A Repair Kit for Grading

15 Fixes for Broken Grades

A Repair Kit for Grading

15 Fixes for Broken Grades

Ken O'Connor

Educational Testing Service
Portland, Oregon ■ Princeton, New Jersey

Cover design: Kaija Juszak

Book design and typesetting: Heidi Bay, Grey Sky Design

Editing: Robert L. Marcum, editorbob.com

Project coordination: Barbara Fankhauser

Printed in the U.S.A.

ISBN 978-0-9655101-8-9

Library of Congress Pre-assigned Control Number (PNC): 2007924006

This book is dedicated to all those brave teachers who have been pioneers in grading for learning. They know who they are and they have fixed broken grades.

Table of Contents

Preface

> The Committee on Grading was called upon to study grading procedures. At first, the task of investigating the literature seemed to be a rather hopeless one. *What a mass and a mess it all was!* Could order be brought out of such chaos? Could points of agreement among American educators concerning the perplexing grading problem actually be discovered? It was with considerable misgiving and trepidation that the work was finally begun.
>
> *—W. Middleton, quoted in Guskey, 1996b, p. 13, emphasis added*

This statement—or something similar—probably could have been made in almost any school or district in North America at any time in the past one hundred years. It was actually made by Warren Middleton in 1933. But it could just as easily have been said in 2007 because, although some schools and districts have forged considerable progress, grading still remains an aspect of school that is clothed in myth, mystery, and magic.

It was with equal "considerable misgiving and trepidation" that I decided to take on this difficult topic in an article for the *NASSP Bulletin* (O'Connor, 1995). That misgiving continued when I wrote *The Mindful School: How to Grade for Learning* (O'Connor, 1999; 2002), and continues to this day.

So why a new book, and why do I still have "misgiving and trepidation"?

This new book is necessary because in many schools grading is still a "mass and a mess." Although teaching has become increasingly standards based, and we know more than we ever knew about how people learn, traditional grading practices persist, especially in middle and high schools. These practices often not only result in ineffective communication about student achievement, but also may actually harm students and misrepresent their learning. Thus I feel the need for this "Repair Kit," in which I identify 15 ways to fix "broken" grades—15 things we should do if we want grades to be effective.

One of the major reasons for writing this book now is the increasing emphasis on standards or learning goals (also called learning outcomes). All American states except Iowa now have academic content standards, as do all Canadian provinces. The mandate is that schools are supposed to be standards based for curriculum, instruction, assessment, and grading and reporting, but what I often see, especially in middle and high schools, is some emphasis on standards for curriculum, instruction, and assessment but very little standards-based grading and reporting. My hope is that this book will help schools and teachers develop standards-based grading and reporting practices.

My previous book was more theoretical in its approach. This book's focus is on classroom implementation. The two books therefore complement each other, and my hope is that teachers and administrators will use them together to develop a deeper understanding of the issues in and solutions to concerns about grading. I also am pleased to report that my "misgiving and trepidation" are greatly reduced because I believe there is an emerging consensus about how grades should be determined in standards-based systems (see especially Stiggins, Arter, Chappuis, and Chappuis, 2004; and Tomlinson and McTighe, 2006, for guidelines similar to those in my works cited previously).

For the bigger picture of how grading fits into classroom assessment, this book complements the books and videos produced by Rick Stiggins and his associates at Educational Testing Service in Portland, Oregon. The principles and ideas presented here are derived with permission from those materials, particularly Stiggins et al., 2004.

I have included several examples of what teachers and schools use to help them implement standards-based grading and reporting and to fix broken grades. I would like to thank all those who gave me permission to use these materials, but I must emphasize that they all should be seen as examples only, not as models.

My hope is that individuals and groups in schools and districts will use this book to help them to reflect on the grading practices they currently use. Each of the 15 Fixes open and close with quotations designed to begin or focus appropriate professional dialogue.

I would like to express my appreciation to Rick Stiggins and Steve Chappuis for agreeing that there is a place for this book. I also thank them for the time and effort they put into reviewing drafts of the manuscript and for their many excellent suggestions that have made it a much better book. I also would like to thank copyeditor Robert L. Marcum of editorbob.com and designer Heidi Bay of Grey Sky Design for their considerable help in crafting the final product. As the author, I take full responsibility for the ideas in the book but the final product has been improved immeasurably by the contributions of those four professionals. I also express my gratitude to the thousands of teachers who have attended my workshops over the past seven years, from whom I have learned so much.

Ken O'Connor
Scarborough, Ontario, Canada
October 2006

CHAPTER 

Setting the Stage

> School improvement expert Bob Marzano asks, "Why [w]ould anyone want to change current grading practices? The answer is quite simple: grades are so imprecise that they are almost meaningless."
>
> —*Marzano, 2000, p. 1*

Every state in the United States except Iowa, every province in Canada, and almost every jurisdiction in most other countries now has educational content *standards*—public, published statements of the expected outcomes of learning; that is, what students are expected to know, understand, and be able to do. The primary goal of a standards-based system is for all students to "meet standards"; that is, to be competent or proficient in every aspect of the curriculum. The key to reaching this goal is to evaluate every student's achievement using similar criteria, consistently applied at all levels.

The two essential questions that all educators should ask about their grades are, "How confident am I that the grades students get in my classroom/school/district are consistent, accurate, and meaningful, and that they support learning?" and "How confident am I that the grades I assign students accurately reflect my school's/district's published content standards and desired learning outcomes?" In most schools/districts the answers to these questions, especially at the middle and high school levels, range from "not very" to "not at all." Because of this I believe that, very often, grades are "broken" and that teachers and schools/districts need a "repair kit." I offer such a

kit here in the form of 15 Fixes that teachers and administrators alike can apply to repair broken grading systems.

Clear evidence of this broken condition can be found in the following example, from Associated Press education reporter Ben Feller:

> Math teacher Ms. R (a high school Math teacher in Maryland) grades on the basics: scores on tests and quizzes; participation in class; and the quality of homework her students hand in. But maintaining objectivity in her grading, she says means fending off outside pressures. School administrators want to see A's and B's. Parents often want answers from her when their children's grades slip. Students themselves make appeals. "They will come up and say: 'Oh, I have an 88.9. Can I have an A?' said Ms. R. Her response 'No. You earned the 88. Work harder next quarter.' " (Feller, 2004, n.p.)

Ms. R obviously has a very traditional view of grading and sees "the basics" as scores on tests and quizzes, participation, and homework. But if class participation and homework intended as practice only are included in grades, those grades will not accurately reflect or support learning, or be meaningful to students or to others. Further, summarizing a term's worth of work in a single grade does not reflect student achievement with the precision Ms. R seems to believe it does. Also, for most students higher achievement, whether evidenced by greater demonstrated learning or a "better grade," is not simply a matter of "working harder." (See Stiggins, 2005, Chapter 11; and Stiggins et al., 2004, Chapter 10 for in-depth discussions of these issues.)

As the first "essential question" indicates, effective grades need to meet four overarching criteria for, or keys to, success: they must be consistent, accurate, and meaningful, and must support learning. I define each of these keys here and then weave them into each of the 15 Fixes throughout the rest of

this book. I believe that most teachers, students, and parents would agree that these are reasonable and necessary expectations; disagreements over how to achieve them within the grade itself are at the root of the debate about grading.

Grades need to be *consistent*. The grades students receive should not be a function of whether they are in teacher X's or teacher Y's class. The question, "How good is good enough?" needs to be the same from classroom to classroom; that is, performance standards need to be the same from teacher to teacher. Students achieving at the same level should get the same grade regardless of context. This clearly is not the case in schools where some teachers are identified as "hard" and others labeled as an "easy A." This should *never* be acceptable. To "fix" grades, especially in standards-based systems, it is at minimum essential that all teachers in every school teaching the same grade or same subject/course should determine grades in similar ways and apply similar or the same performance standards. This consistency in the meaning of grades should be systemic at all levels—school, district, and, ideally, state/province.

Grades need to be *accurate*. Inaccurate grades lead to poor instructional decisions being made by and about any student whose grades are used as the basis of those decisions. When determining grades, many teachers continue the traditional practice of combining a large amount of evidence/data into a single summary symbol. This may involve literally hundreds of decisions; if even one is wrong the grade inaccurately reflects student achievement. Inaccurate grades most commonly occur because teachers determine them by blending achievement with behaviors (effort, participation, adherence to class rules, etc.) (Fix 1), poor-quality assessment (Fix 10), and inappropriate use of the mean (average) in combining data (Fix 11). For grades to be "fixed," each of these practices (and others, discussed in Fixes 4, 5, 8, 9, and 12) needs to be eliminated.

Grades need to be *meaningful*. They must communicate useful information to students and to everyone interested in or needing to know about their learning. Traditionally, teachers have collected evidence using various assessment methods and have organized their gradebooks by type of evidence such as tests, projects, and assignments. So, the grading link to learning outcomes has been tentative at best. The "fix" needed for grades to be meaningful is that they must directly reflect specified learning goals. This requires that teachers set up and organize their gradebooks around those goals or standards—not simply summarize multiple marks into a single grade, or organize grades by the date administered, type of assignment or activity, or type of test—by using the standards or some organizational structure arising from or related to the standards (Fix 7). The evidence categories for mathematics, for example, may include "develops and uses number strategies," "compares and orders whole numbers to 100," and "uses estimation strategies." The evidence structure for English may use strands such as reading, writing, listening, speaking, language, and literature.

Grades need to *support learning*. Students and parents need to understand that achieving in school is not about only "doing the work" or accumulating points. When teachers assign a point value to simply turning in work, or put a mark or number on everything students do and use every number when calculating the grade, the message sent to students is clear: success lies in the quantity of points earned. Any intended message about valuing the *quality* of the learning is blurred. We want students to understand that school is about learning. Grades are artifacts of learning; as such, they should reflect student achievement only (Fixes 2, 3, and 6).

Grades also support learning when the purpose of each assessment is clear. *Formative* assessments are designed

to help students improve, and in most cases are not used to determine grades. *Summative* assessments are designed to measure student achievement, and "are used to make statements of student learning status at a point in time to those outside the classroom" (Stiggins et al., 2004, p. 31). With certain limited exceptions, use only evidence from summative assessments when determining grades (Fix 13). We also must allow new evidence to replace old evidence when it is clear that a student knows or can do something today that they didn't or couldn't previously (Fix 14). Finally, and perhaps most importantly, for grades to support learning, we must learn how to involve students in the grading process (Fix 15).

Key Definitions

One problem is that the terms *marks* and *grades* are often mistakenly used as synonyms, although each involves very different processes. A teacher looking at a single assessment and deciding whether a student should get 7 or 8 out of 10, or a 3 or 4 on a rubric, is doing something very different than when that teacher is looking at the evidence accumulated over a grading period and deciding whether that student gets an A or a B (or whatever summary symbols are used). To avoid confusion, we use the following definitions throughout this book (note, however, that the sources quoted herein may not necessarily follow these definitions):

- A *mark* or *score* is the number (or letter) given to any student test or performance that may contribute to the later determination of a grade.
- A *grade* is the symbol (number or letter) reported at the end of a period of time as a summary statement of student performance.

Purpose(s) for Grades

Traditionally, grades have served a number of purposes—communication, fostering student self-assessment, sorting and selecting, motivation and punishment, and teaching/program evaluation (Guskey, 1996a). As Brookhart (2004, p. 21) points out, "It is very difficult for one measure to serve different purposes equally well." She also states, "The main difficulty driving grading issues both historically and currently is that grades are pressed to serve a variety of conflicting purposes" (p. 31). For example, for communicating effectively in a standards-driven environment where many students are succeeding, we need to be communicating the highest possible achievement in the narrowest possible range—all students are successful. However, for sorting and selecting these same students we need to spread them along the widest possible range, thus ranking some high and some low. These two purposes clearly can be in conflict. Bailey and McTighe (1996, p. 120) state that "the *primary purpose* . . . of grades [is] to *communicate student achievement* to students, parents, school administrators, post-secondary institutions and employers." Brookhart (2004, p. 5) suggests, "*Secondary purposes* for grading include providing teachers with information for instructional planning, . . . and providing teachers, administrators, parents, and students with information for selection and placement of students" (emphases added).

A central premise of this book is that, at the district and school levels, there must be a shared vision of the primary purpose of grades. I believe that primary purpose to be communication about achievement, with *achievement* being defined as performance measured against accepted published standards and learning outcomes.

Underpinning Issues

There are three underpinning issues we must consider before addressing the specifics of how to determine grades. They are fairness, motivation, and objectivity and professional judgment.

Fairness

In education we have tended to think of *fairness* as uniformity. All students have been required to do the same assessments in the same amount of time, and their grades have been calculated in the same way from the same number of assessments. But students are different in many different ways, and so treating them the same can actually be unfair. Patterson (2003, p. 572) points out that "fair does not mean equal; yet, when it comes to grading, we insist that it does." Fairness is much more about equity of opportunity than it is about uniformity. For example, some students need to wear glasses and for equity of opportunity they wear their glasses when they need them; for fairness we do not say, "You are doing a test today, but you cannot wear your glasses because everyone is not wearing glasses," or "Some students in this class need glasses, so you will all wear them (whether you need them or not)."

This concept has been captured in the following statement from the policy about provincial testing in Manitoba. All teachers and jurisdictions would be serving their students well if they had a similar statement in their assessment/grading policy:

> All students are given an equal opportunity to demonstrate what they know and can do as part of the assessment process. Adaptations . . . are available for students *including* students with learning or physical disabilities, to allow them to demonstrate their knowledge and skills,

> provided that the adaptations do not jeopardize the integrity or content of the test. (Manitoba Education, Citizenship and Youth, 2006, p. 1, emphasis added)

The italics emphasize that, for fairness, "adaptations" should not be limited to students who have been specifically identified as needing, for example, more time to complete a test/exam.

Motivation

Grades are often *extrinsic* motivators, meaning that their power to influence student behavior derives from outside the student. Many teachers—and parents, grandparents, and other adults—have used grades as extrinsic motivators ("Everyone who gets an A on this quiz can skip the next homework assignment"; "Get a B or better on that test or you can't go to the concert"). However, this use of grades is not always effective or appropriate. Grades certainly motivate successful students, at least some of the time. But they are definitely not motivators for all students, such as those who get grades that are lower than they expect or think they deserve. For these students, grades in fact often act as *de*-motivators. Many schools and school districts have mission or belief statements that state their desire to develop students who are "independent, self-directed, lifelong learners." To achieve this goal students need to be *intrinsically* motivated, meaning that their desire to achieve and improve must arise from within themselves. Intrinsic motivation is clearly in conflict with the use of grades as extrinsic motivators. Thus, as we think about our current and future grading practices, it is important that we examine and apply our knowledge and beliefs about what does and does not motivate students.

Consider this quote from Nora Rowley, the fifth-grade student who is the main character in Andrew Clements's *The Report Card*:

> Most kids never talk about it, but a lot of the time bad grades make them feel dumb, and almost all the time it's not true. And good grades make other kids think they're better, and that's not true either. And then all the kids start competing and comparing. The smart kids feel smarter and get all stuck-up, and the regular kids feel stupid and like there's no way to catch up. And the people who are supposed to help kids, the parents and the teachers, they don't. (Clements, 2004, pp. 72–73)

Clements, through Nora, makes it clear that he believes that not only do grades not motivate many students, but that they can actually damage both student attitudes toward learning and relationships among students. Both in and out of school we provide elaborate systems of rewards and punishments in the belief that this will lead to more of those behaviors deemed desirable and less deemed undesirable. But the research on motivation shows that continued use of extrinsic motivators leads to two main results. First, extrinsic motivators increase students' focus on the reward or punishment rather than on the desired behavior. Second, they give rise to the need to continuously increase the amount of the reward or punishment to elicit the desired behavior (Covington and Manheim Teel, 1996; Gathercoal, 1997; Ginsberg, 2004; Kohn, 1993; Marshall, 2001a; Rogers, Ludington, and Graham, 1998; Szatanski and Taafe, 1999). Thus it is inappropriate to use grades as extrinsic motivators, either to reward desired behavior or to punish undesired behavior. The primary "reward" for learning should be intrinsic—the positive feelings that result from success. As Stiggins notes, "those who experience . . . success gain the confidence needed to

risk trying. . . . Students who experience . . . failure, lose confidence in themselves, stop trying, and [fail] even more frequently. . . . As it turns out, *confidence is the key to student success in all learning situations*" (2001, p. 43, emphasis in original). Actual success at learning, then, is the single most important factor in (intrinsic) motivation, and it is important to recognize that success is relative—success for each individual is seeing oneself getting better.

Additionally, teachers have other tools available to help them change student behavior. As Marshall (2001b, p. 9) points out, "the most effective ways to change behaviors are: 1. using noncoercion, 2. prompting the person to self-assess, and 3. if authority is necessary, having the student own the consequence. When a consequence is imposed, the student feels the victim. When the consequence is elicited, the student owns it and grows from the decision."

The best classroom practices maximize intrinsic motivation and minimize extrinsic motivation. Teachers in these classrooms help students to the critical understanding that "30 years from now, it won't matter what grades you got. What will matter is what you learned and how you used it."*

Objectivity and Professional Judgment

Teachers often say that they are striving to be as objective as possible in their assessment and grading. In my experience, they most often mean that they are trying to be *consistent* in evaluating student work. Such a process in fact involves subjective judgment. The only aspects of learning that can be assessed objectively are such elements as the correctness of factual content, spelling, and calculation.

Assessments themselves are designed subjectively. Teachers create assessments based on their professional judgment of what is to be assessed and how—a subjective

*From a poster seen on the wall of a high school cafeteria in Council Bluffs, Iowa. Source otherwise unknown.

process. We need to acknowledge this and not apologize. As Wiggins (2001, n.p.) notes, "All scoring by human judges, including assigning points and taking them off math homework is subjective. The question is not whether it is subjective, but whether it is defensible and credible. The Advanced Placement and International Baccalaureate assessments are subjective and yet credible and defensible, for example. So-called objective scoring is still subjective test writing." Thus the real issues are accuracy and consistency, more than objectivity versus subjectivity. We need to develop approaches to help teachers both assess and grade more accurately and consistently. One key to accomplishing this is shared understanding of performance standards—our "How good is good enough?" Another is unified approaches to determining grades at the school or district level.

The problem as identified by an assistant superintendent in a Wisconsin school district is that in grading "every teacher sees himself or herself as an independent contractor and they shouldn't be" (personal communication, n.d.). What is needed is a set of guidelines such as the 15 Fixes in this book provide. Making these Fixes part of district or school policy and providing teachers frequent opportunities both for professional learning and dialogue about these guidelines and to carry out shared marking to arrive at a common understanding of performance standards will greatly enhance the probability of consistent grading across teachers and classrooms.

Student Involvement

Over the past few years it has become increasingly clear that student involvement in teaching/learning and in assessment and communication can make significant contributions to improved achievement and positive attitudes about both learning and school. This issue is so significant I incorporate

suggestions about it into many of the Fixes. Fix 15 is a synthesis and summary of these ideas as presented throughout the rest of the book.

The 15 Fixes

The 15 Fixes appear in Figure 1-1. They are organized into four categories—fixes for distorted achievement, fixes for low-quality or poorly organized evidence, fixes for inappropriate grade calculation, and fixes to support learning. We discuss each Fix in turn in the following chapters.

Figure 1-1 The 15 Fixes

Fixes for Practices That Distort Achievement

1. Don't include student behaviors (effort, participation, adherence to class rules, etc.) in grades; include only achievement.
2. Don't reduce marks on "work" submitted late; provide support for the learner.
3. Don't give points for extra credit or use bonus points; seek only evidence that more work has resulted in a higher level of achievement.
4. Don't punish academic dishonesty with reduced grades; apply other consequences and reassess to determine actual level of achievement.
5. Don't consider attendance in grade determination; report absences separately.
6. Don't include group scores in grades; use only individual achievement evidence.

Figure 1-1 The 15 Fixes (Continued)

Fixes for Low-Quality or Poorly Organized Evidence

7. Don't organize information in grading records by assessment methods or simply summarize into a single grade; organize and report evidence by standards/learning goals.
8. Don't assign grades using inappropriate or unclear performance standards; provide clear descriptions of achievement expectations.
9. Don't assign grades based on student's achievement compared to other students; compare each student's performance to preset standards.
10. Don't rely on evidence gathered using assessments that fail to meet standards of quality; rely only on quality assessments.

Fixes for Inappropriate Grade Calculation

11. Don't rely only on the mean; consider other measures of central tendency and use professional judgment.
12. Don't include zeros in grade determination when evidence is missing or as punishment; use alternatives, such as reassessing to determine real achievement or use "I" for Incomplete or Insufficient Evidence.

Fixes to Support Learning

13. Don't use information from formative assessments and practice to determine grades; use only summative evidence.
14. Don't summarize evidence accumulated over time when learning is developmental and will grow with time and repeated opportunities; in those instances, emphasize more recent achievement.
15. Don't leave students out of the grading process. Involve students; they can—and should—play key roles in assessment and grading that promote achievement.

CHAPTER 2

Fixes for Practices That Distort Achievement

FIX 1

Don't include student behaviors (effort, participation, adherence to class rules, etc.) in grades; include only achievement.

> Reports on student . . . achievement should contain . . . information that indicates academic progress and achievement . . . separate from . . . punctuality, attitude, behavior, effort, attendance, and work habits.
>
> —*Manitoba Education and Training, 1997, p.13*

Grades are broken when they do not accurately communicate achievement. The fix for this is to make grades as pure a measure as possible of student achievement; that is, make them reflect only student performance in mastering the public, published learning goals of the state/province/district/school. This is the only way that grades can act as clear communication. Everyone who has a need to know about a student's performance in school certainly can be told that she or he is "a nice student who tries hard," but they also have a right to know the specific level of her or his knowledge in a particular subject at a given point in time.

We know that the grading practices of some teachers have contributed to grade inflation for some students by including desired behaviors unrelated to achievement, while other students who achieve at a high level have received deflated grades because of their failure to exhibit these same behaviors. For example, consider this evidence from two Canadian provinces:

> Girls consistently outperform boys in high school classrooms across Ontario, and the explanation for the gender gap is a systematic bias against boys, the Fraser Institute says. According to six years of Grade 12 "grades" in advanced-level courses, girls get better grades more than 90% of the time in Language arts and about 60% of the time in Math. "Factors such as promptness in coming to class, willingness to cooperate, and what might be considered [good] work habits are distorting the marks," says Peter Cowley, the report's lead researcher. In other provinces where the institute's ratings have become an annual event, researchers have found girls receive better grades overall than boys even when their exam marks are lower. In B.C., "girls receive higher grades on school-based assessments in subjects regardless of their relative performance on the provincial examinations." (In British Columbia final year high school grades are determined partly from school based teacher assessment and partly from external provincial examinations.) (*Toronto* [ON] *National Post*, 18 April 2001, p. F3)

Similar evidence can be found in the United States from the Commonwealth of Virginia:

> Many students . . . get passing grades by working hard in class but (their) academic weaknesses are pinpointed by the SOL's [Virginia's Standards of Learning exit tests]. (Joyce O. Jones, director of guidance at Gar-Field High School in Prince William County, VA, quoted in Helderman, 2004, p. B01)

Teachers combine achievement and other variables, such as behavior, into grades for several reasons. One is the belief that this practice appropriately rewards students who are well behaved and punishes those who do not behave as expected. When thus combined, grades become extrinsic motivators to control student behavior. As noted previously, this does not work for all students. A second reason, particu-

larly prevalent at the high school level, is that teachers have had no way to communicate separately about the behaviors they think are important, and so have blended them together with achievement. The solution for this faulty communication is to use standards-based expanded format report cards where the desirable behaviors are listed and rated. This has become increasingly common at the elementary level, but is not yet a widespread practice in middle and high school reporting. This is somewhat ironic because at the high school level grades serve high-stakes purposes (rank in class, program and/or scholarship eligibility, college admissions, etc.) and thus should depict achievement as accurately as possible to ensure good decisions.

One of the best examples of this type of reporting is the Provincial High School report card in Ontario (http://www.edu.gov.on.ca/eng/document/forms/report/sec/not1e.pdf). Another excellent example is the Grade 7 and 8 report card used by the Winnipeg School Division in Manitoba (Figure 2-1). This district reports on six aspects of behavior for all students in each subject and the rubric for four levels of performance appears on the report card.

Reporting achievement separately from behaviors means that everyone can know as accurately as possible what a grade means in achievement terms. Another benefit of expanded format reporting is that it enables a school/community to show very clearly and forcefully which behaviors it values in students. Some states and many schools have articulated such statements. For example, the state of Hawaii has identified six General Learner Outcomes (GLO's) and reports on these for all students. See Figure 2-2 for an example of how teachers could record evidence for evaluating students on the basis of each GLO.

As you read and think about Fixes 2 through 6, keep in mind that they are subsets of the larger issue of separating behaviors from achievement. They each address specific behaviors that lead to inflated or reduced student grades, both inaccurate measures of achievement.

Student Involvement

Students benefit from frequent opportunities to identify both the behaviors that help and those that hinder their achievement. They also can self-assess their achievement and behaviors and set goals for both. In furtherance of this, teachers can identify the components of desired behaviors and help students to develop specific goals. For example, students often hear that they "need to improve their effort," but *effort* may seem a vague concept. To help clarify this concept for students, teachers can identify the components of good effort, such as persistence, striving for accuracy, time on task, and trying alternate methods, which students then focus on to identify their relative strengths and weaknesses.

As educators, our beliefs and practices about motivation will have great impact on students. Students who have a sense of control because they know they are free to choose, and who receive frequent descriptive feedback instead of rewards and punishment linked to their behaviors, are much more likely to exhibit the desired behaviors and to value the separation of achievement and behavior.

Figure 2-1 Winnepeg (MB) Schools Grade 7–8 Report Card

The Winnepeg School Division

Student Effort & Behaviour Report
Grade 7–8

Home Room Teacher: **Student Name:**
Grade: 07

KEY TO TERMS	Excellent 4	Good 3	Needs Improvement 2	Unacceptable 1
Organizational Skills	Consistently sets goals. Collects and organizes information and uses time effectively.	Usually sets goals, collects and organizes information and uses time effectively.	Frequently needs assistance in setting goals, organizing information and using time effectively.	Rarely sets goals, information disorganized and frequently wastes time.
Homework	Consistently completes homework.	Usually completes homework.	Frequently does not complete homework.	Rarely completes homework.
Assignments	Consistently brings materials and completes assignments.	Usually brings materials and completes assignments.	Frequently needs to be reminded to complete assignments.	Rarely brings materials and completes assignments.
Citizenship	Consistently respects the rights of others.	Usually respects the rights of others.	Frequently needs teacher guidance in appropriate behaviour.	Rarely demonstrates respectful behaviour.
Teamwork	Consistently participates well in class/group activities.	Usually participates well in class/group activities.	Sometimes participates well in class/group activities.	Rarely participates in class/group activities.
Interpersonal Skills	Consistently resolves conflict in constructive manner.	Usually resolves conflict in constructive manner.	Frequently needs reminder on how to resolve conflict.	Rarely resolves conflict appropriately.

ART 5 — Nixon, C

Term	1	2	3	4
Organizational Skills	3			
Homework	4			
Assignments	4			
Citizenship	3			
Teamwork	3			
Interpersonal Skills	3			

ENGLISH 7 — Palerum, U

Term	1	2	3	4
Organizational Skills	3			
Homework	3			
Assignments	3			
Citizenship	3			
Teamwork	3			
Interpersonal Skills	3			

MATH 7 — Nixon, G

Term	1	2	3	4
Organizational Skills	3			
Homework	3			
Assignments	3			
Citizenship	4			
Teamwork	4			
Interpersonal Skills	3			

MUSIC 7 — Drumstone, G

Term	1	2	3	4
Organizational Skills	3			
Homework	3			
Assignments	3			
Citizenship	3			
Teamwork	3			
Interpersonal Skills	3			

PHYS ED 7 — Pauls, T

Term	1	2	3	4
Organizational Skills	3			
Homework	3			
Assignments	3			
Citizenship	3			
Teamwork	4			
Interpersonal Skills	3			

PRACT ARTS 7 — Practical Arts Churchill

Term	1	2	3	4
Organizational Skills	3			
Homework	3			
Assignments	3			
Citizenship	3			
Teamwork	3			
Interpersonal Skills	3			

SCIENCE 7 — Nixon, C

Term	1	2	3	4
Organizational Skills	3			
Homework	3			
Assignments	3			
Citizenship	3			
Teamwork	3			
Interpersonal Skills	3			

SOCIAL STDY 7 — Pauls, T

Term	1	2	3	4
Organizational Skills	4			
Homework	4			
Assignments	4			
Citizenship	3			
Teamwork	4			
Interpersonal Skills	4			

Source: Copyright © 2002 Winnipeg School Division, Winnipeg, MB. Reprinted by permission.

Figure 2-2 State of Hawaii General Learner Outcomes

Student:

	Achievement Evidence										
Assessments →											**Summary**
GLO's ↓											
Self-Directed Learner											
Community Contributor											
Complex Thinker											
Quality Producer											
Effective Communicator											
Effective and Ethical User of Technology											
Comments:											

Note: A complete rubric for the General Learning Outcomes appears at http://doe.k12.hi.us/standards/GLO_rubric.htm

Summary

Grades are broken when they mix achievement and non-achievement elements. The fix is to report variables such as behaviors separately from achievement, thereby ensuring that the grades reflect student achievement as accurately as possible.

> A grade should give as clear a measure as possible of the best a student can do. Too often, grades reflect an unknown mixture of multiple factors. . . . How effective is such a communication system? The problem transcends individual teachers. Unless teachers throughout a school or district completely agree on the elements and factor them into their grading in consistent ways, the meaning of grades will vary from classroom to classroom, school to school.
>
> *—Tomlinson and McTighe, 2006, p. 133*

Don't reduce marks on "work" submitted late; provide support for the learner.

> Teachers turn things in late all the time, as do workers in every profession. The idea that "You can't get away with turning work in late in the real world, mister" isn't true.
>
> —*Wormeli, 2006, p. 148*

Grades are broken when they include penalties for student "work" submitted late. Penalties distort the achievement record the grade is intended to communicate, can actually harm student motivation, and for many students do not result in changes in behavior. The fix is to not use penalties and to set up support systems that reduce or eliminate the problem of late work.

It is critical to emphasize that we want students to exhibit responsibility and submit assessment evidence in a timely manner. The difficulty we face is, what do we do when students do not demonstrate these qualities? What policies and procedures are most likely to get them to learn as much as possible and exhibit the desired behaviors? Traditionally, we have used penalties such as a reduction of one letter grade or of a number of points for each day a required piece is late.

Many teachers believe that they need a policy with penalties to attempt to ensure that students turn in work on time so the teacher can maintain the pace of instruction necessary to meet tight curricular requirements. Many also use penal-

ties because they believe that it communicates fairness to students: everybody gets the same amount of time. There are, however, at least four problems with this practice. First, and most damaging, they distort the grade's representation of the student's true achievement. Second, they can motivate exactly the opposite behavior than that intended. At some point in the grade reduction scenario, accumulating penalties lead students to conclude that it no longer makes sense to do that work. If it is an important piece of assessment evidence, it is better that the student submit it late than not at all. Third, my own classroom teaching experience and anecdotal evidence from many teachers leads me to conclude that penalties don't work because they do not change behavior—the same student who is late with required evidence in week 2 is frequently late in weeks 18 and 36. Fourth, having absolute deadlines (and penalties) for everything does not prepare students for the world beyond school. In the "real world" timelines are frequently negotiated (real estate, legal matters) or adjusted to circumstances (contractors and consultants); deadlines range from fixed to considerably flexible. (Ironically, "You can't deliver work late in the real world" is the very reason some teachers tell students they have the policy!) We prepare students better for that world when we offer a variety of deadlines in school; work part of an instructional sequence needs to be done tonight for tomorrow, but timelines for long-term assignments might be framed more flexibly.

Furthermore, in the world beyond school, as adults, if we are not able to meet a timeline, we often can communicate with the person/institution to whom we are responsible, arrange a new mutually agreeable timeline, and then work to meet it. This is the responsible, adult behavior that we need to encourage in students and we do this by allowing

them to request extensions. This is preferable to students "hiding in the back corner" as they often do when they have late or missing assessment evidence. If we want students to be responsible and timely, then we can teach them and help them along the way, rather than assume they will learn the lessons through punitive policies.

Again, there is no suggestion here that teachers should condone or ignore lateness in submitting required evidence. Teachers should keep records of students' timeliness and report on this behavior in expanded format report cards. They also can assign consequences as they would for any other unacceptable classroom behavior. Direct parent contact may also be necessary, especially if the lateness is chronic. Students who are late with important assessment evidence could be required to come in before school, at lunchtime, or after school where they will receive both the assistance and time they need. This is similar to the approach suggested by Rick DuFour (DuFour, Eaker, DuFour, and Karhanek, 2004), whose "pyramid of interventions" to help students succeed moves from "limited and voluntary" to "significant and compulsory."

The consequences for submitting required assessment evidence late should be as positive and supportive as possible, although some "negative" consequences, such as detention, may be warranted for repeated or chronic lateness. Supportive approaches do not distort achievement or motivation and more closely mirror practices in the world beyond school. Support should also include identifying at the beginning of the school year students who are organizationally challenged and providing them structure in assignments.

The most appropriate fix for grades is to not use penalties at all. Some teachers (and parents) will see the emphasis on support and communication suggested here as too "soft."

Thus, as we make the transition from traditional to standards-based practices, it may be both acceptable and necessary to use small penalties that do not distort achievement or motivation; that is, penalties that are more apparent than real. One example of this approach is that students who submit required assessments late receive the grade level they "earned" but it is recorded at the lowest form of that level (e.g., a student submits an "A+" paper several days late so the grade is recorded as "A–"). Using a grading scale with percentages linked to letter grades, it is easy to record the lowest percentage at the level earned—e.g., if an A is 90 to 100% the maximum reduction for evidence judged as A quality would be to 90%.

The principle that should be applied to late work is to separate achievement from behavior and communicate both to those who have a right to know about the student. If Rory is a brilliant writer who always hands assignments in late, both aspects are hidden if she gets a C or a D. But if she gets an A and the report says, "Brilliant writer, but always late," then we have accurate information. A daily newspaper or an advertising agency may not want to employ Rory but she may be perfect to write features in a monthly magazine or as a novelist or playwright.

Student Involvement

Students should have input into decisions about timelines for required assessment evidence because when they have input they have ownership, and ownership frequently leads to meeting timelines. As noted, if a student is not able to meet a timeline, the teacher should not use mark penalties, but should encourage the student to acknowledge the lateness and request an extension and/or suggest other appropriate consequences.

Summary

Penalties distort achievement and motivation, and in my experience are generally ineffective. The fix for late student work is a positive, supportive approach that directly affects student behavior, leaving the scores and the resulting grades as pure measures of achievement.

> The appropriate consequence for failing to complete an assignment is completing the assignment. That is, students lose privileges, free time, and unstructured class or study hall time, and they are required to complete the assignment. The price of freedom is proficiency, and students are motivated not by threats of failure but by the opportunity for greater freedom and discretion when work is completed accurately and *on time.*
>
> *—Reeves, 2006, p. 122, emphasis added*

Don't give points for extra credit or use bonus points; seek only evidence that more work has resulted in a higher level of achievement.

> Recently it was "Dress like an Egyptian Day" at my school. If we dressed like an Egyptian we got extra credit. When we didn't (which the majority of the kids didn't) our teacher got disappointed with us because we just "didn't make the effort." . . . One of the most frustrating things in my mind is that we get graded on something that has no educational value. I would very much like to discontinue these childish dress-up days.
>
> —*Starsinic, 2003, n.p.*

Extra credit and bonus points can distort a student's record of achievement—grades are broken as a communication tool if we give points for "dressing like an Egyptian" when such "performances" do not demonstrate achievement of specified academic standards. It is obvious in the quotation that the writer, a high school senior, understands this but that her teachers do not. The fix for this is to not use extra credit or bonus points. If students want to get higher grades teachers can require them to provide "extra" evidence that demonstrates a higher level of achievement.

Over the years I have heard of an amazing array of extra credit activities including cleaning blackboards, bringing in classroom supplies, supplying food for the food drive, or bringing a Mexican dish for the Spanish class. My favorite

story was from someone who said her high school Physics teacher believed very appropriately in students identifying examples of physics in the "real world." He provided them with a "worksheet" with six questions, and for each worksheet they completed 1 percent was added to their grade. Her final grade in Physics was 91 percent and she did 60 extra credit sheets! It is interesting, is it not, to speculate on her level of physics knowledge? One high school Science department even has an "Extra Credit Counter" (like a counter or merchandise display in a store) for each course on the school's website. I have also heard many stories about the availability of bonus points on tests and exams so students finish with grades of 110 percent!

The basic problem with weaving extra credit and bonus points into a grade when they reflect something other than the expected learning is that they distort the record of achievement. Extra credit and bonus points stem from the belief that school is about doing the work, accumulating points, and that quantity is the key—with more being better—rather than about achieving higher levels of learning. But in standards-based systems the main issue should be having enough quality evidence to accurately determine each student's achievement. Extra credit and bonus points come from a culture that emphasizes extrinsic motivation. As with other nonachievement factors that find their way into the grade, they have frequently been used to manage student behavior.

Students should, of course, be able to provide additional evidence of their understanding, knowledge, and/or skill. However, this additional evidence must reveal new or deeper learning—and should be considered along with the previous evidence to determine the student's level of achievement. For example, if previous evidence was a mixture of the

achievement levels of "competent" and "approaching competency" and a student's additional evidence was all "competent," this would allow the teacher to justify assigning this student a final achievement level of "competent" with the appropriate letter grade. (See Fix 8 for more on levels of achievement and performance standards.)

The shift in thinking is illustrated in the following example. Imagine that a student receives the following scores for a series of tests and assignments:

> 5/10, 66/100, 39/50, 27/35, 37/50, 8/10,
> 15/20, 20/25, 8/10, 75/90

The total would be 300/400, and if the grade were calculated as a mean in the traditional way the grade would be 75 percent, which in most schools/districts would be a grade of C. The student then completes three extra credit assignments (which may or may not be in any way related to the learning goals) for which he receives scores of 14/20, 7/10, and 3/5. The total is now recorded as 324/400 (although it is really 324/435) with a mean of 81 percent, so the student receives a final grade of B, which is an inflated grade.

Now imagine that a different student has a teacher who is truly standards based; this teacher records scores as proficiency levels, with 3 as proficient (meets the standard). The scores this student receives on a series of tests and assignments are 1, 2, 2, 1, 2, 2, 2, 2, 3, and 3. The mean, median, and mode for these scores are the same—2—so this student would normally receive a grade of C. The teacher, however, notes that the two more recent scores are 3s so asks the student to provide extra evidence on specific learning goals to see if this is now her level of achievement. The student receives 3, 4, and 3, which shows she is now proficient, so her final grade is a well-deserved B.

Student Involvement

Through self-assessment and teacher communication, students can acquire a clear sense of their level of achievement. If it is less than proficient, or lower than they (or their parents) are willing to accept, teachers can offer students opportunities to provide additional evidence. It must be clear that this will not result merely in points being added to a total. If students are able to show that they now know, understand, or can perform at a higher level, their grade must reflect this. At minimum, students should be partners in identifying appropriate evidence of additional learning, making suggestions about what they will do to show a higher level of achievement. For some it may be a traditional test, for others it will be a product, for still others it will be a performance or a personal communication such as an interview or oral exam. If they have participated appropriately in student-involved assessment they will make the right choice(s).

Summary

Grades are broken when teachers provide extra credit or bonus points that are just about more points, not about higher levels of proficiency. The fix is to eliminate extra credit and bonus points that do not relate to achievement and to communicate clearly to students and parents that better grades come from evidence of higher levels of performance, not just from more points.

> Some teachers add "extra credit" points to the total scores. . . . This does a disservice to students when their test scores rightly show that they did not learn certain key concepts and skills and the extra credit

tasks do not help the students to master those concepts and skills. Sometimes the extra credit work is barely, or not at all, related to the key concepts and skills that are supposed to be the basis of the grade. Not everyone agrees with my position, but I believe it is logical and fair to students.

—*Carr, 2000, p. 53*

FIX 4

Don't punish academic dishonesty with reduced grades; apply other consequences and reassess to determine actual level of achievement.

> No studies support the use of low grades or marks as punishments. Instead of prompting greater effort, low grades more often cause students to withdraw from learning.
>
> —*Guskey and Bailey, 2001, pp. 34–35*

"You cheated, so you get a zero on this test (assignment, etc.)." This has been the typical response to the discovery of academic dishonesty. It is another example of broken grades because it uses the assessment/grading policy as a tool to discipline students for inappropriate behavior, thus distorting student achievement. The fixes for this are to articulate an academic honesty policy with clear behavioral consequences for breaches and to require students to redo the test or assignment without cheating or plagiarizing, to establish an accurate achievement record for grading.

Academic dishonesty is an ongoing problem in middle and high schools and colleges. Dealing with it is often difficult; probably like many of you, I have been part of very emotional arguments about it. As with most behavioral concerns there are two main issues—how to prevent it, and what to do about it when it happens. Most schools try to deal with both of these issues together by having punitive policies that range from zeros on the assignment to loss of credit to expulsion.

These policies arise from the belief that if the punishment is sufficiently severe then students will not risk being caught. Continuing academic dishonesty, however, points to the need for viable alternatives. It is perhaps best to begin with how to prevent it, then develop procedures to deal with it when it happens.

"Prevention is better than cure" is an old but true saying, and it certainly applies here. Tom Solyom, an assistant principal, and teacher-librarian Dawn Keer at Archbishop Macdonald High School in Edmonton, Alberta, have led the development of a policy aimed at decreasing cheating. They believe that teachers must make their expectations clear and explicit and should talk about academic integrity with their students to help them understand why it is so important. They also believe that teachers should not assume that students understand exactly what they mean by the terms *plagiarism* or *cheating*.

The policy statement at the school provides the following "Definitions of Inappropriate Academic Behavior/Academic Misconduct":

> *Plagiarism:*
> Submitting the words, ideas, images or data of another person's as one's own in any academic writing or other project.
>
> *Cheating:*
> a) Possession of unauthorized material,
> b) Substantial editorial or compositional assistance,
> c) Submission of another student's material already graded for credit,
> d) False claims or fabricated references,
> e) Copying off of someone else's exam and/or quiz; or passing answers from a quiz or exam to another student.
>
> (Archbishop Macdonald High School, 2006, p. 27)

The faculty was also provided with a list of "Tips for Preventing Plagiarism and Cheating," that included the following (remember, it is a high school!):

> Assign essay topics that are specific to your course and timely in nature.
>
> Give clear guidelines for format.
>
> Provide bibliography resources. (Websites, guidebooks, etc.)
>
> Give your class an example of a plagiarized paper and have a discussion about it.
>
> Use in-class writing assignments.
>
> Set assignments where the objective is to critique websites, thus avoiding the temptation for students to copy them.
>
> Change your exams every term and/or use alternating formats so students next to each other are not writing identical exams.
>
> Proctor your exams. During mid-term and final exams, make sure there are enough supervisors (proctors) for the number of students.
>
> Where possible arrange students with a seat in between them. (This is obviously difficult with large classes.)
>
> Provide scratch paper.
>
> Whenever possible, use long answer/essay format. (Math could have more open-ended questions.)
>
> Do not allow students access to back packs or their coats during an exam.
>
> Be aware of technologies that could allow cheating, e.g. calculators, cell phones, pagers, etc.
>
> Be explicit about possible sanctions. Have students refer to the Academic Honesty and Integrity Policy in their agenda books.
>
> (Tom Solyom, personal communication, July 2006)

When academic dishonesty is suspected students can be interviewed privately in an attempt to determine whether the transgression was inadvertent or deliberate. If it was inadvertent the student may be counseled and may revise the work as appropriate; school/district policies generally dictate the response to deliberate dishonesty.

Effective policies first and foremost recognize that academic dishonesty is very serious inappropriate behavior equivalent to theft, and as such requires primarily behavioral consequences. These policies also recognize that academic dishonesty deprives everyone of quality evidence of student achievement. The appropriate assessment consequence is to have students redo the work with honesty and integrity.

The Archbishop Macdonald High School policy has the following possible sanctions:

> The grade coordinator in conjunction with the teacher, in whose class the offence occurred, has the authority to impose one or more of the following sanctions.
>
> ***Plagiarism and/or Cheating:***
> A student's academic misconduct will be confidentially communicated to all of his/her teachers.
>
> At the teacher's discretion the students may be required to do another assignment/exam submitting their own original work for grading purposes.
>
> The student must complete the exam/assignment on his or her own time (outside regular class time).
>
> A zero may be awarded for that particular assignment/ exam. In this instance, parents must be informed that the zero is being assigned as a punishment for inappropriate academic conduct and does not represent a true assessment of the student's ability. Assessment is a snapshot of performance, not potential.

> All extracurricular involvement will be suspended until this sanction is lifted at the discretion of the disciplinary committee.
>
> Probation—The probation period will last the remainder of the school year. If a student is discovered cheating and/or plagiarizing a second time during this period further sanctions will be applied.
>
> In addition to the above sanctions, the grade coordinator, in conjunction with the school principal, has the authority to impose one or more of the following sanctions.
>
> Suspension
>
> Expulsion
>
> (Archbishop Macdonald High School, 2006, pp. 26–27)

This sanction list is presented as an example, not as a model; note that the main behavioral consequence is suspension from all extracurricular involvement. This would be a significant consequence for participating students but not for nonparticipants. The policy could be reworded in the following way to avoid this potential problem:

1. (Entry 2)—Require students to redo the assignment/exam with the stated reason being "the provision of accurate evidence of achievement," not "grading purposes." This should not be at each teacher's discretion and probably should be done with supervision in or out of class time.
2. (Entry 4)—Do not include the option of zero. I would also prefer the use of the words "understanding, skill and/or knowledge" to replace "ability" and the last sentence should read, "A single assessment is a snapshot of performance, not a judgment of achievement."

3. (Entry 5)—For offences in the last month or two of the school year, probation probably should be extended for returning students for a period of time into the next school year.

Another important aspect of any school policy that involves judgment is that there must be an appeal process. There is such a process in the Archbishop Macdonald policy:

> APPEAL PROCESS
> Any student has the right to appeal the charges and/or sanctions determined by the teacher and grade coordinator within 1 week. The student will meet with the Appeal Board Committee, which will consist of the Principal, the subject Department Head, a counselor and, if requested, the student's parents. The student must fill out the appropriate form. The parents will be informed. (Archbishop Macdonald High School, 2006, p. 27)

Student Involvement

One way to involve students in academic honesty is for schools to have clear policies and to have frequent age-appropriate discussions about what this means, using specific examples. I have heard of some schools that have an honor code that requires students to attach a statement to all assessments that they have not cheated or plagiarized. Obviously this would not prevent academic dishonesty, but students who have to reflect on this issue for each assessment are likely to develop a clearer understanding of what academic honesty requires.

Summary

Academic dishonesty is unacceptable and must not be tolerated. But grades are broken if the response to cheating is a lowered score or grade, because this renders inaccurate the student's record of achievement. The fix is to remove grading as the vehicle for assigning a consequence to students who cheat, and to have an academic honesty policy that clearly describes inappropriate practices and the consequences for breaches. To emphasize that the learning is most important, the policy would also require that students must redo any assessment that involved academic dishonesty—without cheating or plagiarizing.

> Don't use grades punitively. . . . Without exception, experts in the area of student grading recommend that grades not be used in a punitive sense. When a teacher uses grades as punishment for student behaviors, the teacher establishes an adversarial relationship in which grades are no longer meaningful to students as indicators of their accomplishment. The punitive use of grades only increases the likelihood that students will lose respect for the evaluation system; consequently the appeal to students of subverting such a system will be heightened.
>
> —*Cizek, 2003, p. 100*

FIX 5

Don't consider attendance in grade determination; report absences separately.

> Excused and unexcused absences are not relevant to an achievement grade. There is no legitimate purpose for distinguishing between excused and unexcused absences. For educational purposes, therefore, there need only to be recorded absences.
>
> —*Gathercoal, 1997, p. 151*

Grades are broken when they are directly or indirectly related to a student's attendance record. The simple fix requires absences be reported separately from grades, and that grades be determined only from evidence of achievement.

Most teachers would probably agree that all students should attend school regularly. Most students need to do so to be successful in their learning. However, standards-based learning is not about seat time. It is about what students know, understand, and can do. Grades should be accurate reflections of that and that alone. Attendance therefore is best recorded and reported separately simply as days present (or days absent).

It is common for schools/districts to go to great lengths to distinguish between excused and unexcused absences, with the difference having a significant impact on grades and the ability of students to "make up" for absences. The distinction between excused and unexcused absences may be very important for behavioral and legal reasons, but it is irrelevant

to learning and assessment perspectives. From these perspectives, the only issue is whether students learn and demonstrate it by providing appropriate assessment evidence. Students who have been absent require opportunities to learn what they have missed and subsequently to demonstrate what they know, understand, and can do, regardless of the reason for the absence.

Another reason it is inappropriate to make this distinction is that the difference often depends on the "creativity" of the parent(s) or the student. This is a polite way of saying whether parents or students are willing to lie, which is obviously an inappropriate basis for decisions affecting grades or "makeup" opportunities.

In some of the schools in which I taught it was a common practice (especially in physical education) to include a fixed number of points for attendance and to deduct one or two points for each absence. In subjects such as physical education, drama, and music active participation is essential, but with such a procedure a student would have zero for attendance after 10 absences even though they attended 35 out of 45 days. Such procedures are illogical, distort the meaning of the grade, and should not be permitted.

One aspect of attendance and grades that presents a real dilemma is when there are requirements that students attend certain out-of-school activities or performances, such as concerts in a music course or performances in a drama course. When participation is required for the program to function, however, it is reasonable to suggest that there should be consequences for failure to attend. Policies that state that students receive a failing grade if they miss performances are inappropriate. A student who is proficient or better in all the music or drama goals would have their grade distorted if it was lowered for failing to attend one or

more performances. The best approach in standards-based schools is probably a behavioral consequence, but I acknowledge that it is difficult to determine what is appropriate. As schools/districts make the transition to a true standards base it may be necessary to establish a policy that students must attend X (or all) concerts/performances to receive a grade/credit, and that they will otherwise receive an Incomplete. This policy can be communicated in writing to students and parents at or before the start of the course.

Summary

In standards-based systems all marks and grades (pass/fail, A/B/C/D/F, etc.) should be determined by proficiency, not by seat time. Most students need to attend class to be successful and teachers must ensure that engaging learning activities are being provided so that students feel it is worth their while to attend. However, absences should not directly affect students' grades. Grades are broken if there is a direct impact because a behavioral variable is being allowed to distort achievement. The fix is to deal with attendance separately from achievement by simply reporting days present (or days absent).

Teacher: "Are you telling me that if a student has been ill and another has been skipping, that they both should be able to make up the work missed?"

Gathercoal: "(Yes) both needed an educator when they returned, perhaps the one who skipped more than the other."

—*Gathercoal, 1997, p. 151*

FIX 6

Don't include group scores in grades; use only individual achievement evidence.

> Group scores [grades] are so blatantly unfair that on this basis alone they should never be used.
>
> —*Kagan, 1995, p. 69*

Grades are broken when they include group scores from work done in cooperative learning groups. The fix is to ensure that all evidence used to determine grades comes from individual evidence of achievement.

Cooperative learning is a very powerful teaching/learning strategy; done well and used appropriately it can lead to significant learning gains and improve attitudes about learning and school. But frequently in cooperative learning situations students are required to produce a group product or presentation for which they receive a group score, which is then recorded for each member of the group. This is an inappropriate practice, as illustrated in the "For Better or Worse" cartoon in Figure 2-3.

In Figure 2-4, Spencer Kagan provides seven specific reasons for opposing group scores (grades). His first four reasons are clearly illustrated in Figure 2-3. The situation depicted is obviously unfair, as one student is receiving "credit" for something she didn't do; report cards will be "debased" because these students will receive inaccurate grades; this situation would undermine motivation because the next time these students will feel that their effort is of

Figure 2-3 An Example of an Inappropriate Group Scoring Practice

Source: FOR BETTER OR FOR WORSE © 1996 Lynn Johnston Productions. Dist. By Universal Press Syndicate. Reprinted with permission. All rights reserved.

Figure 2-4 Kagan's Seven Reasons for Opposing Group Scores (Grades)

Group scores (grades)

1. Are no(t) fair
2. Debase report cards
3. Undermine motivation
4. Convey the wrong message
5. Violate individual accountability
6. Are responsible for resistance to cooperative learning
7. May be challenged in court

Source: The data in Figure 2-4 are adapted from "Group Grades Miss the Mark," by S. Kagan, 1995, *Educational Leadership*, *52*(8), pp. 69–70. Adapted by permission of Association for Supervision and Curriculum Development.

dubious value; and this group score sends the wrong message about the purpose and value of teamwork.

But the two most important reasons why group scores should not be used as part of student grades are reasons 5 and 6. With regard to reason 5, many models of cooperative learning (see, e.g., Gibbs, 2000; Johnson and Johnson, 2004;

Kagan, 1995) have individual accountability as a basic principle in the model. Group scores that become part of individual grade determination violate this principle, meaning the cooperative learning model is being implemented incorrectly. Regarding reason 6, not surprisingly, cooperative learning has encountered parental and student resistance in some schools/districts precisely because of group scoring. In the extreme, parents have taken teachers, principals, schools, and districts to court over this issue. The parents generally have won because judges followed the principle that no student's grade should depend on the achievement (or behavior) of other students. Cooperative learning can be a powerful teaching/learning strategy. We want to help students to be successful learners so we need to have all such powerful strategies available. We do not want to impair any strategy's effectiveness by incorrectly measuring the achievement of students who use it.

There is yet another issue with giving scores for products or performance developed in cooperative learning groups. The strategy is cooperative *learning*, which implies that any activities that occur in groups are learning activities and any assessment of them is best considered formative assessment—to help students improve their knowledge, understanding, and skill(s). Such assessment is for practice and should not produce scores that are part of grade determination. (This issue is the subject of Fix 13.)

Summary

Grades are broken if they involve the use of group scores from cooperative learning or group activities. This is so because the group scores may not accurately reflect the achievement of each student and therefore would be unfair for some members of the group. This problem can be addressed by

recognizing that cooperative learning is essentially a learning activity, *not* an assessment tool. After a class has experienced cooperative learning teachers can then assess students individually to find out what they know, understand, and can do as a result. This individual assessment could involve one or more of the following: "teacher monitoring of [cooperative] activity work; an essay response based on questions formulated during the activity; a class discussion of the questions and responses generated; [or a test] on the content of the questions formulated and responses generated" (Benevino and Snodgrass, 1998, p. 146).

> The assessment of individuals within groups begins with setting individual learning goals and involves such procedures as individual tests and products, observing students while working in groups, giving group members a questionnaire to complete, and interviewing group members during group sessions. There is a pattern to classroom life summarized as *"learn it in a group, perform it alone."*
>
> *—Johnson and Johnson, 2004, p. 53, emphasis added*

CHAPTER 3

Fixes for Low-Quality or Poorly Organized Evidence

FIX 7

Don't organize information in grading records by assessment methods or simply summarize into a single grade; organize and report evidence by standards/learning goals.

> The important thing is . . . that everyone in the [school or] district . . . can identify what it is that students are expected to learn.
>
> —*Butler and McMunn, 2006, p. 23*

Grades are broken when evidence of learning from multiple sources is blended into a single grade and the communication fails to show how successful students have been in mastering individual standards/learning goals. The fix is to base grades on published school/district/state standards (learning outcomes/goals, essential learnings, expectations, strands, etc.), and to report them for each standard to create a more complete profile of individual student strengths and weaknesses. Evidence also may be summarized into a grade, and often this is required. But the total communication must also and always report mastery by standard or by some categories derived from the standards.

This requires curriculum, instruction, assessment, and grading and reporting all to be organized around the standards. Many schools have successfully done so with curriculum and instruction, and increasingly assessment has become aligned as well. However, while many schools/districts have embraced standards-based grading and reporting at the

elementary level, there remains much work to be done, especially in middle and high schools. Schools focused on standards only for curriculum, instruction, and assessment are standards *referenced*, not standards *based*.

Traditionally, teachers have organized their evidence of student achievement for all learning goals either simply in the order collected over time or in categories based on the type of data, such as tests, projects, and homework assignments. For each collection, they then distill the individual grades into a single summary grade and report that grade. In either case, what is not recorded and therefore not reported is vital information revealing how well each student has mastered each learning goal. In other words, although each student's performance can be summarized with a single symbol/grade, this approach provides no basis for reporting direct evidence of student performance on each learning goal, unless accompanied by a narrative report that describes learning in relation to the written curriculum. To be standards based in grading, teachers plan each assessment to provide direct evidence of student proficiency on specific learning outcomes/goals and then record this evidence by goal, dedicating columns or blocks of space in their gradebook to each learning goal. Figure 3-1 shows a sample of such a gradebook using some of the State of Oregon reading standards for Grade 4. The standards are the basis for organizing assessments and collecting evidence from tests and performance assessments ("PA" in the figure) to determine an overall grade (if necessary) and a grade for each standard.

In this example the first test was not just 10 questions worth 2 points each with a single score recorded and reported as X out of 20; the test elicited information on two of the five standards shown, and our example student received scores of 15 out of 20 on understanding text read and 19 out of 20

on identifying key facts and information. The performance assessment provided level scores for three standards using a 5-level rubric (0–4). This gradebook format enables teachers and students (and parents) to see a profile of student performance that clearly identifies areas of strength and areas for improvement.

Figure 3-1 Summary of Evidence for Meeting State of Oregon Reading Standards

Student:

	Achievement Evidence										
Assessments → / **Standards ↓**	**9/1** Test	**9/8** PA									Summary
Read aloud grade level text											
Understand, learn and use new vocabulary											
Listen to, read and understand text	**15/ 20**	**4**									
Identify key facts and information	**19/ 20**	**3**									
Identify and analyze text that uses sequential order		**4**									
Comments:	*The standards used here have been selected and adapted from the Oregon Grade 4 reading strand.*										
										Report Card Grade	
										Most consistent level of achievement	

Note: The Oregon reading standards appear at http://www.ode.state.or.us/teachlearn/real/documents/05-06elagrade4.pdf

Often at the middle and high school levels, where teachers interact with a large number of students, teachers feel/believe that recording data at the learning goal level is impractical. An alternative (and transitional) approach is to use strands within a subject as the organizing structure. Figure 3-2 illustrates how such a standards-based gradebook might look for an individual student over a grading period. The top row of numbers is the date of the assessment. Each test (T) and performance assessment (PA) is recorded by strand. Some provide evidence of only one strand; other, more comprehensive assessments provide evidence on several strands. By the end of the grading period there are at least five scores for each strand, which is sufficient evidence to make summary judgments for that strand. To determine a summary grade for each strand all performance ratings must be recorded using a common scale, so test scores are recorded as a number of points out of a total, which is converted to the same 5-point scale used for the performance assessments. If a single summary grade for the subject is required this can be determined by identifying the most consistent level or calculating central tendency (mean, median, or mode). When students perform at the same level on all strands, the summary grade is easy to determine and has clear meaning. But if a student's mastery is inconsistent, then a summary grade is difficult to determine and will lack the detail needed to understand the student's real achievement; that is, their strengths and weaknesses. It is for this reason that I recommend that we always report information about the level of achievement on each standard, backed up by summary subject grades (if required).

These gradebook examples show one page for each student because this is the best way to illustrate this Fix and it is also the best way to collect evidence of student achieve-

Figure 3-2 Summary of Evidence for Meeting State of Florida Mathematics Standards, First Grading Period

Student:

	Achievement Evidence												
Assessments → / **Strands ↓**	8/13 Test	8/20 PA	8/23 PA	8/24 PA	8/25 Test	8/30 PA	9/5 Test	9/8 PA	9/12 Test	9/19 PA	9/21 PA	9/12 Exam	SUMMARY
Number Sense, Concepts and Operations			2		11/20 (1)		16/20 (3)			2		7/10 (2)	C
Measurements	19/20 (4)	4			18/20 (4)			4			4	16/20 (3)	A
Geometry and Spatial Sense		4		2		3	17/20 (3)				3	10/10 (4)	B
Algebraic Thinking	11/20 (1)	2				2				2		14/20 (2)	C
Data Analysis and Probability		1		2		3	20/20 (4)			4		19/20 (4)	A

Comments:

Overall Grade	A

Note: The Florida Mathematics standards can be found at http://www.firn.edu/doe/curric/prek12/pdf/math9.pdf

ment. However, teachers interacting with a large number of students or teaching several subjects to a smaller number of students may find one page per student to be impractical, and may prefer gradebooks such as the one shown in Figure 3-3.

Figure 3-3 shows part of the gradebook that Glenda Greier uses in her Grade 3 and 4 classes in Bay District Schools in Panama City, Florida. There is a column for two of the math

Figure 3-3 Mrs. Greier's Gradebook

Mrs. Greier's Grade book

2005–2006

Math	Number Sense, Concepts, & Operation										Measurement									
	Formative					Summative					Formative					Summative				
Date																				
SSS/GLE																				
Methods																				
Students																				
1.																				
2.																				
3.																				
4.																				
5.																				
6.																				
7.																				
8.																				
9.																				
10.																				
11.																				
12.																				
13.																				
14.																				
15.																				
16.																				
17.																				
18.																				
19.																				
20.																				
21.																				
22.																				
23.																				
24.																				
25.																				
26.																				
27.																				
28.																				
29.																				
30.																				

Source: Copyright © 2005 by Glenda Greier. Reprinted by permission.

strands on this page. Two additional pages are used for the other math strands, with space for each student in her class on each page. Note also that Mrs. Greier records separately information from formative and summative assessments (Fix 13 covers this issue in detail).

Another approach that may work well for some teachers and subjects is a blending of the two approaches described here. For example, there are five strands in the U.S. foreign language standards: (1) Communication, (2) Cultures, (3) Connections, (4) Comparisons, and (5) Communities (*Standards for Foreign Language Learning*, n.d.). Communication has three standards: 1 requires students to engage in conversations; 2 requires students to understand and interpret written and spoken language; and 3 requires students to present information, concepts, and ideas to an audience of listeners or readers. There is a distinct likelihood that some students would achieve very differently on each of these. So for effective communication and instructional decision making, a foreign language teacher records information about achievement on each of these standards separately, but it may be sufficient for the other strands to collect and report information at the strand level. The level of specificity at which teachers collect evidence depends on the nature of the learning goals, the specificity of reporting required, and the teacher's beliefs about what is both necessary and possible. (Examples of gradebooks at different levels of specificity appear in Stiggins et al., 2004, pp. 289 and 290.)

Many, maybe even most, teachers now use computer grading software to (help) manage evidence of student achievement; almost all computer grading programs can be used for standards-based grading because these programs rely on "bins," or categories. Teachers have generally made those bins tests, projects, homework assignments, and so on,

but they may just as easily be standards or strands. The only limitation is the number of bins or categories allowed by the program.

So, wherever teachers are on the technological continuum, from hardcopy paper gradebook to computer software to using one's own spreadsheet, they can put this Fix in place. There may be a lot of work involved at first to get organized to record scores and determine grades in this way. However, teachers then find that assessment and grading are easier to organize, as they are "working smarter, not harder," using the same organizing structure right through the process and not using one structure for curriculum, instruction, and assessment and a different structure for grading and reporting.

Summary

Grades are broken when they are not directly based on standards and do not give information about achievement of standards. Fixing this requires the use of standards-based curriculum, instruction, and assessment, and collecting and reporting student achievement by standards.

> The principal limitation of any grading system that requires the teacher to assign one number or letter to represent course learning is that one symbol can convey only one meaning. . . . One symbol cannot do justice to the different degrees of learning a student acquires across all learning outcomes.
>
> —*Tombari and Borich. 1999, p. 213*

FIX 8

Don't assign grades using inappropriate or unclear performance standards; provide clear descriptions of achievement expectations.

> Performance standards specify "how good is good enough." They relate to issues of assessment that gauge the degree to which content standards have been attained. . . . They are indices of quality that specify how adept or competent a student demonstration should be.
>
> —*Kendall and Marzano, 1997, pp. 16–17*

Grades are broken when they are determined using poorly defined performance standards, such as letter–number relationships (A = 90–100, B = 80–89, etc.), that have traditionally masqueraded as performance standards. The fix is to develop clear and rich criterion-referenced descriptions of a limited number of levels of achievement. Whatever symbols are used to summarize student achievement (e.g., A B C D F; 4 3 2 1; E M N U), each level must be described clearly, with the level considered "good enough" (i.e., competent, proficient, mastered) to justify the assignment of each grade clearly labeled.

The challenge is to create clear descriptors of our overall levels so that we have a delineated achievement continuum with which we can identify when to judge student achievement to be competent or to deserve a certain grade. The range of options is almost unlimited for representing

achievement across grades and subjects. For instance, Guskey (2004) illustrates that we can consider levels of understanding/quality, levels of mastery/proficiency, frequency of display, degree of effectiveness, or evidence of accomplishment. Within each he suggests alternative progressions such as Below Basic, Basic, Proficient, and Advanced for levels of mastery and Rarely, Occasionally, Frequently, and Consistently for frequency of display. Marzano (2006, pp. 56–58) provides another example and a useful discussion of the issues involved in determining performance standards.

The judgments made when developing the descriptors and when evaluating student work are always subjective. These are not matters of learning science, but are common communication conventions. As long as everyone involved accepts that those who have developed the descriptors and levels are qualified, and understands the terms used, we can communicate effectively.

The best performance-standard setting pools the collective experience of a number of educators who are knowledgeable and experienced. When teams of teachers set standards, not only can meaningful performance standards result, but teams also can develop the basis for communicating both the standards and continua in ways that all concerned will understand, including developing student- and parent-friendly versions of the performance standards. Once developed, the resulting depiction of academic success will be published and public for all—administrators, teachers, students, parents, and others—to see from the beginning of instruction. Given that states and provinces have content standards, this is best done at the state/provincial level, but if it has not been, it should be done at the district level. Once the performance standards are in place, teachers need frequent opportunities for professional dialogue about them so they develop shared understanding and apply them consistently.

In a pure standards-based system we would have two levels of performance only—proficient and not proficient. Some schools/districts use only these two levels, but this is rare. Most commonly, performance continua use at least four levels: (1) a level above proficient, to recognize (and encourage) excellence; (2) proficient; (3) below proficient but acceptable; and (4) significantly below proficiency, or insufficient. Once we agree on the number of levels we determine the characteristics of each level and label and describe them clearly and concisely. As much as possible the language chosen should be descriptive, not judgmental. Figure 3-4 shows one example (see Arter and Chappuis, 2006, for a variety of other examples).

It is important to note that the terms used in Figure 3-4 describe the quality of achievement on the learning goals in the public, published curriculum in terms of the knowledge and skills demonstrated at the time of the report card. There are two important aspects to this—the meaning of quality and the timing. With regard to the former, as noted previously, Guskey (2004) illustrates very clearly that we have alternatives. With regard to the latter, Figure 3-4, for example, states directly that the levels describe performance at the time of the report card. The alternative is to consider the standard to be the performance level expected at the end of the year. This choice must be made and communicated clearly to teachers, students, and parents.

The overall performance standard is only a starting point in the standards-setting process. The most important performance standards are those used to give students feedback and scores on their demonstrations of learning. Thus when the overall performance levels and descriptors have been accepted, standard-specific and task-specific classroom rubrics based on these levels must be developed. Ideally these rubrics, written in the actual language that describes

Figure 3-4 Edmonton (AB) Catholic Elementary Schools Levels of Achievement (Performance Standards)

LEVELS OF ACHIEVEMENT
Edmonton Catholic Elementary Schools

ACCEPTABLE PERFORMANCE

LEVELS OF ACHIEVEMENT COMMENT CODE
Demonstrates Excellent Achievement
This level of achievement describes assessment evidence that demonstrates **exemplary** performance in relation to the learner outcomes from the Alberta programs of study. The evidence is characterized by an **in-depth** understanding of subject-area content, and it demonstrates **excellence** in the knowledge and skills at this grade level at the time of the report card.
Demonstrates Proficient Achievement
This level of achievement describes assessment evidence that demonstrates **skilled** performance in relation to the learner outcomes from the Alberta programs of study. The evidence is characterized by a **solid** understanding of subject-area content, and it **proficiently** demonstrates the knowledge and skills at this grade level at the time of the report card.
Demonstrates Basic Achievement
This level of achievement describes assessment evidence that demonstrates **limited** performance in relation to the learner outcomes from the Alberta programs of study. The evidence is characterized by a **generally accurate** understanding of subject-area content, and it demonstrates **basic** knowledge and skills at this grade level at the time of the report card.
Demonstrates Insufficient Achievement
This level of achievement describes assessment evidence that demonstrates **unsuccessful** performance in relation to the learner outcomes from the Alberta programs of study. The evidence is characterized by an **inadequate** understanding of subject-area content, and it demonstrates **insufficient** knowledge and skills at this grade level at the time of the report card.

Teachers use this level of achievement to screen for children not working at grade-level outcomes.

Source: Copyright © Edmonton Catholic School District, Edmonton, Alberta, Canada. Reprinted by permission.

levels of achievement for the applicable subject and grade level, will be supported by exemplars demonstrating requirements for each grade.

As with overall standard development, the best classroom performance-standard setting pools the collective experience of several educators. These teachers, confident, competent masters of the relevant academic discipline who have extensive teaching experience, set standards based on their collective study of samples of student work. Here again, teachers using these standards need frequent opportunities for professional dialogue about them so they develop shared understanding and apply them consistently. Figure 3-5 presents an example of a classroom performance standard.

As with the overall descriptors, these performance standards will be published and public for all to see from the beginning of instruction. The key to success is to describe levels of achievement in terms of the characteristics of the actual kind of academic achievement (or behaviors) being judged. Obviously, therefore, performance standards for feedback and scoring will be very different for different contexts (subjects and grade levels). When such descriptions are accompanied by samples of student work depicting each level of proficiency, we lay a solid foundation for effective judgment of and communication about student achievement.

As professional associations of teachers have established standards and associated performance continua for their particular academic disciplines, and as state assessments have been created that represent the state's standards, all involved have had to decide, How good is good enough? We should rely on these resources whenever available to assist with local standard setting.

After the performance standards are in place, understood, and used competently, in almost all schools summary grades have to be determined and communicated to students,

Figure 3-5 Oral Presentation Rubric

Score	Language	Delivery	Organization
3	Correct grammar and pronunciation are used. Word choice is interesting and appropriate. Unfamiliar terms are defined in the context of the speech.	The voice demonstrates control with few distractions. The presentation holds the listener's attention. The volume and rate are at acceptable levels. Eye contact with the audience is maintained.	The message is organized. The speaker sticks to the topic. The main points are developed. It is easy to summarize the content of the speech.
2	Correct grammar and pronunciation are used. Word choice is adequate and understandable. Unfamiliar terms are not explained in the context of the speech. There is a heavy reliance on the listener's prior knowledge.	The voice is generally under control. The speaker can be heard and understood. The speaker generally maintains eye contact with the audience.	The organization is understandable. Main points may be underdeveloped. The speaker may shift unexpectedly from one point to another, but the message remains comprehensible. The speech can be summarized.
1	Errors in grammar and pronunciation occur. Word choice lacks clarity. The speaker puts the responsibility for understanding on the listener.	The student's voice is poor. The volume may be too low and the rate too fast. There may be frequent pauses. Nonverbal behaviors tend to interfere with the message.	Ideas are listed without logical sequence. The relationships between ideas are not clear. The student strays from the stated topic. It is difficult to summarize the speech.

Note: Samples of the student work illustrating levels of quality are available. Research information on technical quality: Exact agreement rate on scores is about 70%.

Source: Adapted from "Rubric Sampler" (CD-ROM p. 52), in *Classroom Assessment for Student Learning: Doing It Right—Using It Well* by R. J. Stiggins, J. A. Arter, J. Chappuis, and S. Chappuis, 2004, Portland, OR: Assessment Training Institute. Original source unknown.

parents, and other interested parties. This requires that teachers use quality assessment (Fix 10) and combine the evidence of each student's achievement in appropriate ways (see especially Fixes 11, 12, 13, and 14). Traditionally, especially at the middle and high school levels, this has involved grading scales that have linked letter grades with percentages. This has, in effect, created both a system of 101 levels and the illusion that grades are mathematically precise. An effective standards-based system should be built on a limited number of clearly described levels based on proficiency or quality. However, partly because of our traditional use of the percentage system and partly because some aspects of learning are quantifiable, we do need to be able to show the relationships between qualitative performance standards and quantity (see Arter and Chappuis, 2006; Stiggins, 2005). Figure 3-6 shows an example of this relationship. Note that the figure offers a sort of thesaurus to clarify the meaning of each level; the numbers and symbols are there, but as reference points only. When we deemphasize the percentage system both our performance standards and the way we report student achievement will be clearer, more consistent, and richer in specific detail. As an added benefit, their clarity and specificity may make them effective teaching tools.

It is also important to recognize that performance standards are about *achievement*, not about growth or progress. For example, a student could make significant personal growth while making limited progress at a (relatively) low level of achievement; also a student could make little growth while making limited progress at a (relatively) high level of achievement. Achievement, growth, and progress are closely related but different concepts (Figure 3-7). Achievement is an absolute and is the grading variable (the basic ingredient of grades); growth and progress are both relative and can be reporting variables (aspects of student performance that

Figure 3-6 Edmonton (AB) Catholic Elementary Schools—Aligning Achievement Indicators

Wow	*Yes*	*Yes, but*	*No, but*
Excellent Achievement	**Proficient Achievement**	**Basic Achievement**	**Insufficient Achievement**
			
Exemplary Exceptional High quality In-depth Superb Outstanding	Skilled Adept Appropriate Solid Capable	Limited Predictable Within reason Generally accurate	Unsuccessful Partial Well below Inadequate Misconceptions Omissions Errors
*Some students will be within this level, **very well prepared** for the next grade level or course*	*Most students should be within this level, **well prepared** for the next grade level or course*	*Some students will be within this level, needing **more direct teacher support** to succeed at the next grade level or course*	*Students who are achieving within this level should be screened for **alternate programming***
4	3	2	1
A	**B**	**C**	**N**
80–100%	65–79%	50–64%	Below 50%

Source: Copyright © 2006 Edmonton Catholic School District, Edmonton, Alberta, Canada. Reprinted by permission.

should be communicated about but not included directly in grades). The reference point for growth is the individual—one grows *from* where they were previously; the reference point for progress is competency—one progresses *to* or *toward* competency. Although it is critical for intrinsic motivation that students have a clear sense of their own growth and progress, grades must be measures of achievement *only* so that everyone knows what they mean.

Figure 3-7 Definitions and Examples of Achievement, Growth, and Progress

Achievement

"The act of achieving or performing; an obtaining by exertion; successful performance."

Measured as an absolute, e.g., "he/she . . . is 4 feet 6 inches tall" . . . "is reading at grade 2 level"

"Achievement at . . ."

Growth

"The process of growing: increase in size, number, frequency, strength, etc."

Measured against where a child was, e.g., "he/she . . . grew three inches since last measurement" . . . "has moved from grade 1 level in the last month"

"Growth from . . ."

Progress

"Movement, as toward a goal; advance."

Relative achievement measured against a goal, standard, future result, e.g., "he/she . . . is now one inch below average height for age" . . . "is now two grade levels below expected level for age"

"Progress to . . ."

Student Involvement

This is a critical area for student involvement. The better students understand the performance standards, both overall and at the task level, the more likely it is that they will achieve at a high level. Students therefore require opportunities to develop and use task-specific scoring tools containing age-appropriate, student-friendly descriptors of the meaning of the relevant summary symbols, so that they can accurately and usefully self-assess and set goals.

Summary

Grades are broken when any of the following occurs:

- When standards, continua, and levels are not clearly described.
- When standards, continua, and levels are not shared at the beginning of learning.
- When the achievement continua are unclear or inappropriate.
- When evidence is inaccurate.
- When the cutoff scores are arbitrary.
- When the level of proficiency required is unclear to graders or learners.
- When cutoffs vary profoundly across classrooms covering the same material.

The keys to success are thus as follows:

1. Overall and specific performance standards with a limited number of levels, clearly described in the language of the appropriate achievement continuum

2. Professional dialogue about performance standards among teachers, so they develop shared understanding and apply standards consistently
3. Clear, easily understandable student- and parent-friendly versions, made available from the beginning of instruction

> Teacher Responsibilities for Performance Standards
> . . . engage in periodic moderation (group marking with other teachers using work samples, rubrics, and exemplars) to ensure collective agreement about the standards.
>
> —*Cooper, 2007, p. 74*

Don't assign grades based on a student's achievement compared to other students; compare each student's performance to preset standards.

> Grading on the curve makes learning a highly competitive activity in which students compete against one another for the few scarce rewards (high grades) distributed by the teacher. Under these conditions, students readily see that helping others become successful threatens their own chances for success. As a result, learning becomes a game of winners and losers; and because the number of rewards is kept arbitrarily small, most students are forced to be losers.
>
> —*Guskey, 1996a, pp. 18–19*

Grades are broken when they compare students to each other. The fix is to base grades on preset achievement standards—to be criterion referenced, not norm referenced in assigning grades. In doing so, we acknowledge that it is possible for all students to get an A or for all students to get an F. There would be no plan to intentionally distribute grades on a construct such as the bell curve, ensuring a few A's, more B's, even more C's, some D's and a few F's.

You can test the thinking in your school or district by answering this question: "What do you think would happen if you did an outstanding job, all the students in your class did an outstanding job, and all the students received an A?"

If the response is that the grades would be questioned with comments such as, "easy teacher," "no or low standards," or "grade inflation," then you are in a norm-referenced setting. If on the other hand the grades were questioned but the comments were "great," "that's what we want," or "let's celebrate lots of winners," you are in a criterion-referenced setting.

One of the main problems in assigning grades based on student-to-student comparisons is, What should be the reference group? Should an individual student be compared to others in their particular class at that time? All who have taken that class over time? From their particular teacher or from all teachers? What span of "all teachers"? Because there is no easy answer, interpretation of and communication with such grades is difficult at best and impossible at worst.

The rationale often cited for creating a competitive grading environment in standards-driven schools is that it provides motivation for students and that highly motivated students learn more. In fact, however, the motivational effects are not beneficial for all students. Students who finish high in the ranking and therefore have hope of getting good grades are indeed motivated. For those at the bottom, motivation wanes. They set lower standards for themselves in order to maintain their personal sense of self-worth and put forth only the effort required to meet those adjusted standards. Competitive grading systems do not reward such students, who are often "left behind" their peers both in school and beyond.

Summary

Grading students by comparing their performance to one another distorts individual achievement. We need clear, criterion-referenced achievement standards—absolute, not

relative, standards that describe a limited number of levels: at, below, and above proficiency. Teachers in a noncompetitive grading system assign grades to each student based only on that student's own achievement in relation to the applicable standards.

> There are a very few legitimate uses for norm-referenc(ing) . . . in school, and all of them are where students are competing for limited resources.
>
> —*Brookhart, 2004, p. 73*

FIX 10

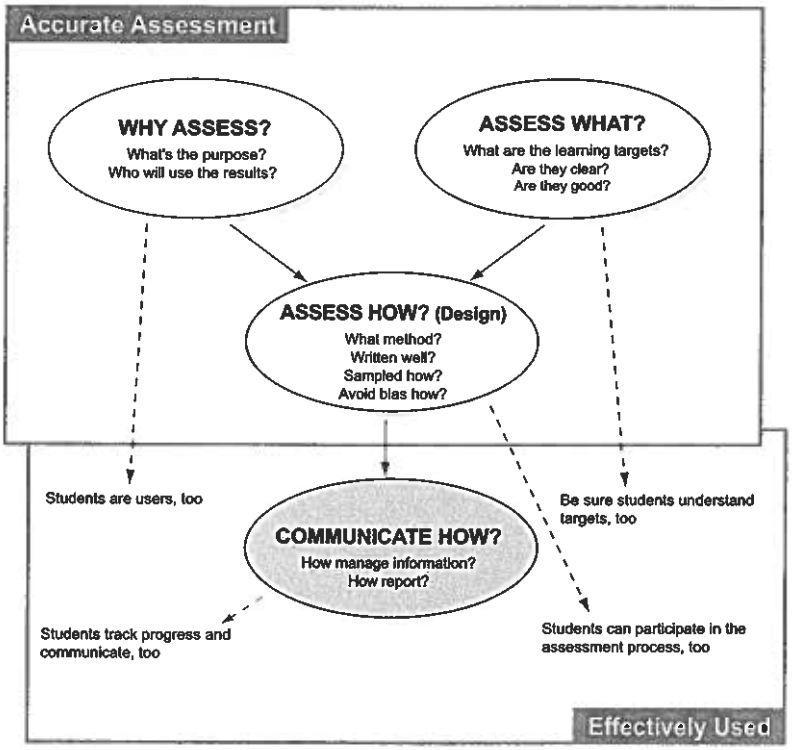
Don't rely on evidence gathered using assessments that fail to meet standards of quality; rely only on quality assessments.

> Quality classroom assessment produces accurate information that is used effectively to maximize student learning.
>
> —*Stiggins et al., 2004, p. 26*

Grades are broken when the evidence used is from poor-quality assessment and so misrepresents student achievement. This is the classic "garbage in, garbage out" syndrome. The fix is to have clear standards of assessment quality and to apply these standards to each and every assessment.

Figure 3-8 provides a framework of standards for ensuring high-quality assessment. Quality assessment requires that we have accurate assessment that is effectively used. Accurate assessment, the focus of this Fix, requires that we pay attention to three questions: Why are we assessing? What are we assessing? How will we assess it? The purpose of each assessment must be clear. In the context of this book, the assessment must be for grading purposes—an assessment of learning. The learning goals to be assessed are those established standards students are to master, specifying both what is expected and how well students must perform to earn each grade. (Students therefore must have these learning targets available and must understand them clearly.) We address the matter of how we will assess by considering the following features of design quality in creating assessments for use in grading:

Figure 3-8 The Keys to Quality Classroom Assessment



Source: From *Classroom Assessment for Student Learning: Doing It Right—Using It Well* (p. 302) by R. J. Stiggins, J. A. Arter, J. Chappuis, and S. Chappuis, 2004, Portland, OR: Assessment Training Institute. Copyright © 2006, 2004 by Educational Testing Service. Reprinted by permission of Educational Testing Service.

1. Use a proper assessment method for the context; that is, a method that will effectively and efficiently gather the needed evidence of student achievement. The proper method depends on the nature of the learning goals. For example, to assess student mastery of content knowledge, we can rely on selected response or essay formats. But to assess mastery of performance skills or the ability to create products that meet certain standards of quality, we must turn to performance assessment.

2. Build assessments out of high-quality ingredients. If the test is to rely on multiple-choice items, they must be good items, not bad ones. Performance assessments must be built of high-quality exercises and rubrics. Classroom assessors must know and understand the differences.
3. Sample student achievement appropriately; that is, gather enough evidence to make valid and reliable judgments of proficiency gained in relation to grade expectations. We know we have enough evidence when we can confidently say that, if we gathered one more item, it would simply confirm what we know now. There is a base or minimum amount of evidence needed from every student, but that amount will not be the same for each student. The more consistent a student is, the less evidence is needed; the more inconsistent, the more evidence is needed. This is just one of several factors that teachers might take into account in deciding how to sample student achievement for grading and how samples might vary across students. (For more detail see Stiggins et al., 2004.)
4. Avoid bias that can distort results. There can be problems with the student, the assessment setting, the scoring process, or the assessment itself that can cause the score to misrepresent student achievement. Problems that can occur with the student include lack of reading skill, emotional upset, poor health, lack of testwiseness, and evaluation anxiety. Problems within the setting that can distract student attention include heat, noise, and lack of light. Problems in scoring include inter-rater disagreement on criteria. Problems that can occur with the assessment include insufficient time for all students to complete the assessment.*

* "Few tasks in life—and very few tasks in scholarship—actually depend on being able to read passages or solve math problems rapidly. As a teacher, I want my students to read, write and think well; I don't care how much time they spend on their assignments. For those few jobs where speed is important, timed tests may be useful" (Gardner, 2002, n.p.).

It is well worth noting here, too, that Black and Wiliam (1998) and others have documented that high-quality, accurate classroom assessments in and of themselves measurably improve student learning. This alone is a potent reason—and perhaps the best reason—to strive to ensure that our assessments meet the highest standards of quality and accuracy.

Summary

Grades are broken when they arise from poor-quality assessment because the evidence is not accurate. The fix is to check every assessment for quality—clear purpose, clear learning goals, sound design, and avoidance of bias. Assessments that do not meet these four standards of quality will mismeasure student achievement and thus will lead to inaccurate grades.

> Evaluation experts stress that if you are going to make important decisions about students that have broad implication, such as decisions involved in grading, then you must have good evidence. . . . In the absence of good evidence, even the most detailed grading and reporting system is useless.
>
> —*Guskey and Bailey, 2001, p. 46, as quoted in Butler and McMunn, 2006, p. 188*

CHAPTER 

Fixes for Inappropriate Grade Calculation

Don't rely only on the mean; consider other measures of central tendency and use professional judgment.

> Educators must abandon the average, or arithmetic mean, as the *predominant* measurement of student achievement.
>
> —*Reeves, 2000, p. 10, emphasis added*

Grades are broken when the summary they provide of student achievement is inaccurate because the procedures used to arrive at the grade are faulty. For example, grades may mislead when they are based on simply calculating the mean (average) of a series of scores, due to the effect of outlier scores. The fix for grades broken in this way is to not use the mean as "the measure" by considering other measures of central tendency, and to recognize that grading should not be merely a numerical, mechanical exercise.

The problem with the mean is well illustrated in this quote from a letter to the editor in one of my local newspapers:

> Whenever I hear statistics being quoted I am reminded of the statistician who drowned while wading across a river with an average depth of three feet. (McMann, 2003, n.p.)

The mean can be very well named—it is truly "mean" to students because it overemphasizes outlier scores, which are most often low outliers. As we see in the following case, the calculation of the mean can distort the final grade. Ten

assessments have been converted to percentage scores to calculate a final grade:

91, 91, 91, 91, 91, 91, 91, 70, 91, 91
Total = 889, Mean = 88.9, Final grade = B

This student performed at an A level 9 times out of 10 and the 70 is clearly an anomaly. But the grade as calculated in most schools would be a B. This example raises both a general issue—How should evidence be summarized?; and one derived from it—Should evidence be summarized by strict mathematical calculation?

Somewhat ironically, in most states and provinces, somewhere about Grade 6 students are taught in math class about three methods of calculating central tendency—mean, median, and mode. They are taught that each measure of central tendency has both virtues and problems, and that, depending on the distribution of scores and your purpose, you carefully choose the appropriate measure for every situation. But somehow teachers frequently ignore this in managing their own gradebooks. If students are very consistent each measure will get the same result and the mean can suffice. But the more inconsistent a student's performance is the less effective any of the measures of central tendency is in accurately summarizing student achievement. Guskey, addressing the issue of summarizing when the record includes extreme scores, notes that "averaging falls far short of providing an accurate description of what students have learned. . . . If the purpose of grading and reporting is to provide an accurate description of what students have learned, then *averaging* must be considered *inadequate* and *inappropriate*" (Guskey, 1996a, p. 21, emphasis added).

So if the mean is "inadequate and inappropriate," what number crunching should be done? The median (middle score by rank) or the mode (the most frequently occurring score) are generally more appropriate than the mean when confronted with extreme scores. Russell Wright has written extensively about the use of the median, arguing that "grading by the median provides more opportunities for success by diminishing the impact of a few stumbles and by rewarding hard work [what I would call *consistency*]" (Wright, 1994, p. 723). In the previous example, if the median or the mode is used, the student would get their deserved A. It is necessary then for teachers to consider all measures of central tendency when determining grades.

At this point we have to ask, "Should grades in fact be determined by straight mathematical computation only?" Given the limitations of measures of central tendency to deal effectively with all score distributions and the many factors affecting student performance I conclude that we have to see grading not as simply a numerical, mechanical exercise, but as *primarily* an exercise in professional judgment. It calls for teachers to demonstrate two key aspects of professional behavior—the application of craft knowledge of sound assessment practice and the willingness and ability to make and be ready to defend one's professional judgment. As teachers we must ask the question, "Based on all the evidence of achievement a student has produced, which summary symbol most accurately represents that achievement?" That is why I always talk about "determining," not "calculating" grades—number crunching may be necessary but ultimately grading requires professional judgment. Each teacher must be prepared to specify: "These were my expectations, here is the evidence of each student's mastery of them. Using the following summary process, here is the grade I determined

for each student . . ." In borderline cases, teachers may allow or encourage students to present additional evidence to persuade the teacher to judge their achievement favorably.

An example of a teacher determining grades by combining number crunching with her professional judgment is illustrated in the following: "I thought your talk at Bronxville was very thought provoking and as I went over my grades for the year over the weekend I was thinking all the time of things you had made us consider. I definitely have a number of students for whom I will reject the average. I feel liberated!! Thanks" (anonymous personal communication, June 2002). The writer went on to say that for 100 of her 105 mostly very high-achieving students the mean was an accurate representation of their achievement, but for 5 students it wasn't and for those 5 she used her professional judgment.

Summary

Grades are frequently broken (inaccurate) when they result only from the calculation of the mean in contexts where extreme scores distort results. They can be repaired by considering other measures of central tendency and using professional judgment. Thus we should think and talk about not the calculation, but the *determination* of grades.

Not everything that can be counted counts, and not everything that counts can be counted.

—Albert Einstein, quoted in ASCD SmartBrief, *19 December 2001, n.p.*

FIX 12

Don't include zeros in grade determination when evidence is missing or as punishment; use alternatives, such as reassessing to determine real achievement, or use "I" for Incomplete or Insufficient Evidence.

> Most state standards in mathematics require that fifth-grade students understand the principles of ratios—for example, A is to B as 4 is to 3; D is to F as 1 is to zero. Yet the persistence of the zero on the 100-point scale indicates that many people with advanced degrees . . . have not applied the ratio standard to their own professional practices.
>
> —*Reeves, 2004, pp. 324–325*

Grades are broken when zeros are entered into a student's academic record for missing evidence or as punishment for transgressions. When combined with other evidence, the resulting grade does not accurately reflect student achievement. There are several fixes for the use of zeros in grades—by far the best is the use of "I" as a final grade, indicating Incomplete or Insufficient Evidence, but as transitional approaches or in situations where calculation "rules," acceptable alternatives are the use of equal difference scales or the use of a "floor" score that makes a percentage scale an equal difference scale.

Zeros most commonly are found in teachers' gradebooks when students fail to submit required assessment evidence, such as turning in assignments. They are also sometimes

used for serious behavioral infractions such as cheating. There are three fundamental problems with zeros:

- Zeros give a numerical value to something that has never been assessed and that therefore has no basis in reality.
- They can have counterproductive effects on student motivation.
- They involve inappropriate mathematics.

But the most important issue is that zeros in the record render grades ineffective as communication.

Assigning a zero to something that has not been seen compromises the accuracy of the grade and does so to an unknown extent. Such misinformation can only lead to poor-quality decisions about students and their learning.

Regarding motivation, as soon as students have more than one zero they have little chance of recovery, increasing the likelihood that they will give up. In high school for some students this can happen as early as the end of the first month of the school year, effectively rendering the remainder of the year a waste of time, at least from a learning perspective. One potential side effect is that students who have given up often have discipline problems. The other motivational problem is that students who are not concerned about grades are willing to "take a zero" and are thus not held accountable for their learning. We are faced with the irony that a policy that may be grounded in the belief of holding students accountable (giving zeros) actually allows some students to escape accountability for learning.

The mathematical problem with zeros is that they represent very extreme scores, and their effect on the grade is always exaggerated. As we have established, this is not acceptable.

The best alternative to the use of zeros is the use of an "I" for Incomplete or Insufficient Evidence. When desired evidence of student achievement is missing, teachers decide whether they have sufficient evidence to determine a grade and if they do not, assign an "I." Guskey and Bailey suggest that this "is both educationally sound and potentially quite effective" (2001, p. 144). One reason is that it clearly places the responsibility where it should be—with the student. It is the student's responsibility to produce sufficient (but not necessarily all) evidence required so the teacher can make a valid summary judgment. It is extremely important that schools/districts have this option available to teachers on each report of student achievement, including the final report card. The "I" has the same impact as an F (in high school = no credit), but it accurately communicates what the problem is. Another benefit is that while zeros can doom students to failure very early in the school year, an Incomplete can always be made complete (sufficient). Schools/districts need to have clear procedures and timelines for students to move from an "I" to a letter grade that accurately represents their achievement. This is a positive, supportive approach that is likely to be much more effective in promoting further learning than is the negative and punitive impact of zeros.

For consistency across curricula, the specific policy adopted should be developed at the school/district level. Teachers may then apply the policy in the manner best fitting their classrooms. Given that many schools/districts are still in transition to a full standards-based approach and still have grading policies that require calculation (often including the use of percentage scales), we also must consider alternatives to zeros in those situations.

Zeros are generally used in grading scales that have unequal differences in the points on the scale so that an included zero has a disproportionate effect. The most commonly used grading scale is A = 90–100 percent, B = 80–89, C = 70–79, D = 60–69, and F = below 60 percent. In this scale there are 11 points for an A, 10 for each of B, C, and D, and 60 points for an F. The problem with using this scale, and three possible solutions, are illustrated in Table 4-1. This student was supposed to do five assessments but does only four; his grades are A, B, C, and D (represented by midpoint percentage scores in Column 1) on the four assessments and a zero for the assessment he does not submit. The mean for the four assessments he did is 80 percent—a B, but the zero lowers the mean to 64 percent—a D, a drastic reduction caused by the range for an F being approximately six times greater than the range for the other grades (Column 2). Alternatives appear in Columns 3 and 4 (Equal Difference Scales) and in the bottom row (Median). Using the 5-point scale in Column 3 results in a summary grade of C; this is still lower than the mean of this student's

Table 4-1 Alternatives to Zeros

			Equal Difference Scales	
	Scores	***101-Point Scale***	***5-Point Scale***	***50-Point Scale***
	95	90–100 (A)	4	95
	85	80–89 (B)	3	85
	75	70–79 (C)	2	75
	65	60–69 (D)	1	65
	0	<60 (F)	0	50
Mean	64 (D)		2 (C)	74 (C)
Median	75 (C)		2 (C)	75 (C)

scores but is a more reasonable summary of his achievement. Column 4 turns the percentage scale into an equal difference scale by having (almost) the same number of points for each grade level, using a floor of 50 percent. (Instead of recording a zero for missing evidence the teacher would record a score of 50 percent. This symbolic percentage is chosen to equalize the points per grade. It does not mean that students have mastered 50 percent of what is expected, but that, if students actually attempt an assessment and receive a failing percentage grade it must be recorded as a percentage between 50 and 59.) The third alternative appears in the bottom row—use the median instead of the mean. Note that each alternative results in a grade of C. (For further information about alternatives to zeros see Guskey, 2005.)

Table 4-2 illustrates the need for more than simply numerical alternatives to the use of zeros. In this example, none of the measures of central tendency provide sufficient accuracy. Students were expected to submit 10 assessments. This student submitted only 3, receiving a 95, 85, and 80 percent. A zero was assigned for each of her missing assignments. If the traditional percentage scale is used, she would receive a failing grade. But the 3 assessments that were submitted clearly indicate that she had a good understanding of the material assessed. If an equal difference scale such as the 5-point scale in Table 4-1 is used, she receives a passing grade, but it is unlikely that she provided sufficient evidence if she only did 3 of 10 required assessments. Also, such an approach would support the undesirable idea that students can pick which assessments they do and choose to take a zero on other assessments. Thus there is a problem with both scales, and the median is clearly not helpful in this situation because it would be zero. This student's appropriate grade would be an "I" for Incomplete or Insufficient Evidence

Table 4-2 The Impact of Zeros

	101-Point Scale	5-Point Scale
	95	4
	0	0
	0	0
	0	0
	85	3
	0	0
	0	0
	80	3
	0	0
	0	0
Total	**260**	**10**
Mean	**26**	**1.0**
Median	**0**	**0**
Letter Grade	**F**	**D**

because it most clearly communicates the problem without distorting her actual achievement.

Another approach to controlling the use of zeros in high schools and middle schools is to use sampling to eliminate the need for them. This starts with developing and announcing assessment plans that identify the learning targets to be mastered, and that specify in advance the summative assessments that will provide the necessary evidence (both what will be assessed and when). The teachers then build overlapping assessments, each replicating part of the evidence provided by the previous one. As long as students complete a reasonable number of the assessments, a sufficient sample of the learning goals is achieved, and they will have produced enough evidence for the determination of a grade. Students

who do not complete enough assessments will receive an "I" for Insufficient Evidence, which will remain until they submit sufficient evidence.

Student Involvement

When students understand the impact of not submitting required assessment evidence, and know what alternatives are in place in their school, they are better able to decide about submitting needed academic evidence and/or making up an Incomplete. Support sessions may be available before or after school or at lunchtime in which they could participate. Students also can be involved in determining the consequences for failure to submit required assessment evidence. For example, they may agree to a contract that requires them to meet certain timelines and/or to attend specific support sessions.

Student-led conferences also may help students recognize their responsibilities, by helping them identify both their strengths and areas needing improvement. Dyck (2002) tells how one student-led conference helped a student to identify his problem with missing assignments, and also helped him recognize his successes (including an excellent *PowerPoint* presentation). Dyck notes that "at the end of the conference Greg left with two proud parents and a plan for finishing those delinquent assignments" (p. 41).

Summary

Grades are broken when zeros are used; zeros distort the actual achievement record and can decrease student motivation to learn. There are, however, many fixes in the form of grading alternatives. Schools/districts develop policies regard-

ing these alternatives, then indicate to their teachers which alternative(s) they can or should use in their classrooms.

> A zero has an undeserved and devastating influence, so much so that no matter what the student does, the grade distorts the final grade as a true indicator of mastery. Mathematically and ethically this is unacceptable.
>
> —*Wormeli, 2006, pp. 137–138*

CHAPTER 5

Fixes to Support Learning

FIX 13

Don't use information from formative assessments and practice to determine grades; use only summative evidence.

> The primary responsibility of our school is teaching and learning.
>
> The individuality of every learner is recognized and welcomed.
>
> The school culture nurtures both the joy of learning and the satisfaction of achievement.
>
> Our shared vision of education empowers us to explore, experiment and grow.
>
> Learners accept responsibility for their own learning . . .
>
> —*School District of Clayton, MO, 2006, n.p.*

Grades are broken if scores for everything students do find their way into report card grades. The fix is to include, in all but specific, limited cases, only evidence from summative assessments intended to document learning, that is, designed to serve as assessments *of* learning.

The primary purpose of grades is to communicate a summary of student achievement at a particular point in time; that is, what students know, understand, and can do as a result of their learning. It is important that teachers, students, and parents recognize that *learning* is a process in which learners increase their knowledge, understanding, and skills as a result of effort, instruction, feedback from

teachers and peers, and self-assessment and adjustment. As Jay McTighe points out, "We know that students will rarely perform at high levels on challenging learning tasks at their first attempt. Deep understanding or high levels of proficiency are achieved only as a result of trial, practice, adjustments based on feedback, and more practice" (McTighe, 1996–1997, p. 11). For this process to work well learners must believe that it is important and worthwhile to try and that it is acceptable to take risks and make mistakes; it is not necessary to always "get it" the first time.

This process is clearly acknowledged in the guiding principles of the School District of Clayton, Missouri, given in the opening quotation. However, it is not recognized when teachers include in grades evidence generated during practice (i.e., learning) activities. Unfortunately, this happens daily in many classrooms, when teachers judge everything students do and then use every piece of evidence to determine grades.

Standards-based teachers distinguish clearly between teaching activities (which include diagnostic and formative assessment) through which students learn and practice, and summative assessments in which students "perform" and show what they know, understand, and can do (Figure 5-1). They are clear about the purpose of every activity, and grades include *only* evidence from summative assessments.

It is critical that both teachers and students recognize when assessment is primarily *for* learning (formative) and when it is primarily *of* learning (summative). Students understand this in band and sports, when practice is clearly identified and separate from an actual performance or game. But often this is not clear in the classroom. If we did in basketball what we frequently do in the classroom, the game would not start 0–0, but each team would start with a score based on

Figure 5-1 Purposes of Assessment

Diagnostic—assessment that takes place prior to instruction; designed to determine a student's attitude, skills or knowledge in order to identify student needs.

Formative—assessment designed to provide direction for improvement and/or adjustment to a program for individual students or for a whole class, e.g. observation, quizzes, homework (usually), instructional questions, initial drafts/attempts. *(Assessments FOR learning)*

Summative—assessment designed to provide information to be used in making judgments about a student's achievement at the end of a sequence of instruction, e.g. final drafts/attempts, tests, exams, assignments, projects, performances. *(Assessments OF learning)*

an assessment of the quality of their practices in the days leading up to the game. Obviously this would be absurd—and it is equally so in the classroom.

A large and growing body of research supports this distinction. The Assessment Reform Group in the United Kingdom, which sponsored "Inside the Black Box," the important paper by Paul Black and Dylan Wiliam (Black and Wiliam, 1998), has sponsored and published much of this research. "Firm evidence shows that formative assessment is an essential component of classroom work and that its development can raise standards of achievement, Mr. Black and Mr. Wiliam point out. Indeed, they know of no other way of raising standards for which such a strong prima facie case can be made" (editor's introduction to Black and Wiliam, 1998, p. 139). Their research and the work of others have shown that improving formative assessment and using assessment *for*

learning raises the achievement of all students, but also that it has the most significant impact on low achievers. Learning gains made through using assessment *for* learning are similar to those achieved through one-on-one coaching.

The key components of assessment *for* learning are (1) sharing the learning target with students from the beginning of the learning, (2) making adjustments in teaching as a result of formative assessment, (3) providing descriptive feedback to students from assessment, and (4) providing opportunities for students to self- and peer assess so that they understand their strengths and what they need to do to improve. This is obviously very different from a summative use of assessment—from putting a grade or number on everything students do and including every bit of evidence when computing grades. Such summative assessment is important, but only when balanced with appropriate formative applications. Students should be assessed regularly; everything (or almost everything) they do can be assessed and/or checked, but everything does not need a score and every score need not be included in the grade. Some student work must be for practice only, and be returned to them accompanied by the kind of feedback that will help them do better the next time.

Black and Wiliam define *formative assessment* as "all those activities undertaken by teachers and by their students [that] provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (Black and Wiliam, 1998, p. 139). To appropriately modify learning, feedback has to be effective; it has to be timely, describe features of the work or performance relating directly to learning targets and/or standards of quality, and be low stakes—i.e., allow for adjustments before it "counts." This means that feedback has to be descriptive, not evaluative. A 7/10 or a 3 (from a rubric) going into a gradebook

is high stakes, provides no useful information about the learning targets, and contributes nothing that will improve learning. One of the important implications of this is that teachers need to identify clearly and record evidence derived from formative assessment separately from evidence from summative assessment. This can be done using separate pages for each in the gradebook, by color-coding entries, or by giving a zero weight to formative assessments in a computerized gradebook or spreadsheet.

One of the most common practices in North American education has been scoring and including all homework as a significant part of grades. This has been done in the belief that it promotes responsibility in students, but in fact it often has the opposite effect. Careful consideration has to be given to the purpose(s) of homework. Sometimes homework requires students to show what they know by extending or integrating their knowledge and understanding through projects or assignments done partially or completely outside the classroom. This is clearly summative assessment and is legitimately part of grades as long as there is careful monitoring to ensure that it is the student's own work. Another purpose for homework is preparation—introducing knowledge, understanding, and skills intended to help students to be ready for subsequent lessons. As this happens *before* instruction any assessment would be diagnostic, which obviously has no place in grades. Most often, however, homework is practice of whatever was learned in class that day—any assessment of this work should be regarded as formative. Practice is valuable only to those students who can have some degree of success on their own without teacher support. It is of little or no value to students that don't need to practice, and it can actually be damaging to students who don't understand because they may embed misunderstandings that will be difficult to correct.

Putting a mark on work done for practice renders it effectively summative, not formative. When homework assigned as practice is scored and included in grades, what becomes most important to students is that it be done because it "counts," not because of any learning that might occur. It becomes an issue of compliance so it really doesn't matter who does the homework—the student, a parent, a sibling, or a friend. If we want homework to be about learning, we need students to understand that it is for practice if they need it, not compliance or grading, because then the person who benefits from the homework is the learner.

One major concern that is often expressed about not including practice homework in grades is, "Students won't do their homework if I don't grade it!" We have done an absolutely superb job of training students into this perspective by putting a number on everything they do and making every number part of the grade. But as we have trained them into it, we can train them out of it. The motivation for practicing and doing homework should come from each student's clear understanding that it will contribute to their learning. We want them to feel a sense of satisfaction from knowing, understanding, or being able to do something better today than yesterday. We want them to think, "If I had done my homework, I would have done better on the test." We can tell them this day after day with no effect, but when they see the assessment evidence speak for itself and understand that practice really does help, they will come to this realization themselves.

As stated, including practice homework in grades can be damaging to struggling students because they may develop misunderstandings that will be difficult to correct. It is also damaging to these students because it reduces their willingness to try. If they know that they are going to get a low

score, then to avoid yet another failure one defense mechanism is not to do it. It is better to keep the stakes low and have students understand, "It is okay to try because if I try I am going to get feedback on what I did well and what needs improvement."

Including practice homework in grades can also be damaging in other ways. Consider this quote from Elinor Burkett, after spending a year observing in a suburban Minneapolis high school: "Nick was fed up; . . . fed up with acing exams but getting C's at the end of the trimester because he refused to do the worksheets assigned in order to help students study so they could ace exams" (Burkett, 2002, p. 124). Nick did not need to do the practice work. Students such as Nick, who refuse to go along and do what for them is busy work, end up with lowered grades that do not reflect their achievement.

Finally, including practice work and/or learning activities in grades can harm students who consistently improve. Consider the high school mathematics class with three formative quizzes and one summative test over a three-week period. Jeremy receives scores of 30, 50, and 70 percent on the quizzes and 90 percent on the test. He has obviously mastered whatever was taught over that unit. But if the quizzes count for one-third of the grade and the test for two-thirds (this is common) he would receive a grade of 77 percent, which in most high schools would be a C or worse.

The fix for all these broken grades is not to include scores from learning activities, including practice homework, in grades. One of the best ways to ensure this happens and to make the process clear to students is to develop assessment plans and (age appropriately) make them known to students.

An assessment plan should start with the desired results—the learning goals derived from the standards. The

summative assessments that are going to be used to determine whether the student "knows and can do," (i.e., the only assessments that will be used to determine grades) follow. Next are the diagnostic assessment(s) that are going to help determine the what and the how for teaching and learning. Finally come the formative assessments that are going to help students achieve the learning goals and through which the teacher will adjust teaching and learning activities. These activities include the homework and quizzes that help students to be successful on tests, the practices that lead to performances, and the series of drafts that help students to produce high-quality products.

Figure 5-2 shows an example of the formative and summative assessment part of such a plan.

Figure 5-2 Sample Assessment Plan

Formative Assessment for Unit 1

Task	Method(S)	Strategy(ies)	Scoring Tool	Assessor
ROLE PLAY Practice(s)	Performance Ass't	Performance	Rubric	self/peer
QUIZZES	Paper and Pencil	Selected Response	Marking Scheme	Teacher
BROCHURE Draft	Performance Ass't	Product	Rubric	peer
BROCHURE Near Final	Performance Ass't	Product	Rubric	self/peer

Summative Assessment for Unit 1

Task	Method(S)	Strategy(ies)	Scoring Tool	Assessor
ROLE PLAY	Performance Ass't	Performance	Rubric	Teacher
TEST(S)	Paper and Pencil	Selected & Constructed Response	Marking Scheme	Teacher
BROCHURE	Performance Ass't	Product	Rubric	Teacher

Note that in this plan there is a clear link between the formative and summative assessments—one or more practices of the role play with descriptive feedback will help students to perform high-quality role plays, one or more quizzes followed by analysis of strengths and weaknesses and appropriate reteaching will help students to be successful on the test(s), and the draft and near final versions of the product with descriptive feedback will lead to high-quality brochures. When a plan such as this is in place and students—and parents—are familiar with it, it is obvious to all that the focus is on the learning, not simply on the accumulation of points.

Quality school and district policy documents distinguish between formative and summative assessment and state clearly the uses of each, as in this from Manitoba:

> The thrust of formative assessment is toward improving learning and instruction. Therefore, the information should not be used for assigning [grades] as the assessment often occurs before students have had full opportunities to learn content or develop skills. (Manitoba Education and Training, 1997, p. 9)

As a final idea in this Fix, I would like to note that this statement in the Manitoba policy and that made in the first paragraph of this Fix ("include only evidence from summative assessments intended to document learning") state the principle very strongly and clearly. However, once teachers have become clear about the appropriate uses for formative and summative assessment, and abandoned the practice of including everything in grades, especially homework, it is acceptable to *consider* formative assessment evidence when determining grades. This, of course, also requires that teachers are determining, not simply calculating, grades (see Fix 11). I acknowledge that I overstate when I say summative "*only*," but given traditional grading practices it seems

to me that we have to establish this "strong" position; when teachers have developed a deeper understanding of grading issues, they can take a more holistic view of the evidence of achievement that each student has produced.

Student Involvement

This is the most critical area for student involvement because students have often been "trained" in classrooms where no distinction was made between practice and performance and where there was little feedback or opportunities to make adjustments in learning (or teaching) based on formative assessment. Students who are actively involved in every aspect of assessment are more able to themselves distinguish between practice and performance. This can be achieved by encouraging self-monitoring and self-adjustment through assessment for learning and by avoiding rushing to judgment (summative assessment) for as long as possible. Stiggins and Chappuis (2005) describe strategies that teachers can use to involve students, including the following:

> 1. Engage students in reviewing strong and weak samples in order to determine attributes of a good performance or product. . . .
> 3. Students practice using criteria to evaluate anonymous strong and weak work.
> 4. Students work in pairs to revise an anonymous weak work sample they have just evaluated. (2005, p. 15)

Teachers can also help students to be reflective learners by providing them with opportunities to think about their performance on summative assessments. Stiggins and Chappuis suggest one way to do this:

9. Teacher arranges items on a test according to specific learning targets, and prepares a "test analysis" chart for student, with three boxes: "My strengths," "Quick review," and "Further study." After handing back the corrected test, students identify learning targets they have mastered and write them in the "My strengths" box. Next, students categorize their wrong answers as either "simple mistake" or "further study." Then, students list the simple mistakes in the "Quick review" box. Last, students write the rest of the learning targets represented by wrong answers in the "Further study" box. (2005, p. 15)

Summary

Grades are broken when they are merely about accumulating points. To make it obvious that they are about learning, the fix is to distinguish between formative and summative assessment and to include only results from the latter directly in grades.

> The test of a successful education is not the amount of knowledge that a pupil takes away from school, but his appetite to know and his capacity to learn. If the school sends out children with the desire for knowledge and some idea about how to acquire it, it will have done its work. Too many leave school with the appetite killed and the mind loaded with undigested lumps of information.
>
> *—Sir Richard Livingstone, President of Corpus Christi College, Oxford, 1941, quoted in Wiliam, 2003, n.p.*

FIX 14

Don't summarize evidence accumulated over time when learning is developmental and will grow with time and repeated opportunities; in those instances, emphasize more recent achievement.

> The key question is, "What information provides the most accurate depiction of students' learning at this time?" In nearly all cases, the answer is "the most current information." If students demonstrate that past assessment information no longer accurately reflects their learning, that information must be dropped and replaced by the new information. Continuing to rely on past assessment data miscommunicates students' learning.
>
> —*Guskey, 1996a, p. 21*

Grades are broken when learning is developmental (likely to improve over time with practice and repeated opportunities) and the final grade does not recognize the student's final level of proficiency. The fix for this type of broken grade is that for any developmental learning we must emphasize the more recent evidence and allow new evidence to replace, not simply be added to, old evidence.

Guskey says, "Educators generally recognize learning as a progressive and incremental process. Most also agree that students should have multiple opportunities to demonstrate their learning. But is it fair to consider all these learning trials in determining students' grades? If at any time in

the instructional process students demonstrate that they have learned the concepts well and mastered the intended learning goals, doesn't that make all previous information on their learning of those concepts inaccurate and invalid? Why then should such information be 'averaged in' when determining students' grades?" (Guskey, 2002, pp. 777–778).

Two very important issues emerge within this quotation as we think about grading developmental learning. First, by emphasizing the more recent information we acknowledge learning as a process and we can give students the message, "It is never over until it is really over!" One of the most unfortunate effects of simply adding up all the scores and calculating the mean is that many students will never be able to overcome the impact of early failures/very low scores. This is a terrible situation for both students and teacher because students who have no hope of success give up—and, as noted previously, frequently become discipline problems. If, however, our message to students is that we will acknowledge their learning whenever it occurs, then they have no reason to give up. In fact, this approach is a powerful motivator for students achieving at any level, because every student will know that improved achievement will get full recognition. Reeves (2000, p. 11) points out that the effective schools research shows that "one of the most consistent practices of successful teachers is the provision of multiple opportunities to learn . . . . The consequence for a student who fails to meet a standard is not a low grade but rather the opportunity, indeed the requirement—to resubmit his or her work."

The second issue Guskey's quotation raises concerns averaging. He indicates that with developmental learning, more recent information should not be simply averaged with outdated evidence when determining grades. Only the

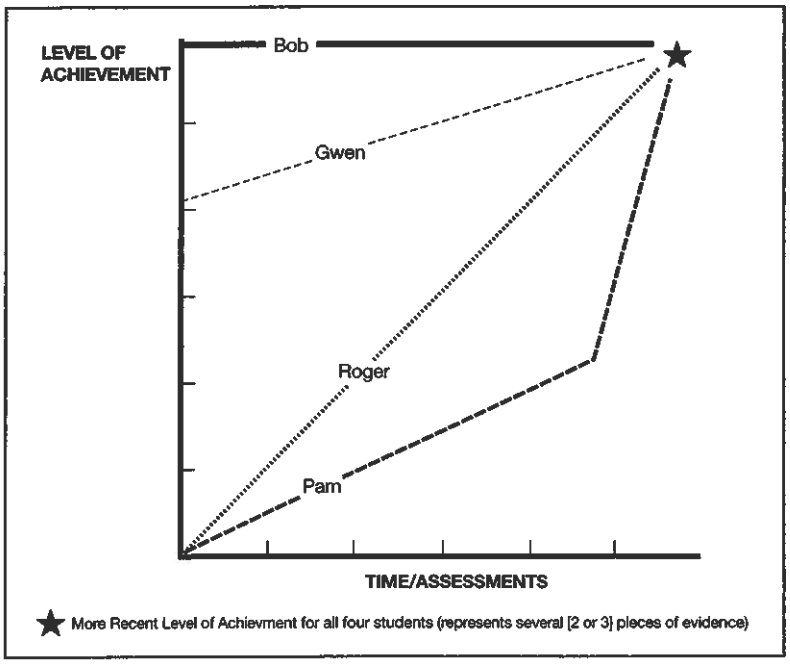
more recent data should be used. Fix 11 is relevant to this context. In it I propose that we see grading as an exercise in professional judgment, not just as a numerical, mechanical activity. The practical implication of this is that a common practice in high schools and middle schools must be abandoned—that of determining final grades by adding the grade for each grading period and dividing by the number of grading periods. Attempts have been made to devise a "power formula" or weighting to be applied when calculating grades that would emphasize more recent achievement while still allowing teachers to calculate grades, but none of these is as good as teacher judgment.

By emphasizing more recent evidence we acknowledge the impact of good teaching on student success. Consider the learning achievement curves of the four students in the graph shown in Figure 5-3.

Bob is basically the student who doesn't need a teacher, Gwen is a fairly typical student, while Roger and Pam need a lot of help. If all four of these students get the same grade, it will acknowledge their equally high achievement at the end. If Roger and Pam get lower grades, as they might traditionally, their achievement is misrepresented, which is academically unjustifiable.

If teachers really help students to meet standards and students then do so, this achievement needs to be recognized when grades are determined. For example, as part of the "effort-based intelligence" model within their "Framework for Improving Teaching and Learning," the Montgomery County (Maryland) Public Schools applies the criterion, "Staff shows tenacity to get students to meet standards"; the evidence that this tenacity bears fruit is then indicated by "a variety of student work that matches desired outcomes" (Montgomery County Public Schools, n.d., n.p.).

Figure 5-3 Levels of Achievement Over Time



For such a model to be truly effective, teachers will then assign students grades that accurately reflect their final achievement levels.

Summary

When learning is developmental and results from a process that unfolds over time so that student achievement increases with practice, the more recent evidence should "count" for the student's grade; old, outmoded evidence should be discarded. Grades are broken when this is not done. The fix is

to emphasize more recent achievement, with more recent evidence replacing previous evidence.

> What matters is not what you have at the starting point, but whether and how well you finish.
>
> —*Gardner, 2002, n.p.*

FIX 15

Don't leave students out of the grading process. Involve students; they can—and should—play key roles in assessment and grading that promote achievement.

> We must constantly remind ourselves that the ultimate purpose of evaluation is to enable students to evaluate themselves. Educators may have been practicing this skill to the exclusion of learners; we need to shift part of that responsibility to students. Fostering students' ability to direct and redirect themselves must be a major goal . . . or what is education for?
>
> —*Costa, 1991, p. 313*

Grades—and assessment—are broken if teachers simply "run the show." Students must be involved in all stages of the assessment process and should understand (age appropriately) from the outset how grades will be determined. Students can learn how to monitor their own progress, and how to communicate that progress to others. In so doing, they understand more deeply their own strengths and areas needing improvement, and can use that understanding to guide specific, meaningful goal-setting about what they can learn/do next. Ideas about student involvement have appeared in several of this book's Fixes. Fix 15 is a summary and restatement of the key ideas.

Grades should communicate achievement status, and both assessment and grading need to help students achieve at higher levels and develop positive attitudes about learning. These things are more likely to happen when students are

involved as active participants in ongoing assessment and grading, so that they see the entire process as something that is done with them, not to them. Teachers also benefit when they share with students from the beginning how they will determine grades. It is also important that students (and parents) receive short, clear written statements about grading policy/procedures. Figure 5-4 shows an example of such a policy.

We must be mindful of the fact that students are users of the information that comes from assessments, so the purpose of each assessment must be clear to them. We must also be sure that students understand the targets; there are many strategies that can be used to help them with this,* but one of the most powerful is to involve them in developing the rubrics we use to provide feedback and/or scores. Probably the most important aspect of student involvement is having them track their progress and achievement and then communicate about their learning with other students, teachers, and significant adults in their life. One way to do this is to use assessment plans, such as the one shown in Figure 5-2. It is the teacher's responsibility to evaluate the summative assessments, but students should be involved in peer and self-assessment of formative assessments. This allows them to practice the skills of self-assessment and to deepen their understanding of the conditions of quality. When students have become self-assessors who are reflective learners they then communicate with parents or significant adults about their strengths, areas for improvement, and next steps in their learning. This means that schools/districts need to set up their communication system to include student-involved or student-led conferences from kindergarten through high school. This type of conference has been found to have a significant impact on students taking responsibility for their own learning and to result in better parent attendance.

* See, for example, the "Seven Strategies of Assessment *for* Learning" in Stiggins et al., 2004, pp. 42–46.

Figure 5-4 Mrs. Greier's Grading Practices

Grading: Grades will be based on mastery of the Sunshine State Standards.

Formative Assessment

This type of assessment is for practice only. It will not be averaged for the report card grade.

EXAMPLES INCLUDE BUT ARE NOT LIMITED TO THE FOLLOWING:

- Daily Work—Center work, group assignments, math practice, etc.
- Homework—Every Monday your child will write their homework assignments for that week in their homework folder. Please sign the homework sheet that is in his/her folder and have your child bring it back to school on Tuesday. Your child will have math homework every night and is due the next morning. On Thursday morning, all other homework assignments will be collected. These assignments must be completed in order to receive privileges such as Fun Friday and other events.

You will be contacted via phone or letter if your child is not completing assignments.

Today's Homework Makes
Tomorrow's Home Work

Summative Assessment

This type of assessment is those that "sum" or measure what your child has learned. A grade will be assigned to the work and goes in the grade book following several opportunities for the student to practice the skill.

EXAMPLES INCLUDE BUT ARE NOT LIMITED TO THE FOLLOWING:

- Pencil and paper test
- Performance task—The student will be asked to perform a skill such as properly measuring liquids.
- Presentation—Student presents material he or she has learned in the form of book talks, reports, etc.
- Rubric—Rubrics are used on many summative assessments. The student is assessed on a number scale according to their achievement.

Grading Scale

90–100 A 80–89 B 70–79 C 60–69 D 0–59 F

Figure 5-4 Mrs. Greier's Grading Practices (Continued)

Expectations

Be respectful and responsible
Encourage others
Always do your best
Care and cooperate
Have the courage to try

I have read and understand this grading policy.

Student_____ Parent_____

Source: Copyright © 2005 by Glenda Greier. Reprinted by permission.

Summary

Grades are broken when students do not understand how their grades have been determined, and when they have been excluded from assessment, record keeping, and communication. The fix is to ensure that students understand how grades have been determined and to involve them as much as possible in all phases of learning and assessment.

> As students become more involved in the assessment process, teachers find themselves working differently. . . . Many teachers are spending less time marking at the end of learning and more time helping students during the learning.
>
> —*Davies, 2000, p. 9*

CHAPTER 6

Summary and Reflection

> The best thing you can do is make sure your grades convey meaningful, accurate information about student achievement. If grades give sound information to students, then their perceptions (and) conclusions about themselves as learners, and decisions about future activity will be the best they can be.
>
> —*Brookhart, 2004, p. 34*

Grades are summary symbols that should communicate only about student achievement at a point in time. To be effective, they must be consistent, accurate, and meaningful, and should support learning. Unfortunately, because grading has often been an unexamined and private practice, grades have frequently not met these standards and as a result are very often broken. In this book I have described 15 Fixes for broken grades—fixes for ingredients that distort achievement, low-quality or poorly organized evidence, inappropriate grade calculation, and lack of support for learning.

Linking the Fixes to these standards, for consistency Fix 8 needs to be in place. For accuracy, Fixes 1 to 6 and 9 to 12 need to be used. For grades to be meaningful, Fix 7 needs to be applied. And to support learning, Fixes 13, 14, and 15 need to be implemented.

This is a long list, and implementing the Fixes is not easy. Achievement in standards-based systems equals mastering those standards. Required content and performance standards must be clear and must be the focus of curriculum, instruction, assessment, grading, and reporting. Assessment must be accurate and the *process* of learning emphasized by involving students in assessment, record keeping, and communication.

Practically, any one (or more) of the Fixes can be used as a starting point. But effective grading in standards-based systems really flows from Fix 7 because as soon as one is truly standards based in assessment and grading the other Fixes become logical extensions. When grading is only about achievement of standards it quickly becomes obvious that it is inappropriate to include factors other than achievement (1–6), that it is necessary to have quality evidence (8–10) that accurately summarizes student achievement (11, 12), and that the emphasis needs to be on the learning process itself (13–15).

Teachers, schools, and districts need to examine their grading procedures and policies to see if they "fit" with what is expected in standards-based systems. Changes in grading practices will occur when teaches engage in professional dialogue about grading and agree on guidelines that avoid the inappropriate use and interpretation of grades. Fixes such as those described here can be the basis for such guidelines and those guidelines can then be incorporated into school board and school grading policies.

It is my intent and hope in writing this book that teachers will become reflective practitioners in the area of grading and that the 15 Fixes will form the basis for grades that are consistent, accurate, and meaningful, and that support learning.

> Neither computerized calculations nor rigorously applied grading systems are enough to save schools from some of the most common and egregious errors in grading. Amazingly, teachers regularly use and leaders tolerate grading systems that may appear to be accurate but are devoid of the most basic elements of mathematical reasoning and are neither fair nor effective.
>
> —*Reeves, 2006, p. 119*

References

Abbott, J. 1994. *Learning that makes sense: Recreating education for a changing future*. Letchworth, UK: Education 2000.

Archbishop Macdonald High School. 2006. *2006–2007 Student handbook.* Edmonton, AB: Author.

Arter, J. A., & J. Chappuis. 2006. *Creating and recognizing quality rubrics*. Portland, OR: Educational Testing Service.

Bailey, J., & J. McTighe. 1996. Reporting achievement at the secondary school level: What and how? In T. R. Guskey (Ed.), *Communicating student learning: The ASCD yearbook 1996* (pp. 119–140), Alexandria, VA: ASCD.

Benevino, M., & D. Snodgrass. 1998. *Collaborative learning in middle and secondary schools: Applications and assessments*. Larchmont, NY: Eye on Education.

Black, P., & D. Wiliam. 1998. Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2): 139–148.

Brookhart, S. 2004. *Grading*. Upper Saddle River, NJ: Merrill/Prentice Hall.

Burkett, E. 2002. *Another planet: A year in the life of a suburban high school*. New York: Perennial.

Butler S.M., & N. D. McMunn. 2006. *A teacher's guide to classroom assessment: Understanding and using classroom assessment to improve student learning.* San Francisco: Jossey-Bass.

Carr, J. 2000. Technical issues of grading methods. In E. Trumbull & B. Farr (Eds.), *Grading and reporting in an age of standards* (pp. 45–70). Norwood, MA: Christopher Gordon.

Chappuis, S., & R. J. Stiggins. 2002. Classroom assessment for learning. *Educational Leadership, 60*(1): 40–43.

Cizek, G. J. 2003. *Detecting and preventing classroom cheating: Promoting integrity in assessment.* Thousand Oaks, CA: Corwin.

Clements, A. 2004. *The report card.* New York: Simon & Schuster.

Cooper, D. 2007. *Talk about assessment.* Toronto, ON: Thomson Nelson.

Costa, A. 1991. *Developing minds: A resource book for teaching thinking.* Alexandria, VA: ASCD.

Covington, M., & K. Manheim Teel. 1996. *Overcoming student failure: Changing motives and incentives for learning.* Washington, DC: American Psychological Association.

Davies, A. 2000. *Making classroom assessment work.* Merville, BC: Classroom Connections.

DuFour, R., R. Eaker, R. DuFour, & G. Karhanek. 2004. *Whatever it takes: How professional learning communities respond when kids don't learn.* Bloomington, IN: Solution Tree.

Dyck, B. A. 2002. Student-led conferences up close and personal. *Middle Ground, 6*(2): 39–41.

Feller, B. 2004, 29 April. Students' rising grades raise questions. Retrieved 14 June 2004 from http//www.boston.com/

Gardner, H. 2002, 18 July. Testing for aptitude, not for speed. *New York Times*, n.p.

Gathercoal, F. 1997. *Judicious discipline*. San Francisco: Caddo Gap.

Gibbs, J. 2000. *A new way of learning and being together.* Windsor, ON: CenterSource Systems.

Ginsberg, M. 2004. *Motivation matters: A workbook for school change*. San Francisco: Jossey-Bass.

Guskey, T. R. (Ed.). 1996a. *Communicating student learning: The ASCD yearbook 1996*. Alexandria, VA: ASCD.

Guskey, T. R. 1996b. Reporting on student learning: Lessons from the past—prescriptions for the future. In T. R. Guskey (Ed.), *Communicating student learning: The ASCD yearbook 1996* (pp. 13–24). Alexandria, VA: ASCD.

Guskey, T. R. 2002. Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan, 83*(10): 777–778.

Guskey, T. R. 2004. The communication challenge of standards-based reporting. *Phi Delta Kappan, 86*(4): 326–329.

Guskey, T. R. 2005. Zero alternatives. *Principal Leadership, 5*(2): 49–53.

Guskey, T. R., & J. Bailey. 2001. *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin.

Helderman, R. S. 2004, 21 June. Thrown for a loss by SOL's. *Washington Post*, p. B01.

Johnson, D. W., & R. T. Johnson. 2004. *Assessing students in groups; Promoting group responsibility and individual accountability*. Thousand Oaks, CA: Corwin.

Kagan, S. 1995. Group grades miss the mark. *Educational Leadership, 52*(8): 68–71.

Kendall, J., & R. J. Marzano. 1997. *Content knowledge: A compendium of standards and benchmarks for K–12 education*. Aurora, CO: McREL.

Kohn, A. 1993. *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise and other bribes*. New York: Houghton Mifflin.

Manitoba Education and Training. 1997. *Reporting on student progress and achievement: A policy handbook for teachers, administrators and parents*. Winnipeg, MN: Author.

Manitoba Education, Citizenship and Youth. 2006. *Policies and procedures for standards tests, 2006–2007*. Winnipeg, MB: Author.

Marshall, M. 2001a. *Discipline without stress*. Los Alamitos, CA: Piper.

Marshall, M. 2001b. *Promoting Responsibility Newsletter*, 1(4).

Marzano, R. J. 2000. *Transforming classroom grading*. Alexandria, VA: ASCD.

Marzano, R. J. 2006. *Classroom assessment and grading that work*. Alexandria, VA: ASCD.

Marzano, R. J., & J. Kendall. 1996. *A comprehensive guide to developing standards-based districts, schools, and classrooms*. Aurora, CO: McREL.

McMann, G. 2003, 15 October. Letter to the editor. *Toronto Globe and Mail*, n.p.

McTighe, J. 1996–1997. What happens between assessments. *Educational Leadership*, *54*(4): 6–12.

McTighe, J., & K. O'Connor. 2005. Seven practices for effective learning. *Educational Leadership, 63*(3): 10–17.

Montgomery County Public Schools. n.d. *Framework for improving teaching and learning*. Rockville, MD: Author. Available at http://www.mcps.k12.md.us/departments/dsd/leadership/handouts/lookfors.pdf

O'Connor, K. 1995. Guidelines for grading that support learning and student success. *NASSP Bulletin, 79*(5): 91–95.

O'Connor, K. 1999. *The mindful school: How to grade for learning*. Arlington Heights, IL: Skylight.

O'Connor, K. 2001. The principal's role in report card grading. *NASSP Bulletin, 85*(625): 37-46.

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

Patterson, W. 2003. Breaking out of our boxes. *Phi Delta Kappan, 84*(8): 569–574.

Reeves, D. B. 2000. Standards are not enough: Essential transformations for school success. *NASSP Bulletin, 84*(620): 5–19.

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

Reeves, D. B. 2006. *The learning leader: How to focus school improvement for better results*. Alexandria, VA: ASCD.

Rogers, S., J. Ludington, & S. Graham. 1998. *Motivation and learning*. Evergreen, CO: Peak Learning Systems.

School District of Clayton, MO. 2003. *2003 Annual report.* Clayton, MO: Author.

Standards for foreign language learning. n.d. Yonkers, NY: American Council on the Teaching of Foreign Languages. Available at http://www.yearoflanguages.org/files/public/execsumm.pdf

Starsinic, J. 2003, 21 November. Letter to the editor. *Harrisburg* (PA) *Patriot News*, n.p.

Stiggins, R. J. 2001. *Student-involved classroom assessment*, 3rd ed. Upper Saddle River, NJ: Merrill/Prentice Hall.

Stiggins, R. J. 2005. *Student-involved assessment* for *learning*, 4th ed. Upper Saddle River, NJ: Merrill/Prentice Hall.

Stiggins. R. J., J. A. Arter, J. Chappuis, & S. Chappuis. 2004. *Classroom assessment* for *student learning: Doing it right—using it well.* Portland, OR: Assessment Training Institute.

Stiggins, R. J., & J. Chappuis. 2005. Using student-involved classroom assessment to close achievement gaps. *Theory into Practice, 44*(1): 11–18.

Szatanski, B., & C. Taafe. 1999. *Classrooms of choice: A teacher's guide to creating the dynamic classroom.* Ottawa, ON: Cebra.

Tombari, M., & G. Borich. 1999. *Authentic assessment in the classroom.* Upper Saddle River, NJ: Merrill/Prentice Hall.

Tomlinson, C. A., & J. McTighe. 2006. *Integrating differentiated instruction and understanding by design.* Alexandria, VA: ASCD.

Toronto (ON) National Post. 2001, 18 April. Title unavailable. p. F3.

Trumbull, E., & B. Farr (Eds.). 2000. *Grading and reporting in an age of standards.* Norwood, MA: Christopher Gordon.

Wiggins, G. 2000. Posted comment. Retrieved 19 January 2000 from http//www. chatserver.ascd.org

Wormeli, R. 2006. *Fair isn't always equal: Assessing and grading in the differentiated classroom.* Portland, ME: Stenhouse.

Wright, R. 1994. Success for all: The median is the key. *Phi Delta Kappan, 75*(9): 723–725.