mid-1990s with the development of new curricula and the creation of an assessment agency, the Education Quality and Accountability Office (EQAO). Assessments of all students in Grades 3 and 6 (reading, writing, and mathematics), Grade 9 (mathematics), and Grade 10 (literacy) are now conducted yearly by the EQAO. The Grade-10 literacy test is the only assessment that has high stakes for students in that a passing score is required for graduation. Standards-based reform is ongoing with curricular revisions conducted in a regular cycle. Mathematics curriculum for Grades 9 and 10 was released in l999 and revised in 2005. All curricular documents include learning expectations and achievement charts, which are four-level, generic rubrics designed to guide classroom assessment. The Mathematics Achievement Chart for Grades 9 and 10 has four categories: (1) knowledge and understanding, (2) thinking, (3) communication, and (4) application (OME 2005).

A standardized report card was phased in across secondary grades from 2000 to 2003, and it was revised again for implementation in September 2010. When this study began, information about classroom assessment was scattered through several OME documents. Following a consultative process with stakeholders, assessment policies were integrated into one document, *Growing Success,* which was released in May 2010 (OME 2010a). In summary, student assessment in Ontario largely depends on teachers' professional judgment within a framework that has gradually been standardized with achievement

charts, large-scale assessments, common report cards, and increasingly explicit assessment policies.

Results

The results are presented in five sections. The first section describes the awareness and use of grading principles reported by the participants. The subsequent sections look at indications in their responses about the extent to which they follow the four grading principles.

Explicit Awareness and Use of Grading Principles

Many of the participants' responses suggested that their assessment practices were guided by pedagogical content knowledge gained over their teaching careers. This was particularly evident as they explained their reasons for specific practices, such as why a rubric needed to be adapted for observations during one-hour classes, or why certain categories on the achievement chart were most relevant for assessing a particular strand of mathematics. A different view of the participants' knowledge emerged when they were asked to rate their awareness of grading principles (or measurement and evaluation theories). Less than one-third reported considerable awareness (28.6 percent), and the largest percentage indicated some degree of awareness (40.3 percent). The balance rated their awareness of grading principles as either small (16.9 percent) or none (13.0 percent). A similar pattern was seen in their use of grading principles for calculating grades. Some agreed that they followed grading principles (23.4 percent), and the largest percentage somewhat agreed (42.9 percent). The balance somewhat disagreed (13.0 percent), disagreed (2.6 percent), or felt that this did not apply to them (10.4 percent). In brief, the percentage of participants who indicated considerable awareness and clear use of grading principles was relatively small. Although some participants had clear reasons for their practices, these results suggested that grading principles did not play a strong role in how students' grades were determined.

When participants were asked directly about grading principles, they responded in two ways. Some referred to principles that were not specifically related to grading, but that guide fair assessment in general terms. For example, the idea that students should have ample opportunity to learn was stated by several participants. One noted that teachers in her school "eat lunch in the classroom where we're helping students … so that's I guess a principle that we make ourselves available so that every opportunity we can provide the means to be successful, we do it." One participant listed eight principles related to purpose, reflection, and collaboration in assessment. While these principles reflected current educational ideals, especially in terms of using assessment for learning, they were not related specifically to the process of grading.

The second way that participants responded when they were asked about grading principles was to refer to grading policies. For example, one participant tepidly agreed that he followed grading principles based on his interpretation of an OME grading policy: "Yeah, kind of I guess. Yeah, the most recent mark counts for the most." This policy states that a "grade should reflect the student's most consistent level of achievement throughout the course, although special consideration should be given to more recent evidence of achievement" (OME 2010a, 41). Policies, or pieces thereof, were taken at face value

as principles, or they were interpreted more intuitively as principles. This was seen in another participant's response to the same question:

> *Grading principles. … Well, to me that just means to be as fair as you can to the student. That's really what it comes down to … if it's the most recent, most consistent, and you knew that they had a really rough start to the semester, then I'd probably lean towards more taking the last few marks that they got.*

Only a few participants openly acknowledged that they were not certain about the difference between policies and principles, but confusion was heard in almost all of their responses.

Frame of Reference for Grading

When the purpose of grading is to report on student achievement, grades should be referenced to the learning expectations (criterion-referenced). To understand the extent to which teachers follow this principle, the responses in Table 1 were used.

Almost all participants agreed or somewhat agreed that the grades they assigned indicated the degree to which students had achieved the learning expectations (94.8 percent). The vast majority also agreed or somewhat agreed their students were able to receive high grades if they met the learning expectations (93.5 percent). These results suggested that the principle of criterion-referenced grading was not controversial among the participants. The same conclusion was supported by the participants' lower rates of agreement for the items about grades following the bell curve (35.1 percent) and grades indicating students' ranking in relation to peers (27.3 percent). A marked difference was seen, however, in the responses about improvement and grades. Most participants indicated that in the process of calculating grades they considered how much a student had

Table 1. Participants' Responses Relating to Frame of Reference (in Percentages)

Variable	Does Not Apply	Disagree	Somewhat Disagree	Somewhat Agree	Agree
Grades indicate degree of achievement	0.0	0.0	3.9	28.6	66.2
Grades follow bell curve	10.4	28.6	23.4	28.6	6.5
Grades indicate ranking	7.8	46.8	15.6	22.1	5.2
Grades reflect improvement	1.3	9.1	13.0	57.1	18.2
High grades for meeting objectives	0.0	0.0	5.2	35.1	58.4

improved during a course (75.3 percent). This meant that, although the idea of criterion-referenced grading generated strong agreement, students' grades were sometimes also based on their improvement (self-referenced).

Further clarification about the frame of reference used for grading was given during the interviews. One participant, who "somewhat agreed" on the survey that students' grades followed the bell curve, explained that "they might or they might not; it depends on the group of kids I have coming in. I don't make them fit; if it fits then it does. If it doesn't, well that's what happens." It may be that other participants interpreted this item in a similar manner, so agreement did not necessarily mean they were using a norm-referenced grading system. However, with almost one-third of the participants voicing some agreement with the statement that the grades they assigned were an indication of ranking in relation to peers, it was clear that comparisons were made. One explanation offered was that "by grade 11 and 12, we're going to be sharing those marks with other institutions, so we have to be on the same playing ground as everybody else." This suggested that teachers may have shifted to a norm-referenced grading system in later secondary years because senior grades were used to select students from a cohort for post-secondary placements.

There was also some indication that the frame of reference might have depended on teachers' expectations for groups of students. For example, one participant noted that academic and applied courses differed "with the way the marks are gathered." Another elaborated further on this point in explaining his grading process for different courses:

> *If it's an academic course, I'd probably put a little bit more on knowledge, understanding, application, and thinking, and a little less on communication … that's just trying to set up the students for success. With the applied students, if you spread it out, that's the best chance for them to succeed. With the academics, they are going off to university, so it's more theory based. So we have to put more weighting on those types of learning*

While these participants were not advocating for self-referenced grading, it was clear that their practices flexed according to what they perceived to be best for different groups of students.

Elements Represented in Grades

A grade should be an accurate representation of achievement, so non-achievement factors should be reported separately to permit valid interpretation by stakeholders. Table 2 shows the survey responses relating to the elements represented in grades.

Most participants reported that they did not consider students' attitude, motivation, or participation in calculating grades (81.8–87.0 percent). Three of the five participants interviewed also pointed out that group work, which was not included on the survey, should not be considered because students' grades should represent individual achievement.

There was less consensus among participants about effort, with one-third reporting that they considered students' effort in calculating grades (32.5 percent). One participant explained: "If a student has struggled but tried very hard, I will raise the mark at

Table 2. Participants' Responses Relating to Elements Represented in Grades (in Percentages)

Variable	Does Not Apply	Disagree (No)	Somewhat Disagree	Somewhat Agree	Agree (Yes)
Zero given for incomplete assignments	2.6	24.7	10.4	26.0	35.1
Zero given for late assignments	10.4	74.0	6.5	6.5	1.3
Attitude included in grade		84.4			14.3
Effort included in grade		66.2			32.5
Motivation included in grade		87.0			11.7
Participation included in grade		81.8			16.9
Other included in grade		88.3			10.4
None of above included in grade		35.1			63.6
Grade lowered for cheating		59.7			37.7
Grade lowered for tardiness		97.4			0.0
Grade lowered for absences		94.8			2.6
Grade lowered for lack of effort		88.3			9.1
Grade lowered for incomplete assignments		48.1			49.4
Grade lowered for late assignments		85.7			11.7
Grade lowered for other		92.2			5.2
Grade lowered for none of above		63.6			33.8

times—by a few percent, maybe 5 percent." Of the participants who provided written suggestions for changing assessment policies, several suggested that effort should be considered as evidence of students' work ethic and included on report cards. There was a notable difference between the percentage of participants who indicated that they *included* effort in grades by *raising* a grade for good effort (32.5 percent) and the percentage who indicated that they *lowered* grades for lack of effort (9.1 percent). One participant wrote that, "Assignments that are not completed are included in their marks—but I would *never* lower a mark based on effort." This suggested that grading was an additive process where effort was taken into account as a means of boosting grades upward.

Although some participants reported that they lowered grades for late assignments (11.7 percent), the most common reason for lowering grades was incomplete assignments (49.4 percent). Two items that asked about the use of zeros in calculating grades also showed a difference in how late and incomplete assignments were treated. Very few participants agreed or somewhat agreed that they included zeros for late assignments (7.8 percent), but over one-half agreed or somewhat agreed that they included zeros for incomplete assignments (61.1 percent). Teachers were discouraged from using zeros or lowering grades, especially for late assignments, in a transitional policy document in Ontario (OME 2008). While these results suggested general compliance regarding late assignments, many of their comments indicated that they did not agree with this policy. One

noted that heavy grade deductions "may not have been fair," but she still felt that some penalty was needed to discourage lateness. Several participants indicated that accepting late assignments without penalty conflicted with their sense of professional responsibility. One commented on the survey:

> *Current System—Failure is not an option. Should Be—You get what you earn. We need to prepare students for the real world, not baby them. We are doing students a grave injustice by not preparing them for life.*

Another participant explained his feelings during the interview:

> *I guess with this policy you have to accept submissions regardless of how late they are. If I don't pay my taxes on time, there's a consequence. You know, there's a penalty to a lot of things. I think if we continue to move toward a system where consequences don't really seem to appear anywhere, then how do you educate good versus not so good decisions. And that's part of our job, right?*

These participants assumed responsibility for teaching not only the subject matter, but also the life skills or habits they believed were important. In sum, the majority of participants reported that they usually focused on academic achievement in calculating students' grades, but many were concerned about the long-term consequences of overlooking non-achievement factors. Furthermore, the determination of grades was often based on the completion of assignments, and they may have also reflected variation in teachers' perceptions of students' effort.

Combining Assessment Results for Grades

Results from multiple assessments should be combined carefully, with weighting that reflects the learning expectations, to ensure that the grade accurately summarizes achievement. To understand the extent to which teachers follow this principle, the authors looked at items related to the process of calculating grades (see Table 3).

Almost all participants agreed or somewhat agreed that they weighted some assessments (e.g., projects, tests, homework) and some skills or categories (e.g., knowledge recall, communication) more than others (89.6 and 93.5 percent). This indicated a fairly strong acceptance of the idea that not all results were of equal importance in calculating grades. What may not have been as clearly understood relates to the alignment of classroom teaching, learning, and assessing. If the opportunity to learn is compromised (e.g., lessons missed due to inclement weather, lock-down, teacher illness), the weight of an assessment should be adjusted. Many participants agreed or somewhat agreed that they considered what was taught in determining weights for grading (67.6 percent). However, almost one-third disagreed, somewhat disagreed, or felt the statement did not apply (29.9 percent), which suggested that they did not recognize this aspect of weighting assessment results.

Of the eight options indicating which assessment results might be dropped in calculating grades, participants most frequently selected none of the above (48.1 percent). This

Table 3. Participants' Responses Relating to Combining Results (in Percentages)

Variable	Does Not Apply	Disagree (No)	Somewhat Disagree	Somewhat Agree	Agree (Yes)
Include only latest results	9.1	67.5	16.9	5.2	0.0
Eliminate atypical results	1.3	22.1	23.4	39.0	13.0
Professional judgment in grading	1.3	1.3	9.1	35.1	49.4
Grades weighted by material taught	13.0	6.5	10.4	32.5	35.1
Some skills/categories weigh more	2.6	2.6	5.2	20.8	68.8
Some forms/types weigh more	0.0	2.6	2.6	19.5	74.0
Drop highest and lowest results		93.5			3.9
Drop highest results only		97.4			2.6
Drop lowest results only		84.4			13.0
Drop any that are inconsistent		70.1			27.3
Drop earliest results		89.6			7.8
Drop other results		83.1			14.3
Drop none of above		49.4			48.1

suggested that almost one-half were including *all* assessment results when they calculated students' grades, which runs counter to the principle of combining results to accurately reflect achievement in relation to learning expectations. Although some participants reported that they dropped inconsistent results (27.3 percent), exactly which results were considered inconsistent was not evident: A few dropped the earliest (7.8 percent), even fewer dropped the highest and lowest (3.9 percent), almost none dropped the highest only (2.6 percent), and some even kept the latest only (5.2 percent). Variation in how grades were calculated may have rested on teachers' interpretations of "most consistent" (OME 2010a, 41). While one interviewee explained that she eliminated the highest and lowest results to obtain the most consistent, two others believed that this was achieved by averaging all results gathered during a course. Understanding the reason for using the most consistent results, which was to produce the *most accurate* representation of a student's learning, may have been lost as teachers devised different procedures for calculating final grades.

Some participants reported dropping the lowest results (13.0 percent), and several explained that they did this when they "feel there is a reason," such as "illness, extended absence," or "personal problems." One participant explained that her department "struggled" in trying to establish procedures for calculating grades. They felt that consistently dropping lowest results would produce inflated grades, but they did not want students penalized by an "off day." Essentially, they were concerned with the "integrity of the message" communicated with grades. Some participants also felt that this was a problem because of the practice in some schools of assigning minimum grades (e.g., all grades of 35 percent and below were recorded as 35 percent). One said, "I don't think that is fair to the student because it's giving them a false sense of accomplishment … and I see these students going off to university and even college and they are failing miserably."

Again, these participants were concerned about the long-term consequences of grading practices.

The participants' explanations about why and how they combined assessment results indicated that they did not always follow a particular formula or grading principle, but relied instead on their understanding of particular students or circumstances in exercising their professional judgment. Most participants agreed to some degree that they use professional judgment in calculating final grades (84.5 percent). However, with one-third not fully agreeing (35.1 percent), it seemed that they were not entirely confident in their use of professional judgment for grading. This may have related to the limited awareness of grading principles reported, but it may also have had something to do with the reform process. Several participants offered reflections similar to this:

> *I appreciate that the Ministry documents are trying to leave the method of grading up to our professional judgment. However, it would have been very useful to have more direction on how to convert my traditional practices to line up with what the ministry expects.*

On the other hand, many of the participants' comments about borderline cases in grading (i.e., where students' grades were close to failing) noted the importance of teachers' professional judgment, suggesting that some students would not be well-served with one-size-fits-all assessment policies. As such, it seemed that while specific direction in the reform process might have helped teachers, some decisions necessarily depended on the development of professional judgment.

One participant in this study demonstrated how professional judgment and technology could work in tandem for grading. She explained all the data analysis options in her marking software, and noted that it was more precise than "eyeballing." Nonetheless, she felt that professional judgment was still needed because students' grades should ultimately "correspond with what [is] going on in the classroom." In contrast, another interview revealed how problematic grading can be when professional judgment involved unexamined assumptions. The following statements were from one participant explaining how she calculated final grades:

> *Say they did get level 4s on their exam, then that means that they should get like an 82, or 90 sorta thing. But if they've done nothing all semester … and haven't been able to show me a level 4, why should I give them a level 4 magically at the end?*

> *I basically look across, what's the most consistent … they couldn't do this, one was a 3 minus, but they could have worked with a partner … and the exam was a 1. To me the exam probably showed their true understanding, so I gave them a 1.*

The interpretation of one student's high results as magical and another student's low results as true understanding raises the question about the principles, or values, at play. While this participant seemed to want students' grades to accurately summarize their achievement, it was not clear that the basis of her professional judgment was sound.

Communication About Grading

Information about grading should be clearly communicated so that grades are justified and their meaning is understood by students, parents, and other teachers. To understand the extent to which teachers follow this principle the authors looked at the responses in Table 4.

In general, it was agreed that grades convey information about students' academic achievements. Almost all participants agreed or somewhat agreed that grades convey information about academic achievement to students (97.4 percent) and to their parents (93.5 percent). Many also felt that grades convey information to other teachers (83.1 percent), but their responses were more dispersed for this item. This may have been partly due to the format of secondary school report cards, where one percentage grade summarized achievement for a course, rather than providing results for each strand (i.e., algebra, geometry). Teachers may find other ways of sharing information about students' achievement more effective. Alternatively, some participants might have disagreed with this item because communication between teachers in their schools was limited.

Almost all participants also agreed or somewhat agreed that they could justify the process that they used to determine students' grades (98.7 percent). This confidence contrasted sharply with the lack of certainty expressed about grading principles. Participants were also less enthusiastic about communicating with parents. Although many "agreed" that they shared information about the grading process with parents (55.8 percent), one-third selected "somewhat agree" (32.5 percent). It may be that participants were taking into account that course outlines, which were commonly identified during the interviews as a vehicle for grading information, were not always delivered to parents by students. These results may have also reflected that secondary school teachers communicated less with parents because students were expected to assume increasing responsibility as learners. However, weaker responses for sharing information, in comparison to justifying grading, suggested that communication about grading by teachers to parents may have been more reactive than proactive.

Almost all participants agreed or somewhat agreed that they shared the process used to determine grades with students (98.7 percent). One participant explained that "surprises

Table 4. Participants' Responses Relating to Communication About Grading (in Percentages)

Variable	Does Not Apply	Disagree (No)	Somewhat Disagree	Somewhat Agree	Agree (Yes)
Grades communicate to student	0.0	0.0	0.0	18.2	79.2
Grades communicate to parents	0.0	1.3	2.6	19.5	74.0
Grades communicate to teachers	2.6	6.5	3.9	23.4	59.7
Can justify grading process	0.0	0.0	0.0	10.4	88.3
Share grading with students	0.0	0.0	1.3	26.0	72.7
Share grading with parents	1.3	2.6	7.8	32.5	55.8

aren't part of the game," so students "need to know exactly what my expectations are." Another made a similar claim, stating that in his class there was "nothing new" by the time students "get to the test." Talking to students about assessment categories, levels or exemplars came up during all the interviews, which provided further evidence that teachers share assessment information with their students. However, the quality of information shared may vary widely. For example, one participant explained that she showed students the "kind of picture I need to have for a level three," but another participant assumed the criteria for a level was "general knowledge," so she did not "really make them aware of that." One benefit of sharing information with students was noted by an experienced teacher who had recently included category weights on tests (e.g., knowledge, communication). He observed that his conversations with students about tests had changed as a result. While he and his students had focused on "that final percentage at the top" in the past, they now discussed strengths and areas for improvement. This highlights that communication about assessment, including test results, can support learning. In conclusion, the principle that information about grading should be clearly communicated generated very little controversy, but some questions remained about the quality or usefulness of information shared with stakeholders.

Summary and Discussion

Grading principles are vetted either by committees that represent the educational community or by specialists in classroom assessment. Ideally, principles should guide the process teachers use to determine students' final grades. Although many teachers in this sample reported at least some awareness and use of grading principles, they had difficulty identifying relevant principles, and the grading practices they reported suggested that underlying principles were not well-understood. The extent to which four essential principles were followed is summarized and discussed in relation to Ontario's educational context and fair assessment in general:

- *Frame of reference*: Teachers agreed with statements about criterion-referenced grading, but they also considered the type of course and individual improvement in grading (self-referenced), and the use of grades to rank students for post-secondary selection (norm-referenced) may have influenced their grading practices.
- *Elements represented*: Grades were often based on work completion, and some reflected teachers' perceptions of students' efforts. Many teachers were concerned about the consequences of overlooking students' work habits (non-achievement factors) in grading.
- *Combining results*: Teachers calculated grades in different ways with many including all results. Most relied on professional judgment, and they believed it was especially necessary for particular situations. However, professional judgment was not always informed by an understanding of underlying principles, especially regarding the alignment of classroom teaching, learning, and assessing.
- *Communication*: Almost all teachers agreed that they could justify their grading process, and most report sharing information about grading with students. However, the quality and usefulness of information shared with all stakeholders was not as clear.

These results were consistent in several ways with another research project in the same educational context. Suurtamm and Graves (2007) studied the implementation of inquiry-oriented curriculum by mathematics teachers in Ontario (N = 1,096), and they found that traditional assessment methods (i.e., paper-and-pencil tests and quizzes) were used more frequently to determine grades in academic rather than applied courses in Grade 10. The teachers in this study also indicated that their grading practices differed for these courses. The authors of this paper question whether these differences were based on assumptions about groups of students, rather than the principles for quality differentiation (Tomlinson 2005). Although teachers may be attempting to support group learning, they may not be avoiding student stereotypes, which is one of the key aspects of fairness for classroom assessment (McMillan 2004; Airasian 2005).

Difficulty with assessment policies has repeatedly been expressed by teachers in Ontario in recent years, not only through research (Hargreaves, Earl, and Schmidt 2002; Suurtamm and Graves 2007), but also in the local media (e.g., Laucius 2009). In this study, teachers were particularly concerned about grading policies related to the separation of achievement and work habits. Teachers seemed to accept professional responsibility for "developing students' potential" (OCT 2008, 9), but many struggled with the idea that good work habits can be encouraged without grading penalties, despite the alternatives that have been suggested by classroom assessment specialists (e.g., Guskey 2000; Brookhart 2004).

Commonsense and real-world examples tend to support the notion that good work habits lead to success. While encouraging students to complete assignments on time may have obvious benefits, ultimately it is more important to orient students toward learning (Shepard 2006). Grading systems should not encourage students to accumulate marks at the expense of learning, and evidence of learning should be evaluated in relation to expectations or standards, rather than the number of assignments completed. Despite the appeal of real-world philosophy, it can lead teachers to interpret assessment policies in such a way that their practices do not effectively support student learning. As such, the authors concur with a conclusion drawn by Suurtamm and Graves (2007, 129) that the "intent of policy may not be clear" if teachers do not understand the underlying philosophy. Without a better understanding of the most essential grading principles, teachers may continue to struggle with assessment policies in a standards-based system.

The disagreement with achievement-only grading expressed by the participants in this study was consistent with the results of research from the early 1990s to the present. Teachers in a variety of educational contexts have indicated that student effort, in particular, influenced the determination of grades, including teachers in Canada (Hunter, Mayengab, and Gambell 2006; Duncan and Noonan 2007), England (Biddle and Goudas 1997), Israel (Resh 2009), and the United States (e.g., Cross and Frary 1996; Cicmanec 2001; McMillan and Lawson 2001; Zoeckler 2005; Green et al. 2007). McMillan and Nash (2000, 7) found teachers "pulling for students" in their assessment practices—meaning that they were oriented to helping students succeed. While this is crucial in assessment for learning, it can lead to inequity if teachers are not aware of the values underlying their grading decisions. An example of this is when effort is considered only in certain

situations, or for certain groups of students, but not for others (Brookhart 1993; Rich 2002; Howley, Kusimo, and Parrott 2000; Resh 2009). With the growing body of research on grading over the past few decades, it has become increasingly evident that grading is inescapably value laden, and that the value placed on effort needs to be acknowledged for policies to support practice. In Ontario, the standardized report card has been revised to emphasize achievement and learning skills, and similar efforts are being pilot tested in other educational contexts (Guskey, Swan, and Jung 2010). At this point, there remain many questions about the role of non-achievement factors in grading. Students may vary considerably in how they display effort, and teachers may be influenced by other factors in evaluating student effort. More work is needed to understand the relationship between students' characteristics, non-achievement factors, and teachers' grading practices, especially in educational contexts where teachers are expected to report on achievement and work habits separately.

Conclusion

This study focused on teachers' grading practices in a standards-based educational system; it did not attempt to study the issue of fairness. However, the participants repeatedly referred to "being fair" in their comments and explanations about grading. It seemed that teachers were driven in their practices by a sense of what was fair for students, which may have included a host of unexamined assumptions, rather than a sound understanding of grading principles. For students' grades to accurately reflect student achievement, teachers needed a better understanding of essential principles and clearer definitions of key concepts, such as the meaning of *most consistent*. Teachers are expected to rely on their professional judgment through the grading process, partly because detailed direction could never address all of the situations that might arise in grading. Nonetheless, this work suggested that teachers would benefit from, and appreciate, more guidance in their effort to produce fair grades. This could be addressed with more focus on assessment principles to support the development of professional judgment in teacher education and professional development programs. Teachers may then be less worried about being fair and more confident that their grading practices are consistent with current principles for fair assessment in standards-based educational systems.

References

Airasian, P. W. 2005. *Classroom assessment: Concepts and applications*, 5th ed. New York: McGraw-Hill.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. 1954. T*echnical recommendations for psychological tests and diagnostic techniques*. Washington, DC: American Psychological Association.

Anders, P., and V. Richardson. 1992. Teacher as game-show host, bookkeeper, or judge? Challenges, contradictions, and consequences of accountability. *Teachers College Record* 94(2): 382–96.

Biddle, S., and M. Goudas. 1997. Effort is virtuous: Teacher preferences of pupil effort, ability and grading in physical education. *Educational Research* 39(3): 350–55.

Bonesronning, H. 2004. Do the teachers' grading practices affect student achievement? *Education Economics* 12(2): 151–67.

Brookhart, S. M. 1991. Grading practices and validity. *Educational Measurement: Issues and Practice* 10(1): 35–36.

Brookhart, S. M. 1993. Teachers' grading practices: Meaning and values. *Journal of Educational Measurement* 30(2): 123–42.

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/Merrill/Prentice Hall.

Bursuck, W. D., D. D. Munk, and M. Olson. 1999. The fairness of report card grading adaptations: What do students with and without learning disabilities think? *Remedial and Special Education* 20(2): 84–92.

Cicmanec, K. B. 2001. High school mathematics teachers: Grading practice and pupil control ideology. Paper presented at the annual meeting of the American Educational Research Association, April 10–14, Seattle, WA.

Cross, L. H., and R. B. Frary. 1996. Hodgepodge grading: Endorsed by students and teachers alike. Paper presented at the annual meeting of the National Council on Measurement in Education, April 9–11, New York.

Dalbert, C., U. Schneidewind, and A. Saalbach. 2007. Justice judgments concerning grading in school. *Contemporary Educational Psychology* 32(3): 420–33.

Delandshere, G., and J. H. Jones. 1999. Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision* 14(3): 216–40.

Duncan, C. R., and B. Noonan. 2007. Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research* 53(1): 1–21.

Eggen, A. B. 2004. *Alfa and omega in student assessment: Exploring identities of secondary school science teachers*. Oslo, Norway: University of Oslo. Available at: *www.ils.uio.no/forskning/pdh-drgrad/doktoravhanlinger/docs/AstridEggenAvhandling1.pdf.*

Gipps, C. 1999. Socio-cultural aspects of assessment. In *Review of research in education*, eds. A. Iran-Nejad and P. D. Pearson, 24: 355–92. Washington, DC: American Educational Research Association.

Green, S., R. L. Johnson, D. H. Kim, and N. S. Pope. 2007. Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education* 2(7): 999–1011.

Guskey, T. R. 2000. Grading policies that work against standards … and how to fix them. *National Association of Secondary School Principals Bulletin* 84(620): 20–29.

Guskey, T. R. 2005. Stability and change in high school grades. Paper presented at the annual meeting of the American Educational Research Association, April 11–15, Montreal, Canada.

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, April 7–11, San Francisco, CA.

Guskey, T. R. 2009. *Practical solutions for serious problems in standards-based grading*. Thousand Oaks, CA: Corwin.

Guskey, T. R., and L. A. Jung. 2009. Grading and reporting in a standards-based environment: Implications for students with special needs. *Theory Into Practice* 48(1): 53–62.

Guskey, T. R., G. M. Swan, and L. A. Jung. 2010. Developing a statewide, standards-based student report card: A review of the Kentucky initiative. Paper presented at the

annual meeting of the American Educational Research Association, April 30–May 4, Denver, CO.

Hargreaves, A., L. Earl, and M. Schmidt. 2002. Perspectives on alternative assessment reform. *American Educational Research Journal* 39(1): 69–95.

Howley, A., P. S. Kusimo, and L. Parrott. 2000. Grading and the ethos of effort. *Learning Environments Research* 3(3): 229–46.

Hunter, D., C. Mayengab, and T. Gambell. 2006. Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing* 11(1): 42–65.

Joint Advisory Committee. 1993. *Principles for fair student assessment practices for education in Canada*. Edmonton, Alberta, Canada: University of Alberta. Available at: *www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf.*

Joint Committee on Standards for Educational Evaluation. 2003. *The student evaluation standards: How to improve evaluations of students*, ed. A. Gullickson. Thousand Oaks, CA: Corwin.

Laucius, J. 2009. Students aren't failing, but system is, teachers say. *The Ottawa Citizen*, April 22: A1–A2.

McMillan, J. H. 2001. Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice* 20(1): 20–32.

McMillan, J. H. 2003. Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice* 22(4): 34–43.

McMillan, J. H. 2004. *Classroom assessment: Principles and practice for effective instruction*, 3rd ed. Boston: Pearson.

McMillan, J. H. 2009. Synthesis of issues and implications for practice. In *Practical solutions for serious problems in standards-based grading*, ed. T. R. Guskey, 105–20. Thousand Oaks, CA: Corwin.

McMillan, J. H., and S. R. Lawson. 2001. *Secondary science teachers' classroom assessment and grading practices*. Richmond, VA: Metropolitan Educational Research Consortium.

McMillan, J. H., and S. Nash. 2000. Teacher classroom assessment and grading practices decision making. Paper presented at the annual meeting of the National Council on Measurement in Education, April 25–27, New Orleans, LA.

O'Connor, K. 2002. *How to grade for learning: Linking grades to standards*, 2nd ed. Thousand Oaks, CA: Corwin.

Ontario College of Teachers. 2008. *Foundations of professional practice*. Toronto, Canada: Author. Available at: *www.oct.ca/publications/PDF/foundation_e.pdf.*

Ontario Ministry of Education. 2005. *The Ontario curriculum grades 9 and 10: Mathematics*. Toronto, Canada: Queen's Printer for Ontario. Available at: *www.edu.gov.on.ca/eng/curriculum/secondary/math.html*.

Ontario Ministry of Education. 2008. *Growing success: Assessment, evaluation and reporting: Improving student learning*. Toronto, Canada: 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, Canada: Queen's Printer for Ontario. Available at: *www.edu.gove.on.ca/eng/policyfunding/success.html.*

Ontario Ministry of Education. 2010b. *Quick facts: Ontario schools, 2007–2008*. Toronto, Canada: Queen's Printer for Ontario. Available at: *www.edu.gov.on.ca/eng/general/elemsec/quickfacts/2007-08/quickFacts07_08.pdf*.

Resh, N. 2009. Justice in grades allocation: Teachers' perspective. *Social Psychology of Education* 12(3): 315–25.

Resh, N., and C. Dalbert. 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): 322–42.

Rex, L. A. 2005. Testing as shaping: A response to "The idea of testing: Psychometric and sociocultural perspectives." *Measurement* 3(2): 103–07.

Rich, R. H. 2002. *Hidden factors in teachers' secondary grading practices*. Proquest Digital Dissertations UMI No. AAT 3036925.

Shepard, L. 2006. Classroom assessment. In *Educational measurement*, 4th ed., ed. R. L. Brennan, 623–46. Westport, CT: National Council on Measurement in Education, American Council on Education, and Praeger.

Silva, M., D. D. Munk, and W. D. Bursuck. 2005. Grading adaptations for students with disabilities. *Intervention in School and Clinic* 41(2): 87–98.

Stiggins, R., D. Frisbie, and P. Griswold. 1989. Inside high school: Building a research agenda. *Educational Measurement: Issues and Practice* 8(2): 5–14.

Suurtamm, C., and B. Graves. 2007. *Curriculum implementation intermediate math: Research report*. Available at: *www.edu.gov.on.ca/eng/studentsuccess/lms/files/CIIMResearchReport2007.pdf*.

Tomlinson, C. A. 2005. Grading and differentiation: Paradox or good practice? *Theory Into Practice* 44(3): 262–69.

WordNet 3.0. 2006. Princeton, NJ: Cognitive Science Laboratory, Princeton University. Available at: *http://wordnet.princeton.edu*.

Yip, D. Y., and D. Cheung. 2005. Teachers' concerns on school-based assessment of practical work. *Journal of Biological Education* 39(4): 156–62.

Yung, B. H. W. 2001. Three views of fairness in a school-based assessment scheme of practical work in biology. *International Journal of Science Education* 23(10): 985–1005.

Zoeckler, L. G. 2005. *Moral dimensions of grading in high school English*. Proquest Digital Dissertations UMI No. AAT3183500.