PUTTING CLASSROOM GRADING
ON THE TABLE:
A REFORM IN PROGRESS

Author

KENI BRAYTON COX, PH.D., is an Assistant Professor for the Department of Educational Leadership in the College of Education at California State University in Fullerton, California.

Abstract

In a standards-based instructional program, what does a course grade convey? What should it convey? What is the role of homework in assigning grades? What is the role of common assessments? This case study examined the responses of two groups of high school teachers during a district wide reform of grading practices. The first was a focus group of seven advocates of non-traditional grading practices aligned with specific district grade reforms: a) 50% minimal score for a failing grade, b) retesting without penalty, c) acceptance of late work, and d) course-alike, standards-based grading agreements. The second group of nine teachers, including seven randomly selected teacher leaders from across the district, participated individually in semi-structured interviews. Findings confirmed earlier research on the role of individual teacher beliefs on grading practice and the emphasis that many teachers place on student effort when they assign grades. Additional findings have implications for leadership actions that may influence grading practices of secondary teachers.

During the standards-based movement in K-12 education, there remains one arena where few educational leaders and reformers have ventured: classroom grading. Although content standards and local assessments may have improved consistency and coherence in curriculum and instruction, grades and grading remain largely the domain of individual teachers, par-

67

ticularly at the secondary level. Standards-based report cards may have become commonplace at the elementary level, but at the secondary level, report cards look pretty much as they did when the Committee of Ten convened in 1892 to consider high school reform. Letter grades (A-F) designate relative levels of student performance, and students' grade point averages (GPA) are computed on a 4-point scale.

Literature focusing on grading practices spans multiple fields of inquiry, including educational psychology, theories of learning and motivation, sociology, and tests and measurement (Brookhart, 2004; Stiggins, 1999) and was prominent a generation ago (e.g., Crooks, 1988; Dorr-Bremme & Herman, 1986; Stiggins & Conklin, 1992; Terwilliger, 1989). A prevalent theme in the literature then and today is the discrepancy between actual and recommended practice (e.g., Brookhart, 1994; Guskey, 2006, 2009; McMillan, 2001; Ohlsen, 2007; Stiggins, 1999, 2005; Stiggins, Frisbie, & Griswold, 1989).

Researchers examining the rationale behind teacher grading practices have found that individual teacher beliefs and values are significant influences (Brookhart, 1997; Graham, 2005; McMillan, 2001, 2003; Senk, Beckman, & Thompson, 1997; Zoeckler, 2007). In the current context of high-stakes testing, standards-based instruction, and the increased use of local assessments, researchers and authorities in the field have addressed the potential for reform in grading practice (Brookhart, 2003; Carifio & Carey, 2009; Guskey, 2009; Reeves, 2004; Stiggins, 2005; Wormeli, 2006). The standards movement may well have set the stage for renewed consideration of this vital educational element (Stiggins, 2005). In the beginning of the 21st century, however, the field lacks compelling evidence of reform.

Background of the Study

The district that I call Lincoln Secondary School District (all names in this study are pseudonyms), consists of five comprehensive high schools a continuation school, and an adult education program. Primarily Hispanic and challenged by poverty and language diversity, this district has a compelling history. Over the past eight years, the district has established itself as a district in reform. Once a proudly decentralized institution, the district now boasts of a consistent, standards-based instructional program. District-wide common assessments, curricular pacing charts, data-based collaboration meetings, and daily tutorials are part of the culture across all sites.

District leaders began strategizing how to move toward reform in classroom grading practices in 2006. Their goal was to alter how students are graded and, in that process, to reduce failure and improve learning. The

theory of action guiding this change was consistent with past efforts and began by enlisting the support of teacher leaders.

In 2006, district leaders attended an Association of Supervisors of Curriculum Development (ASCD) conducted by Thomas Guskey, Professor of Educational Measurement and Evaluation at Georgetown College, Georgetown, Kentucky. In the spring and fall of 2007, district administrators began to "plant the seeds." "The Case against the Zero" (Reeves, 2004) and *How to Grade for Learning* (O'Connor 2002), a teacher guide for altering grading practices, were widely distributed. The target group was English, mathematics, science, and social science teachers.

In spring 2008, teacher leader meetings focusing on grading practices continued, followed by a web-based survey. A district-wide in-service day, with more than 80% of the over 500 teachers in attendance, featured teacher-led workshops and was introduced by the superintendent via a video conference with Doug Reeves.

In the course of this effort, district leaders, true to the culture of the district, issued no mandates. Leaders promoted the collaborative development of site-based, course-alike agreements for grading students as the natural extension of standards-based instruction. A frequently voiced phrase was to "keep hope alive" by adopting policies, within each course, that promote student learning and reduce the numbers of students who get so far behind early in the semester that passing a class becomes impossible. Four specific reforms were put "on the table."

- Course-alike grading policies: A policy that establishes consistency in "what counts" for a course grade among teachers of the same course at the same school.
- A minimum 50% F: Eliminating the zero on a traditional 100-point scale in favor of an equal intervals between grades (e.g., 90 = A, 80 =B, 70 = C, 60 = D, 50 = F).
- Test retake policy: Permitting students to retake tests, holding them accountable for learning the standards, and giving them credit for that learning.
- Late work acceptance policy: Acceptance of late assignments for credit.

As evidence of the transparency of the endeavor, these desired practices were featured in the district's Guide to Instructional Practice, a document widely distributed across the community.

69

Significance of the Study

Reform in grading lags behind other reforms necessary for standards-based practice, at a cost to students and the collective efforts of the profession to achieve both equity and excellence (Guskey, 2009; Stiggins, 2005). Although the field is replete with examples of individual teacher practice regarding classroom assessment and grading (Graham, 2005; Ohlsen, 2007; McMillan, 2001, 2003; Zoeckler, 2007), examples of large-scale, directed reform in grading practice are minimal. Increasingly, districts have been recognized as necessary actors in implementing and sustaining reform (Agullard & Gouhnour, 2006; Hightower, Knapp, Marsh & McLaughlin, 2002; Waters & Marzano, R.J. (2006). Leaders interested in grading practice reform, however, have few examples from which to draw as they consider their own options and strategies. This case study sought to address that gap by examining the strategies and impact of a district-wide effort to effect change in classroom assessment practice.

Purpose of the Study

The purpose of this study was to describe the classroom grading practices of individual teacher leaders in Lincoln Secondary School District (LSSD) and to understand their thoughts and concerns as they considered the four district initiated proposed reforms. For this study, classroom-grading practice was defined as the method and criteria that teachers use for assigning end-of-course grades for a given course. The researcher pursued answers to the following questions:

1. What are the current grading practices of "high implementers" in LSSD as identified by the district, and how do they perceive and explain their decisions regarding student grades?
2. What are the current grading practices of selected teacher leaders, and how similar are these practices to practices employed by others teaching the same course?
3. What are the things teachers think about in making or resisting the changes being advocated?

Perspective and Researcher Role

I approached this investigation as an advocate of grading reform and from the perspective of a researcher with more than 20 years of district leadership experience, part of that experience in a neighboring district. (I also served as a consultant to the district, in an unrelated area, for five months in 2008). In the design of this study, I was assisted prominently by a participant ob-

70

server, also an advocate and one of the architects of the reform effort.   Our
shared interest was to gain insight regarding the potential of a district led
grading reform effort and how that effort might be strengthened.

In planning the approach to this case study, we were influenced by
Stiggins and Bridgeford's (1985) application of the Concerns Based Adop-
tion Model (CBAM). CBAM is an appropriate diagnostic tool (Hall, 2010;
Roach, Kratochwill, & Frank, 2009) for assessing the status of educational
change in terms of individual implementer concerns (information, personal,
task, consequence, collaboration, and refocusing). CBAM can also be con-
ceptualized as three broader stages: concern for "self," concern for "task,"
and concern for "impact" (Hall & Hord, 2006). Expressed concern for self
focuses on the individual teacher and his or her practice and beliefs; task
is a focus on implementation, time requirements, and ease of use. Impact
encompasses consequences in the classroom and for students, collabora-
tion with other teachers, and efforts to refine what is already being done. In
designing this study, we sought to identify where teachers were in terms of
self, task, and impact.

Methodology

This inquiry consisted of two parts, each with a separate group of partici-
pants, instruments, and procedures.

Part I

Part I was designed to answer the first of three research questions: what
are the current grading practices of high implementers as identified by the
district, and how do they perceive and explain their decisions regarding
student grades? This initial question was designed to establish a reference
point for the remainder of the story.

Participants. In December 2008, seven Algebra I teachers at Wilson
High School, a course-alike cadre identified by site and district leaders as
exemplars of grading reform, agreed to participate as members of a focus
group. The years of teaching experience among the participants ranged
from two to thirty, with a median of six years; there were four male and three
female teachers. Don, the mathematics department chair at Wilson is a
primary leader in the grading reform movement and issued the invitation to
participate. He did not participate in the focus group interview, but he did
offer, "You know, we're not about 'Stand and Deliver;' we are about building
a team."   Focus group participants were the Algebra I team.

Instrument. The focus group protocol was structured around the fol-
lowing questions:

71

1. Speaking as you would to a parent or student, please explain how my child's performance in your class will be assessed and evaluated. To what degree are these practices shared?
2. Why do you do what you do? What factors have contributed to your current practices?
3. Was there a time when these practices were uncomfortable, and if so, what were your thoughts and feelings at that time?
4. What does a grade [in your course] convey?

I conducted the focus group and directed questions to the group as a whole, following each initial response with probes such as, "Is this true of others? How do others see this? What else do you want to add?"

Procedures. I met participants in a mathematics classroom at Wilson High School and distributed and reviewed written guidelines for focus group protocol. During an extended lunch of approximately 55 minutes, with lunch provided by the researcher, each teacher signed an informed consent form. I recorded the conversation on two digital recorders and took field notes.

Part II

Part II addressed research questions 2 and 3: What are the current grading practices of selected teacher leaders, and how similar are these practices to those employed by others teaching the same course? What are the things that teachers think about in making or resisting the changes being advocated?

Participants. Nine teachers participated in individual interviews. Seven were randomly selected from the population of teacher leaders called "course leads," and two were selected purposefully. "Course lead" is a position that is funded by the district to support the work of department chairs. Course leads facilitate course-alike meetings that include the analysis of common assessment results, curriculum planning, and the sharing of instructional strategies. These teachers were appropriate participants because their positions placed them in the forefront of the reform effort. An LSSD district administrator explained the purpose of the research during a district-wide meeting of course leads and alerted the teachers to the possibility that they might be asked to participate in a study focused on grading practice.

The number of course leads per high school ranges from 13 to 18. Using a random numbers generator, I identified two course leads from each comprehensive high school, excluding Wilson High, the site of the focus group. I contacted each teacher by telephone, leaving messages and a follow-up email. Five of the eight responded quickly, one did not respond, and one was eliminated due to a scheduling conflict. Replacements were

selected randomly, resulting in seven participants representing four schools. I emphasized the voluntary nature of their participation and the promise of confidentiality in each communication.

Two additional teachers were selected purposively. One interviewee, although not a course lead, was a veteran teacher at Wilson High School with a 30-year history at the same school and one of the high implementers in the focus group in the school. The other teacher, a course lead and science department chair at Taft High School was identified by district leaders as "someone you need to talk to." Table 1 presents the interview participants according to years of experience, school, and teaching/course lead assignment.

Table 1. Interviewed Teachers Rank Ordered According to Years of Teaching Experience				
Participants	School	Department	Course Lead	Years Teaching
Teacher 1	Roosevelt	Social Science	World Civilizations	32
Teacher 2	Wilson	Math	NA	30
Teacher 3	Taft	Math	Geometry	15
Teacher 4	Madison	English	American Lit.	12
Teacher 5	Adams	Social Science	World Civilizations	10
Teacher 6	Madison	Science	Biology	10
Teacher 7	Taft	Science	Biology	10
Teacher 8	Taft	Math	Algebra (Gr. 9)	7
Teacher 9	Adams	Math	Algebra 1 (Gr. 10-12)*	4

Note. *For students who failed Grade 9 Algebra.

Instrument. The nine individual interviews were semi-structured and centered on three key inquiries:

- What is your current practice in assessing student achievement and determining course grades?
- How similar or dissimilar is your practice from others teaching the same course?
- What kinds of things do you think about when considering your grading practices?

I asked teachers to focus on the course for which they were responsible as course leads as the basis for their answers. The question, "What kinds

73

of things do/did you think about when making the choices that you do?" provided an opportunity for multiple probes related to the specific practices "on the table": course-alike grading policies, retesting, minimum score for an F, and acceptance of late work. Additional questions, determined by participant responses, included, "If you were 'boss,' what kind of grading policies would you design?" "Where do you see grading practices moving over the next two years?"

   Procedures. I recorded each interview on a digital audio recorder and took field notes. Interviews took place in February 2009, in a private setting on teachers' respective campuses during a conference period, before or after school. Each interview lasted between 45 minutes and 2 ½ hours as dictated by teacher interest and the time available. I assured teachers that their participation was voluntary, that their confidentiality would be protected, and that no answer would be associated with any individual respondent.

Data Analysis

I transcribed recordings of the focus group and individual interviews and coded the transcripts according to specific areas of practice: retesting, minimum F, acceptance of late work, and "what counts," as reported by focus group participants and each interviewee. Although the question, "What kinds of things do you think about when making the choices that you do about student grades?" was intended to illicit expressions of concern, most teachers interjected feelings and opinions throughout the exchange. All comments expressing opinions, rationale for grading practices, or feelings regarding grading were included in the qualitative data analysis.

Results

   Grading practices of the high implementers. The focus group of high implementers confirmed for me their chair's assertion that they are "a team." They were eager to share.

Figure 1. Summary of course-alike practices reported by the high implementers' focus group.
• 70% tests and 30% teacher discretion. Tests include common assessments and the final exam. An A on the final exam = an A for the course. "Advanced" or "Proficient on the California Algebra 1 Standards test results in raising student grades to a B or an A.
• Students may retake any test, with the highest score being recorded.
• The lowest score that a student will get on any test is 50%.
• Late work is accepted without penalty (6 of the 7).

74

As illustrated in Figure 1, this group of Algebra I teachers have a course-alike agreement: they permit students to retake tests; 50% is the norm for a failing grade; and every teacher accepts late work without penalty. The rationale for each of these grading practices became clear in the course of a fast-paced conversation.

Shared grading policy. Relative to the development of a shared grading policy, the teachers agreed that common assessments played an important role. Craig, a 30-year veteran, explained it succinctly:

> Before we had common assessments . . . students would complain that if they got teacher A, they could get an A, but if they had teacher B, they'd only get a C . . . Teachers weren't weighting things the same; it wasn't fair for the students. [Now] they're pretty much going to end up getting a similar grade.

Juan (4 years) added, "With common assessment there needs to be an agreement within the department as to what we want the students to understand."

As expressed by Crystal (11 years) a lack of alignment between grades and state assessment results affected their initial discussions on grades.

> I know when we looked at the data and the CST [California Standards Test] scores had risen quite dramatically, and our grades didn't reflect that . . . that's when I think we really looked at alternative grading practices to match those things . . . our kids were making strides, yet in our own personal grades, it wasn't reflected.

Minimum 50% F. "Keeping hope alive" is a continuing refrain in the district, and focus group members reiterated that goal, specifically in reference to the 50% minimum F policy and permitting students to retake tests.

Agreeing to the 50% minimum for an F as a course-wide policy was, at first, uncomfortable for Craig, with 30 years of teaching experience and the most veteran of the group.

> At first, it didn't sit well with me. It was brought to us by the district, and I thought if a kid's not going to prepare for the exam, why should he be rewarded with getting a higher grade? Then I thought about it; it's still an F. It's still 50%, and it does give the kid the opportunity to get out of the basement...

75

"If a kid fails a test," added Joe (9 years), "they get a minimum 50% . . . giving them that hope that they can get it up to a D."

Test retake policy. Algebra I teachers administer 14 common assessments a year. In accordance with the group's grading practices agreement, students may retake any test without penalty; the highest score "counts." In this discussion, hope was cited once again.

> Some people say if it was in the real world, people don't get a chance to make up a test. . . but I just don't see the harm in this process . . . in giving them the hope. (Thad, 9 years)

Juan, a 4th year teacher, asserted that "hope" benefits both students and teachers:

> There's hope for the student but it's also hope for the teacher because when you have too many students who get an F, who are on the bottom, it's pretty hard to motivate.

Craig added,

> Like Juan said, once a kid gives up, you can't get through to them . . . they don't come to school; they come late . . .

Crystal (11 years) acknowledged, among general agreement, that retesting is also a matter of fairness; teachers and teaching are not always perfect:

> We know we can definitely drop the ball on certain topics and we don't want our students to pay the price for that; and if there is something I totally misjudged or I assumed the kids would get quickly and they don't, a retest is a fair opportunity for all to make up for that.

The final question for the focus group concerned what a grade in their class represents: Absolute achievement or student growth? Absolute achievement or effort? Absolute achievement or achievement relative to ability? The seven participants responded "absolute achievement" every time.

Grading practices of selected teacher leaders. Individual interviews of the teacher leaders (course leads) provided a diverse set of responses. The nine teachers were forthright and often impassioned about their grading practices, the importance of course grades, and what grades do or should represent. They appeared engaged by the questions and eager to be heard.

76

Table 2 provides a summary of the participants' current assessment practices, relative to the proposed reforms. Individual teachers are identified by subject, school, and years of teaching experience. Teachers are listed according to the number of reforms they reported using.   Each column identifies a desired assessment practice, beginning with "course-alike grading agreement" and concluding with "acceptance of late work." Practices reported by the high implementer focus group are included at the top of the table to provide a comparison.

Table 2. Grading Practices Reported by the Focus Group and Nine Teacher Leaders						
		Elements of Grading Reform				
Teacher Course School	Teaching Experience	Grading Agreement (Same course/ same school)	Retesting Permitted w/ no penalty	50% minimum on all tests	Acceptance of late work; no penalty	Total Elements Implemented
High Implementer Focus Group	Median 6 Years	YES	YES	YES	YES	4 OF 4
Teacher 2 Algebra 1 Wilson	30	YES	YES	YES	YES	4 of 4
Teacher 3 Geometry Taft	15	YES	YES	NO	YES	3 of 4
Teacher 8 Algebra 1 Taft	7	YES	YES	NO	YES	3 of 4
Teacher 1 World Civ. Roosevelt	32	YES	YES	NO	NO	2 of 4
Teacher 6 Biology Madison	10	YES	NO	NO	NO	1 of 4
Teacher 9 Algebra 1 Adams	4	YES	NO	NO	NO	1 of 4
Teacher 7 Biology Taft	10	NO	YES	NO	(does not grade homework)	1 of 3
Teacher 4 AP Am. Lit. Madison	12	NO	NO	NO	NO	0 of 4
Teacher 5 World Civ. Adams	10	NO	NO	NO	NO	0 of 4
	TOTAL	6 of 9	5 of 9	1 of 9	3 of 8	

As shown in Table 2, the most highly implemented reform among these teachers, was grading agreements (6 of 9); the least implemented reform was the 50% F (1 of 9). Though not shown in Table 1, both course-alike agreements and individual grading practices reported by teachers tended to give more weight to tests than to any other single factor, a characteristic they had in common with the high implementers.

What teachers think about relative to the grading reform. The contextual variations among teachers, the rationale for what they do, and their level of concern are not easily summarized. Analysis of interview data resulted in the identification of three emergent themes: (a) what a grade conveys; (b) the nature and purpose of schooling; and (c) common or "shared" grading practices. I discuss each theme below accompanied by individual comments.

What a grade conveys. Achievement as the primary basis for grading students is recommended in the literature (Brookhart, 1994; Guskey, 2009; Senk, Beckmann, & Thompson, 1997; Stiggins, 2005), and a major intent of the grading reform effort in LSSD. Leaders sought to reduce the impact of non-achievement criteria (e.g., late work) on course grades. Consistent with other research (e.g., McMillan, 2001; Stiggins, Frisbie, & Griswold, 1989), teachers in LSSD consider both achievement and indicators of effort to compute grades with a primary focus on achievement.

While the high implementers "counted" tests a minimum of 70%, the individual teacher leaders varied greatly, with tests counting as little as 48% and as much as 100%. Seven of the individual teachers were particularly adamant about the importance of including evidence of student effort as a grading criterion, with comments similar to the following:

> We are really lucky in my department because we all have similar philosophies because we know there are kids that work their tails off and don't do well on tests . . . but I feel it is really important to assess them on the amount of work they put in. (Teacher 6, Madison)

For me, every kid should be getting a C or better if they turn in their homework . . . we're [department] really big on trying to reward students on effort and work effort. (Teacher 5, Adams)

For Teacher 8 (Taft), effort and the completion of homework are requisites for getting an A, regardless of how well the student might perform on tests.

> If they get A's (on tests) and little or no homework, they'd pass with a C or better because they're proficient. An A? *No. No. No one* [original emphasis] would say that! It's just unfair.

79

Contrary to the belief of Teacher 8, there were teachers in LSSD who indicated that they discount homework as a grading criterion and award A's based on tests alone. The high implementers reported that they raised student grades (to A or B) if student performance on the final exam is an A or B, regardless of previous test scores and/or other assessments (Figure 1). Perhaps more surprising, students who score as Advanced or Proficient on the California Standards Test (CST) in Algebra, a test administered in May, also merit grade change to A or B, assuming that their pre-CST grade is lower.

Teacher 7 whom I labeled the "innovator" proved to be the outlier on several fronts. Similar to the high implementer group, the Biology course lead at Taft focused on achievement, allowed retesting, and changed course grades based on end of year standardized state tests. However, he rejected his own department's grading agreement, assigned homework but did not grade it, "counted" common assessments minimally and rejected the 50% minimum F. Instead, he "tests" students daily with short, open-ended, "standards-based questions," and these tests as well as quarter and end-of-semester exams constitute 100% of the evidence for grades. He assigns, but does not collect or grade homework and reported that in 2008, he raised 15 grades as a result of students' scores on the California Standards Test in biology. He explained his rejection of the course-alike grading agreement:

> I realized I had given my kids a test and some of them had done quite well, but when it came down to giving them grades they had really poor grades, just because they weren't doing the busy work, the homework, all the little things that we expect kids to do on a daily basis that don't reflect knowledge at all . . . I asked myself, what are my grades reflective of? I was tired of what I call 'point grubbing.' (Teacher 7)

He added, "I want my grades to represent one thing, and one thing only: Do they know the material?" At the invitation of district leaders, the innovator shared his grading practices at a district inservice; he expressed doubt that he had any converts.

The nature and purpose of schooling. Two areas of the reform elicited more expressions of dismay than any other: the 50% minimum test score and acceptance of late work and reforms intended to reduce failures and encourage student effort. Six teachers expressed strong objections to these policies. The recurring refrain was, "we are not preparing students for the real world."

> It's a district push for every school to have a unified late work policy . . .

80

it's a nice idea but how does that prepare them for real life? (Teacher 4) If our students get the idea that in a job they're going to get paid for no work, well, they're not going to get paid for no work and if you're going to work on a job and do half the work, then you're going to get half the pay. (Teacher 5)

Even the innovator, who does not count homework at all, expressed concern, "If you do 25% of the work, you're going to get fired; in fact, if you do 65% of the work you're probably going to get fired" (Teacher 7). Teacher 1 stated, "It's [50% F] not accepted in my department; we all have pretty strong feelings that a student gets what he's earned."

In contrast, one of the high implementers challenged the idea that the "real world" is unforgiving. "This is my second career. Once you start any kind of job, you have second chances; we all make mistakes. You have more than one chance to learn." Another high implementer stated, "We're not talking about adults; these are kids who are still building their confidence and figuring out who they are."

Common grading practices. In urging schools to develop course-alike grading policies, district leaders presented a compelling case for supporting standards-based instruction with consistent grading practices. Common assessments and pacing charts, norms in most departments, supported this argument. Six of the nine teacher leaders reported that they were part of grading agreements. Teacher 4, the sole English teacher in the sample, was singularly adamant in his objection to every aspect of the district's reform effort, including the call for consistency:

Consistent grading practices? Suffice to say that it has not garnered unilateral, by any stretch of the imagination, support across all departments, or even within departments . . . If you're going to respect the value of individual teachers, their professionalism, their expertise, their ability, if you're going to not just pay lip service to it but really honor it, you've got to give teachers the benefit of the doubt when it comes to why this is 25% for you and only 20% 'for another.'

For most, however, having an agreed-upon set of criteria for grading had advantages. Teacher 8 expressed her support:

We wanted to be rigorous; we were finding out that sometimes students would come to your [advanced math] class and they'd have a C, and they came in without knowing any of their algebra. So we started look-

81

ing at everybody's grading policy. Some were giving too much credit for homework; that's how we came up with the rubric [agreement] we have now.

More generally, teachers viewed common grading practices as a matter of fairness. Students taking the same course and the same tests should receive the same grades.

Discussion

The purpose of this investigation was to describe the current classroom practices of individual teacher leaders in LSSD, a district intent on changing the norms of teacher grading practice, and to understand the thoughts and concerns relative to grading. Results indicate that teachers' current grading practices remain largely traditional, based on a combination of achievement and effort. Nevertheless, there are indications that the reform effort has made progress. Grading practices are definitely "on the table," a topic of discussion and debate. Furthermore, each of the teacher leaders represented a course-alike cadre at their respective sites, and six of the nine reported the establishment of shared grading agreements. Conservatively, these six agreements include more than 70 teachers. Moreover, the district focus on grading practices led at least one teacher to review his own practice and the meaning of grades. Although the innovator rejected his department's course-alike agreement, he implemented grading policies in line with the intent of the reform: a focus on achievement and second chances.

Measures of achievement. In identifying achievement as "what grades should convey," most of the teachers relied on traditional tests. The common assessments adopted across the district, including the 14 tests in Algebra I, are multiple-choice, easily scanned and scored for fast analysis. These exams and the collaborative analysis of assessment results have had a significant impact on teacher practice, as reported by a majority of teachers. Only the innovator (Teacher 7) rejected the multiple-choice format, opting for short answer quizzes.

At the time of this study, teachers were working on the development of alternative testing protocols. Common assessments were entirely multiple choice, which facilitated quick turnover in results and were patterned after the high-stakes annual exams required by the State of California.

Expressed concerns of teachers. An initial intent of this research was to examine teacher responses relative to Stages of Concern (Hall & Hord, 2006). We hoped to determine whether teachers expressed concerns focused on self, task, or impact. Although the data did not support a robust

82

CBAM analysis, interesting quandaries arose.

Concerns voiced by the high implementers appeared to be at the impact level, focusing on students. They articulated clearly their focus on motivating students by, for example, "keeping hope alive," but they also said, "it's hope for the teacher, too." Impact or personal level of concern? Similarly, did the comment, "we are not preparing students for the real world," indicate a predictable resistance to changing traditional practice, or is it at impact level, a concern for students? These and similar comments may be more easily understood as expressions of personal values and beliefs, individual perceptions of the purpose and nature of schooling as discussed above. This interpretation of the data would be consistent with earlier research (McMillan, 2001; Graham, 2005).

Surprisingly, expressed concerns regarding task or management issues were minimal. One teacher, Teacher 6, said that logistics and time had contributed to her department's reluctance to implement a retesting policy.

Implications

Without suggesting that the results of this case study of a single high school district is representative of districts similarly engaged, preliminary analysis of results suggests implications in four areas: the need for alternative assessment methods; prerequisites for effective district reform; building a strong knowledge base and philosophical foundation for assessment; and revision of report cards.

Alternative assessment methods. Common assessments are a norm in LSSD, supported by the ease of development and scoring. The district has supplied each course lead with a scanner for immediate scoring of the exams, and time is allocated for test development, revision, and collaborative analysis of results. Assessments for learning, however, require a variety of methods (Stiggins, 2005). Similar recommendations have been made by the National Council of Teachers of Mathematics (Senk, Beckmann, & Thompson, 1997; Ohlsen, 2007).

Grading policies, like assessments, need to be sensitive to the complexity of different subject areas (Duncan & Noonan, 2007). Results suggest that content areas made a difference in the level of implementation in LSSD. Neither Teacher 5, World Civilizations, nor Teacher 4, American Literature, or their course-alike colleagues were supportive of the reforms. Teacher 1, another World Civilizations teacher, indicated compliance with two reforms: a common grading agreement and retesting; but "compliance" was the emphasis of his remarks, not support. Teacher 4, an acknowledged resister, voiced concerns that merit attention. Valid assessments may look

83

very different in the humanities as compared to those in mathematics.

Requisites for district reform. Lincoln Secondary School District has been a district in reform since 2000. Several characteristics and norms of this district and its leadership supported the current reform effort:

- Common assessments and curricular pacing charts
- Dedicated time for teacher collaboration and data analysis
- A strong system of distributed teacher leadership, e.g., course leads

Perhaps most relevant in the context of a values-based reform, such as grading, LSSD consistently avoids mandates. Such was the case in this reform: grading practices were rolled out and presented for discussion and debate, not as directives.

More difficult to describe is the climate in LSSD, a climate built on a culture of strong, mutual respect. Throughout this investigation, teachers expressed positive feelings regarding both school and district leadership.

Building a knowledge base and philosophical foundation. In LSSD, the philosophical drive behind their collective reform efforts has been, in the words of the superintendent, a "laser-like focus on achievement and doing 'whatever it takes' to close the achievement gap." This strong moral base is articulated among all stakeholders, and progress in achievement has merited recognition. Next steps for LSSD, and these are in progress, are to continue to build teachers' knowledge base regarding assessment practice. Results of this study suggest that traditional grading practices, practices that continue to reward effort and compliance at the expense of student learning (Brookhart, 1994; Deci, Koestner & Ryan, 2011; DeLisle & Hargis, 2003; Guskey, 2009; Stiggins, 2005), are based on philosophical beliefs regarding schooling and the role of grades (Graham, 2005; McMillan, 2001; Zoeckler, 2007). LSSD and other districts engaged in these efforts must be prepared to put these beliefs, along with the reforms, on the table. LSSD, to its credit, is doing just that; the superintendent has urged all teachers to have the discussion, to talk about grades, grading, and the impact on students.

Revision of high school report cards. Report cards remain the principal vehicle for communicating course grades, grades that become part of a student's permanent file and transcript. The current design in LSSD, and not unusual for secondary schools, is a letter grade for achievement (A-F) and a citizenship grade, satisfactory or unsatisfactory. Neither of these marks provides an opportunity for teachers to report students' work habits. A single additional category may provide a satisfactory opportunity for teachers to express their values regarding the work ethic without compromising the sub-

84

ject grade. Guskey (2006, 2009) has written persuasively about this need, a need that addresses perceived needs to take into account what McMillan (2001) refers to as academic enablers, such as homework and participation.

Limitations

The limitations of this study include the focus on a single urban school district. The size of the sample and the unintended bias, such as an over-representation of mathematics teachers, also limited the study. Time was limited, especially with the focus group, although their contributions in a short time frame were instrumental in framing the rest of the study.

Recommendations for Future Research

Few studies of secondary grading practices have been reported in the last decade as the focus turned toward standards and formative assessment measures. Nevertheless, final course grades continue to play a significant role in high schools and have long-lasting consequences for high school students. Even less studied are central office or school-wide initiatives intended to reform grading practice. The field would benefit from increased attention to the role of the organization in effecting grading reform. The stories behind these efforts may encourage others to put classroom assessment "on the table."

References

Agullard, K., & Gouhnour, D.S. (2006). Central office inquiry...assessing organization, roles, and actions to support school improvement. San Francisco, CA: West Ed.

Brookhart, S. (1994). Teacher grading: Practices and theory. *Applied Measurement in Education, 7*(4), 279-301.

Brookhart, S. (1997). A theoretical framework for the role of classroom assessment in motivating student effort. *Applied Measurement in Education, 10*(2), 161-180.

Brookhart, S. (December, 2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5-12. Doi:10.1111/j/1745-3992.2003.tb00139.x

Brookhart, S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teacher College Record, 106*(3), 429-456.

Carifio, J., & Carey, T. (2009). A critical examination of current minimum grading policy recommendations. *The High School Journal, 93*(1), 23-37.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Deci, E. L., Koestner, R., & Ryan, R.M. (2011). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research 2001, 71*(1). doi 10.3102/00346543071001001

DeLisle, R., & Hargis, J. (2003). The big fish. *Education, 125*(4), 701-705.

85

Dorr-Bremme, D. W., & Herman, J. L. (1986). *Assessing student achievement: A profile of classroom practices*. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles.

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

Graham, P. (2005) Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education, 21*, 607-621.

Guskey, T. R. (2006). Making high school grades meaningful. *The Phi delta Kappan, 87*(9), 690-675.

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

Hall, G. E. (2010). Technology's Achilles heel: Achieving high quality implementation. *Journal of Research on Technology in Education, 42*(3), 231-253.

Hall, G. E., & Hord, S. M. (2006). *Implementing change, patterns, principles and pot-holes* (2nd ed.). Boston, MA: Allyn and Bacon.

Hightower, A. M., Knapp, M. S., Marsh, J. A., McLaughlin, M. W. (Eds.). (2002). *School districts and instructional renewal*. New York, NY: Teachers College Press.

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. doi:10.111/j.1745-3992.2003.tb00142.x

O'Connor, K. (2002). *How to grade for learning*. Thousand Oaks, CA: Corwin Press.

Ohlsen, M.T. (2007). Classroom assessment practices of secondary school members of NCTM. *American Secondary Education, 36*(1), 4-14.

Reeves, D. R. (2004). The case against the zero. *Phi Delta Kappan, 86*(4), 324-325.

Roach, A. T., Kratochwill, T.R. & Frank, J.L. (2009). School-based consultants as change facilitators: Adaptation of the Concerns-Based Adoption Model (CBAM) to support the implementation of research-based practices. *Journal of Educational and Psychological Consultation,* (19), 300-320.

Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics. *Journal for Research in Mathematics Education, 28*(2), 187-215 [Electronic version]. Retrieved from JSTOR database.

Stiggins, R. J. (1999). Assessment, student confidence and school success. *Phi Delta Kappan, 81*, 191-198.

Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324-328 [Electronic version]. Retrieved from Academic Search Premiere database.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271-286 [Electronic version]. Retrieved from Academic Search Premiere database.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, N.Y.: State University of New York Press.

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

Terwilliger, J. S. (1989). Classroom standard setting and grading practices. *Educational Measurement: Issues and Practice, 8*(2), 15-19.

Waters, J. T., & Marzano, R.J. (2006). School district leadership that works: The effect of superintendent leadership on student achievement. [A working paper]. Denver, CO: Mid-continent Research for Education and Learning.

Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education, 34*(3), 14-27.

Zoeckler, L. G. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education, 35*(2), 83-102.

87