Introduction

"If I were asked to enumerate ten educational stupidities, the giving of grades would be at the top of the list."—DeZouche (1945, p. 339)

"Schools use grades because it's one of those things somebody once decided on and now everybody just goes along with it."—Littky (2004, p. 154)

"Grading and reporting student achievement is a caring, sensitive process that requires teachers' professional judgment."—Cooper (2007, p. 240)

"Teachers are often very sensitive about their grading practices—they regard them as sacred ground." —Brown (2004, p. 29)

"Of all the things we do as teachers, few have the potential for creating more problems and miscommunication than grading."—Stiggins et al. (2004, p. 304)

"In a perfect world there would be no grades—at least as we know them now."—Brookhart (2004, p. 2)

"Official guidelines on grading are often vague, nonexistent or ignored."—Mathews (2005c, p. A10)

"Schools have not changed [from] toxic and counterproductive grading policies."—Reeves (2007, p. 231)

"Students . . . see their schoolwork as a game they play for grades." —Winger (2005, p. 62)

"Grades are momentary inferences at best."—Wormeli (2006, p. 11)

"Informative assessment really isn't about the grade book." —Tomlinson (2007/2008, p. 10)

"A profound cultural transformation [would be] classrooms in which both students and teachers focus on learning, not grades."—Shepard (2005, p. 70)

WHAT GRADING TERMINOLOGY IS NEEDED?

Reflecting on . . . Quotes

Ask yourself the following questions:

- What is your reaction to the titles and quotes about grades on page 1?
- What do they say about grading?
- How do you think your colleagues, students, parents, and community would react to them?

As the quotes about grading on the previous page show, grading raises many concerns. One communication concern is grading terminology. The term *grading* carries different meanings for different people, while other words, such as *marking* and *scoring,* may sometimes mean grading, too. As McTighe and Ferrara (1995) state, "Terms [are] frequently used interchangeably, although they should have distinct meanings" (p. 11). Discussion of any issue or principle must proceed from a clear understanding of the meaning of the terms being used. In support of this goal, a glossary is provided at the end of the book. At this point, readers need a shared understanding of three critical terms: *grades* (or grading), *marks* (or marking), and *scores* (or scoring). These terms are often used almost interchangeably, although *grading is* used more frequently in the United States and *marking* more commonly in Canada, while *scoring* is most commonly associated with large-scale external assessment.

Reflecting on . . . Terminology

Ask yourself the following questions:

- What do you understand by the terms *grades* and *grading*?
- What do you understand by the terms *marks* and *marking*?
- What do you understand by the terms s*cores* and *scoring*?
- Are they the same? Are they different? How?

The main problem is that these terms, especially *grades* and *grading,* are often used with two meanings. For careful analysis, it is critical to have a clear meaning for each term. In this book, *grade(s)/grading, mark(s)/marking,* and *score(s)/scoring* are used as follows:

Grade(s) or grading—The number or letter reported at the end of a period of time as a summary statement of student performance

Mark(s) or marking and score(s) or scoring—The number, letter, or words placed on any single student assessment (test, performance task, etc.)

Airasian (1994) uses *grading* to mean "making a judgment about the quality of a pupil's performance, whether it is performance on a single assessment or performance across many assessments" (p. 281). In most writings, the context

makes clear which of these two meanings is intended. However, this is not always the case, and when the meaning is not clear, confusion and lack of clarity in analysis and discussion requires that the two activities be distinguished by using separate terms.

Anderson and Wendel (1988) define *marks* and *grades* oppositely to the definitions used here. They agree, though, that defining terms is essential so that "everyone operates under the same assumptions and knows exactly what meanings underlie those assumptions" (pp. 36–37).

Another definition is provided by Paul Dressel (as cited in Kohn, 1993b):

> A grade can be regarded only as an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite amount of material. (p. 201)

WHAT IS THE CONTEXT OF GRADING?

> The role of schools has changed. . . . Our assessment practices historically have been designed to promote accountability by separating the successful from the unsuccessful learners and highlighting their differences. However, given the new mission of ensuring universal competence, assessments now must support the learning of all students so that all can succeed at meeting standards. The result must be balanced assessment systems and a fundamental rethinking of the dynamics of assessment in effective schools. (Stiggins, 2006, p. 3)

This quote from Rick Stiggins eloquently summarizes the shift in thinking about assessment that has occurred since the 1980s and shows that a different understanding has developed about the purpose of assessment. This change has led to the understanding that it is essential to make clear distinctions between assessment *for* learning and assessment *of* learning (Stiggins, Arter, Chappuis, & Chappuis, 2004) and that schools and districts must ensure a balance between both purposes in the assessments used by teachers. This distinction has been taken a step further by Lorna Earl and Steven Katz (2006) in the document they wrote for the Manitoba Ministry of Education, Citizenship and Youth titled *Rethinking Classroom Assessment,* where they make distinctions between assessment for, of, and as learning. In this view of the purposes of assessment, assessment *for* learning has been split into two:

1. Assessment *for* learning—Basically done by others who provide students with descriptive feedback to move their learning forward

2. Assessment *as* learning—basically done by the students themselves through reflection, self-assessment, and goal setting

CONSTRUCTIVIST THEORIES OF LEARNING

Another important understanding has been the development of constructivist theories of learning. Constructivism recognizes that learning is a process in

which the learner builds personal meaning by adding new understanding to old on the basis of each new experience. This means that "learning is not linear. . . . Instead, learning occurs at a very uneven pace and proceeds in many different directions at once" (Burke, 1993, p. xiv).

Individuals experience meaningful learning when they have the opportunity to process information and relate it to their own experiences. This has significant implications for how the teaching/learning process takes place in schools.

> Learners should be able to construct meaning for themselves, reflect on the significance of the meaning, and self-assess to determine their own strengths and weaknesses. Integrated curricula, cooperative learning, problem-based learning, and whole language are just a few examples of curricula that help students construct knowledge for themselves. (Burke, 1993, p. xiv)

Each of these approaches requires more complex assessment than traditional approaches, which emphasize simple scoring of answers or behaviors as right or wrong. More varied approaches to assessment imply that teachers will not always have neat numbers that can be "crunched" and converted into grades. Grading, therefore, also becomes a more complex activity. Teachers need to consider carefully how they will incorporate data from a broader array of assessments into the determination of their students' grades. Guidelines presented in this book help teachers do this because they are designed to support varied approaches to learning and assessment and to encourage student success, however it is demonstrated.

BRAIN-BASED RESEARCH

The constructivist view of learning has been supported and expanded by what is often called "brain-based research." This research has demonstrated that the way the brain works is much more complex than theories of learning previously acknowledged. Brain research shows that the ability to learn is significantly influenced by coping with emotions and the environment, by being taught the skills of thinking, and by the encouragement of metacognition—that is, thinking about thinking.

> The classroom environment that best facilitates the full development of the intelligences is sometimes called "brain compatible." For the brain to function fully, it is beneficial for the classroom to provide five elements: trust and belonging, meaningful content, enriched environment, intelligent choices, and adequate time. (Chapman, 1993, p. 9)

Assessment and grading practices need to be "brain compatible." Brain-compatible assessment results from paying attention to the same elements:

Trust and belonging occurs when students are comfortable undertaking assessment activities. They need risk-free assessments where they feel they

want to try because the marks will not "count" in grades, as it is understood that they may make mistakes on early attempts and/or take risks that may not be successful. Students need to be in a familiar environment with opportunities to practice each assessment type before the real assessment. It has been shown that unless students have had an opportunity to practice a high-stress activity in an unfamiliar environment, they do not perform at their real level of achievement. For example, it has been demonstrated that students perform better on the SAT when they do the test in their own classroom rather than in the school gym or cafeteria. Grading can be made brain-compatible by using the scores from second chance assessment and by using more recent information to determine grades (see Chapter 5).

Meaningful content and *enriched environment*, from an assessment point of view, mean that teachers provide assessment that promotes learning, not just assessment that is easy to score.

> "Two rules of thumb come from the field of brain research and enrichment. One is to eliminate threat, and the other is to enrich like crazy."
>
> —Jensen, 1998, p. 2

Intelligent choices in assessment means that teachers do not require each student to demonstrate achievement in the same way as other students; students have some choice in how they are assessed. Davies (2007a) goes as far as to suggest that "potentially every student could have a different collection of evidence because every student has taken a different learning journey towards the same standards or outcomes" (p. 49).

Adequate time refers to students' need for time to become comfortable with approaches to instruction and assessment that are new to them. It also means that students need sufficient time to be able to demonstrate their knowledge and skills in assessment situations. Students should only be required to perform in strictly time-limited assessment situations if time is a critical element of the achievement being assessed. Reflective learners and slow writers often receive lower grades than they deserve as a result of being required to perform in inappropriately time-limited assessments (see Chapter 7).

These ideas are supported by the work of Judy Willis (2007), who is both a neurologist and a teacher. She states,

> Brain-based research in learning has given educational researchers the means to translate neuroimaging data into classroom strategies that are designed to stimulate parts of the brain seen to be metabolically activated during stages of information processing, memory, and recall. And what has emerged from the neuroscience of learning over the past two decades is a body of highly suggestive evidence that
>
> - successful strategies teach for meaning and understanding, . . .
> - learning-conducive classrooms are low in threat and high in reasonable challenge, and that
> - students who are actively engaged and motivated devote more brain activity (as measured by metabolic processes) to learning. [bulleted list format added] (p. 698)

Each element of brain-compatible assessment requires that teachers be very flexible in their approach to assessment and grading. If teachers are more flexible, then a greater variety of information will be available to incorporate into their summary judgments. The guidelines in this book are designed to provide teachers with an approach to grading that allows for more than mere number crunching.

MULTIPLE INTELLIGENCES

In the past, intelligence was seen as a singular entity, relatively fixed and easily measured. Gardner (1983) demonstrated that, rather than one fixed entity, there are at least eight intelligences:

- Verbal/linguistic—Words, listening, speaking, dialogues, poems
- Visual/spatial—Images, drawings, doodles, puzzles, visualization
- Logical/mathematical—Reasoning, facts, sequencing, judging, ranking
- Musical/rhythmic—Melody, beat, rap, pacing, blues, classical, jingles
- Bodily/kinesthetic—Activity, try, do, perform, touch, feel, participate
- Interpersonal—Interact, communicate, charisma, socialize, empathize
- Intrapersonal—Self, solitude, create, brood, write, dream, set goals
- Naturalist—Nature, observe, classify, hike, climb, trees, ecosystem

Knowledge of multiple intelligences requires that teachers focus on how smart students are in different ways; the focus is no longer on "how smart" but "how one is smart." Gardner believes that each person's mix of intelligences produces a unique cognitive profile. Educators should ensure that children learn by building on their strengths. Teaching to or through each of the intelligences gives students whose strengths have been undervalued in schools far greater opportunity to succeed.

Understanding multiple intelligences also means that teachers use a wide variety of instructional and assessment activities. One of the best ways to acknowledge individual differences is to encourage students to develop portfolios—purposeful collections of their work—that can show strengths, weaknesses, growth, and progress over time. See Figure 0.1 for ways to use multiple intelligences in both instructional and assessment activities. This figure dramatically illustrates the links that can be made and the incredible variety of activities that are available to teachers to promote student success.

Each of these areas of understanding—constructivism, brain-based research, and multiple intelligences—has contributed to the realization that, in the past, educators have held a very narrow view of learning and knowledge and that this view now needs to be broadened dramatically. Teachers, for example, have focused most commonly on only two intelligences, verbal/linguistic and logical/mathematical, to the exclusion of the other six; students whose strengths are in the other intelligences have frequently not done well in school.

Figure 0.1 Portfolio Activities and Assessments for the Multiple Intelligences

Verbal/ Linguistic	Logical/ Mathematical	Visual/ Spatial	Bodily/ Kinesthetic
• tape recordings of readings • reactions to guest speakers • autobiographies • reactions to films or videos • scripts for radio shows • list of books read • annotated bibliographies	• puzzles • patterns and their relationships • mathematical operations • formulas/abstract symbols • analogies • time lines • Venn diagrams • original word problems	• artwork • photographs • math manipulatives • graphic organizers • posters, charts, graphics, pictures • illustrations • sketches • props for plays • storyboards	• field trips • role playing • learning centers • labs • sports/games • simulations • presentations • dances
Musical/ Rhythmic	**Interpersonal**	**Intrapersonal**	**Naturalist**
• background music in class • songs for books, countries, people • raps, jingles, cheers, poems • musical mnemonics • choral readings • tone patterns • music and dance of different cultures • musical symbols	• group videos, films, filmstrips • team computer programs • cooperative task trios • round robins • jigsaws • wraparounds • electronic mail • class and group discussions • group projects • group presentations	• problem-solving strategies • goal setting • reflective logs • divided journals • metacognitive reflections • independent reading time • silent reflections time • self-evaluations	• outdoor education • environmental studies • field trips • photographs of nature • research on ecosystems • debates on environmental issues • poems about nature

SOURCE: Burke, K., Fogarty, R., and Belgrad, S. (2001). *The Portfolio Connection: Student Work Linked to Standards, 2nd Edition*. Thousand Oaks: Corwin. Used with permission.

WORLD ECONOMY

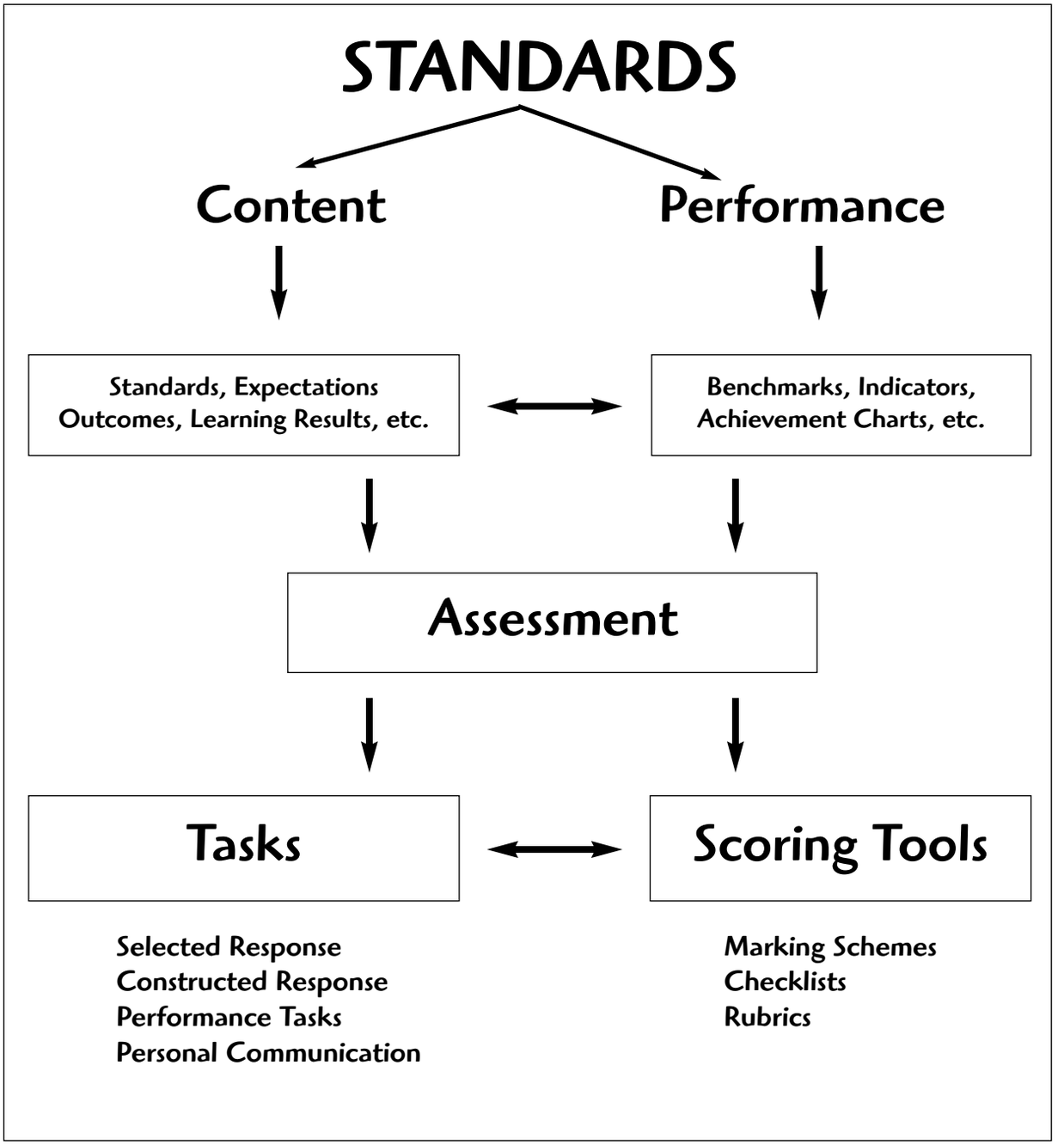
The world economy changed dramatically in the 1990s, and the pace of the change has continued into the new century. Globalization has given unprecedented freedom based on comparative advantage to the flow of capital and jobs between countries. For the developed world, the manufacturing sector has declined, and the service, or tertiary sector, which requires higher levels of skill and knowledge, has enjoyed a huge increase. Thus, far fewer jobs are available for those who do not complete high school.

The sorting function of schools—creating categories of those who leave early and find low-skill jobs, those who complete high school, and those who go on to postsecondary education—does not have the value that it did in the past. What schools now have most commonly is the orientation and expectation that students will succeed. Educators consider themselves to be in "the success business," ensuring that students have real opportunity available to them and that the economy has sufficient skilled and knowledgeable people to continue to function efficiently and effectively. This means that, at least at the school and classroom levels, we have to operate a criterion-referenced system in which the standards are absolute and all can succeed (or fail), *not* a norm-referenced system where standards are relative. The result of a norm-referenced system is a dependable rank order in which a somewhat fixed percentage of students are successful and some students fail regardless of their actual level of performance. Everyone, especially parents, must be helped to understand that in a standards-based, criterion-referenced system, we compare each student's performance to the "standard," not to other students or groups of students, and that there is no place for the normal, or bell, curve in twenty-first-century schools.

STANDARDS

The 1990s saw a huge change in how curriculum is determined. By the end of the decade, 49 of 50 American states, most educational jurisdictions in Canada, and many jurisdictions in other parts of the world had developed mandatory standards for curriculum content. (The 50th state—Iowa—approved, what are in effect, state standards in May 2008.) These standards have a variety of titles, including *standards, expectations, outcomes, learning results,* or *learning goals.* They describe, with varying degrees of clarity and specificity, what students are expected to know and be able to do at different stages in K–12 schooling. The distinguishing characteristic of these statements as compared with previous organizers for curriculum content is that the focus is on outputs—what students will know and do—rather than on inputs—the opportunities that will be provided to students and/or what teachers are expected to do. Generally speaking, standards consist of *content standards* at various grade levels—the what—and *performance standards*—descriptions of how good is good enough. These two types of standards should form the basis of both classroom-level and large-scale assessment. The connections are shown in Figure 0.2.

Figure 0.2 *Assessment, Evaluation, and Reporting Connections Chart*



Schmoker (2000) believes that the "standards movement [is] among the most radical and promising movements in the history of education" (p. 49). Not everyone would give standards such a ringing endorsement, but there is widespread agreement that standards provide a number of benefits:

- Clear focus on what students should know and be able to do
- Common direction for all schools in an educational jurisdiction
- Greater equity in learning goals for all students
- Consistent basis for communication about student achievement to and among stakeholders
- Explicit and external basis for judging the success of teaching and learning

Marzano (2000) notes that the standards movement is not "problem-free." Standards are criticized as

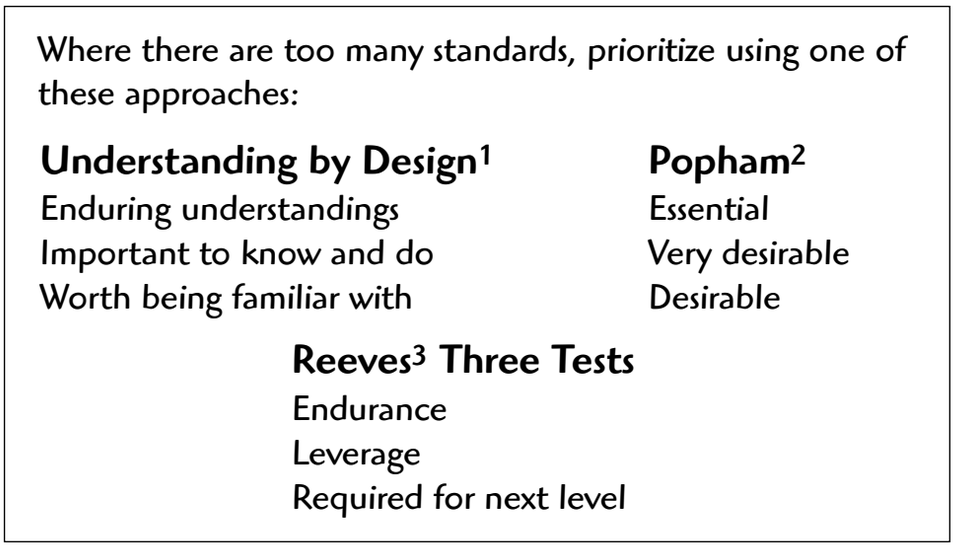
- "glorified wish-lists" (Popham, 2000a)—there are too many standards, and many are not well written or sufficiently succinct;
- straitjackets for teachers who take the life out of the classroom;
- hoops for students to jump over because "the bar has been raised"; and
- responsible for the explosion of testing at district, state, and national levels.

Reeves (2001) suggests that such criticisms of standards and their use are "a good rationale for the improvement of standards[, but] they are not arguments for the rejection of standards" (p. 6). He suggests that "the adoption of standards retains large amounts of . . . discretion, and individual judgment" (1996/1998, p. 2) but that it is appropriate to have some limits on teachers' individual freedom in curricular decision making. Another way of putting this is that the standards determine the *what,* but teachers still have great freedom in determining the *hows.*

Undoubtedly, improvements need to be made in standards, such as a review and revision of the content and number of standards in each jurisdiction. But even though there are too many standards, school districts and teachers can prioritize them so that the more important standards receive appropriate emphasis. This has been done by some states and school districts through the identification of "power standards"—the most important standards from the original, overly long lists. Figure 0.3 shows suggestions for possible classifications and rationale for power standards. Standards will be appropriately emphasized when they are seen as the primary focus for classroom assessment rather than for large-scale assessment (Reeves, 2001).

To be effective, standards-based reform will require the previously mentioned improvements in detail or usage, as well as an approach to lesson and unit design that replaces teachers' absolute individual freedom and the tyranny

Figure 0.3 Prioritizing Standards



[1] From Wiggins, G., and J. McTighe. 1998. *Understanding by Design,* Association for Supervision of Curriculum Development, Alexandria, VA. Pg. 15. Used with permission.
[2] From Popham, W. J. 2000. "Assessing Mastery of Wish-List Content Standards." *NASSP Bulletin,* December, 30–36. Used with permission.
[3] From Reeves, D. B. 2001. "Standards Make a Difference: The Influence of Standards on Classroom Assessment." *NASSP Bulletin,* January, 5–12. Used with permission.

of the textbook with a "design-down" or backwards design approach. This involves the following sequence:

1. Selection of the standard(s) as a base for planning
2. Identification of how and how well students will be expected to demonstrate their knowledge and skills
3. Instructional planning that is focused on "how to get them there"; that is, the instructional strategies, topics, theme, and resources that will be used to illuminate the standards

The logic of "design down" suggests a planning sequence for curriculum. This sequence has three stages:

1. Identify desired results.
2. Determine acceptable evidence of achievement.
3. Plan learning experiences and instruction.

It is important to remember that "[achievement] standards can be raised only by changes that are put into effect by teachers and students in classrooms"

(Black & Wiliam, 1998, p. 148). The move to standards-based systems holds promise, if teachers are assisted appropriately in aligning curriculum instruction, assessment, grading, and reporting. If this alignment occurs, teachers will truly be able to "work smarter, not harder." It will also be easier for teachers to separate their dual classroom roles of coach/advocate and judge because of the clear focus on publicly articulated learning goals known to all. In this context, it should also be easier for students to see assessment as something that is done *with* them (to improve their learning) rather than something that is done *to* them (to find out what they don't know). For maximum benefit to be obtained, the purpose for grades and reporting must clearly be the communication of achievement of the standards. "If grading and reporting do not relate grades back to standards, they are giving a mixed message. Our grading practices must reflect and illuminate those standards" (Busick, 2000, p. 73).

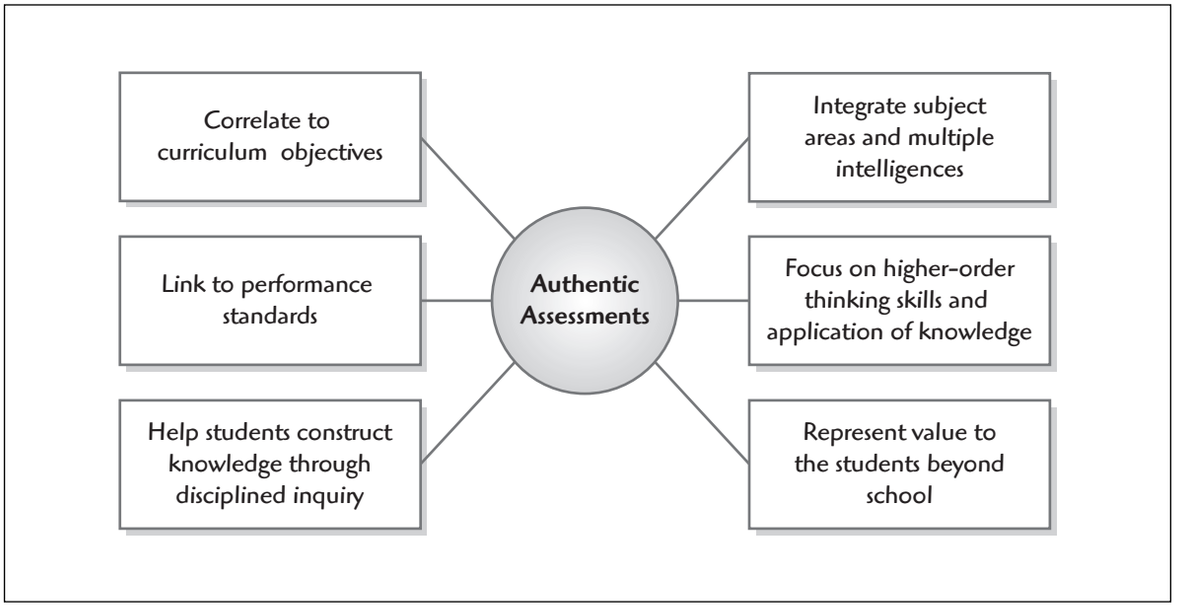
HOW DO THESE CONCEPTS AFFECT CLASSROOM ASSESSMENT?

Global economic changes, together with the development of standards and our new understandings about learning, are leading to significant changes in the ways children are taught and the ways in which they are assessed. There has been a move to authentic learning—learning that is relevant to students and to the real world—and to authentic assessment—assessment that provides students with opportunities to demonstrate what they know, can do, and are like. (See Figure 0.4 for a graphic illustration of the characteristics of authentic assessment.) These approaches have moved classroom assessment away from emphasis on paper-and-pencil methods (especially an almost exclusive reliance on multiple-choice questions) toward the use of a broader array of assessment methods with an emphasis on performance assessment.

All of these changes and their impact on schools lead to the conclusion that "the primary. . . purpose of classroom assessment [must now be] to inform learning, not to sort and select or justify a grade" (McTighe & Ferrara, 1995, p. 11).

The focus of traditional grading practices is to sort, select, and justify. Traditional grading practices emphasize the use of scores from assessments that are easy to quantify, such as selected-response items, especially multiple-choice questions. Teachers "become 'bean counters' . . . adding up all the marks, bonus points, and minus points before using the calculator to divide by the total number of entries—to the second decimal point, of course" (Burke, 1993, p. 140). This approach was consistent with the competitive mentality prevalent in schools and society. However, as McTighe and Ferrara suggest, this approach is not compatible with the role grading could play, given what is now understood about the nature of

Figure 0.4 *Criteria for Authentic Assessment*



SOURCE: Adapted from Burke, Fogarty, and Belgrad (2001).

learning and the type(s) of assessment that encourages and supports real learning. It is, therefore, necessary to move away from traditional grading and, as much as possible, use grading in the service of learning. This book provides many suggestions about ways in which grading can be used to inform learning.

Reflecting on . . . Assessment Methods

Use the checklist shown in Figure 0.5 to identify the assessment methods you use in your classroom.

Figure 0.5 Assessment/Evaluation Checklist

TYPES OF STUDENT ASSESSMENT

Personal Communication

- ❑ Instructional questions
- ❑ Conferences
- ❑ Questionnaires
- ❑ Response journals
- ❑ Learning logs
- ❑ Oral tests/exams

Performance Assessment (using rubrics, checklists, rating scales, and anecdotal records)

- ❑ ***Written Assignments***
  - ❑ Story
  - ❑ Play
  - ❑ Poem
  - ❑ Paragraph(s)
  - ❑ Essay
  - ❑ Research paper
- ❑ ***Demonstrations*** (live or taped)
  - ❑ Role play
  - ❑ Debate
  - ❑ Reading
  - ❑ Recital
  - ❑ Retelling
  - ❑ Cooperative group work
- ❑ ***Presentations*** (live or taped)
  - ❑ Oral
  - ❑ Dance
  - ❑ Visual (photos or video)
- ❑ ***Seminars***
- ❑ ***Projects***
- ❑ ***Portfolios***

Paper-and-Pencil Tests/Quizzes

- ❑ True/false
- ❑ Matching items
- ❑ Completion items
- ❑ Short answer
- ❑ Visual representation
- ❑ Multiple choice
- ❑ Essay style

WHY GRADE?

Reflecting on . . . Grading Purposes

Reflect on why educators grade students and their achievement. List as many purposes as you can. When you have finished your list, number each purpose in your order of priority (1 for highest priority).

Through such reflection and discussion with colleagues, you will almost certainly recognize that clarity about purpose is critical to everything we do. Purpose is like a compass—it provides direction. If there is ever any doubt about what we should do, if we are thinking logically, we focus on our agreed-upon purpose, and what we should do usually becomes clear. Your reflection and discussion will lead you to find that there are many purposes for grading. To understand this fully, it is helpful to consider classifications from two sources. According to Gronlund and Linn in their classic text, *Measurement and Evaluation in Teaching* (1990), there are four general uses for grading:

1. *Instructional uses*—To clarify learning goals, indicate students' strengths and weaknesses, inform about students' personal-social development, and contribute to student motivation

2. *Communicative uses*—To inform parents/guardians about the learning program of the school and how well their children are achieving the intended learning goals

3. *Administrative uses*—To include "determining promotion and graduation, awarding honors, determining athletic eligibility, and reporting to other schools and prospective employers" (Gronlund & Linn, p. 429)

4. *Guidance uses*—To help students make their educational and vocational plans realistically

A second source, Guskey (1996), summarizes the purposes of grading as follows:

- *Communicate* the achievement status of students to parents and others.
- *Provide information* that students can use for self-evaluation.
- *Select, identify, or group* students for certain educational paths or programs.
- *Provide incentives* to learn.
- *Evaluate* the effectiveness of instructional programs (p. 17).

Both of these classifications were developed relative to the broader, double meaning of grading. However, when the narrower, single meaning of grading employed in this book is used, all of the purposes still apply, although some uses apply more to marks than to grades, for example, self-assessment. Also

note that the use of grades, especially traditional grades, for accountability purposes is of limited value.

It is clear from these two classifications that grades have served many different purposes. Therein lies the basic problem with grades—to serve so many purposes, one letter or number symbol must carry many types of information (achievement, effort, behavior, etc.) in the grade. Putting together such a variety of information makes it very difficult to clearly understand what grades mean. To achieve this clarity, a definitive prioritization of the purpose of grades, such as "The primary purpose for grading . . . should be to communicate with students and parents about their achievement of learning goals" (Brookhart, 2004, p. 5), is needed. Brookhart also provides the important caution that we must recognize that "grades are only one way to communicate student achievement and should be used with other feedback methods" (p. 5).

Communication is also the purpose that best fits with what grades are—symbols that summarize achievement over a period of time.

The premise on which this book is based is that communicating student achievement is the primary purpose of grades. If clear communication about student achievement does not occur, then none of the other purposes of grades can be effectively carried out. Communication is also the purpose that best fits with what grades are—symbols that summarize achievement over a period of time. Communication is most effective when it is clear and concise; grades are certainly concise, and they can be clear communication vehicles if there is shared understanding of how they are determined and, thus, what they mean. Instructional and guidance uses not only need to be based on grades with clear meaning, but also are best served by much more information than symbols provide. The administrative uses of grades are really a form of communication and are best served when communication is clear. The other purposes of grades are also best served when communication is the focus. Clarity about student achievement enables all the participants in the educational endeavor to do what is needed to support learning and encourage success.

Acknowledging that the primary purpose of grades is communication helps to point teachers in some very clear directions concerning the ingredients of grades and the use of grades at different levels within the school system. It is essential that there be a shared understanding, at least at the school level but preferably at the district level, about this primary purpose. It is critical, therefore, that teachers have opportunities for professional dialogue about purpose so this shared understanding can be developed. Emphasizing communication about achievement means that clarity is also needed about what achievement is (see Chapter 3). This emphasis is reflected in the analysis of grading and the grading guidelines presented in this book.

WHAT ARE THE UNDERLYING PERSPECTIVES ON GRADING?

The following sections explore seven perspectives, which were developed from a variety of assessment specialists, including Stiggins, McTighe, and Guskey. They provide both a clear indication of the philosophy that underlies the approach to grading advocated in this book and a vehicle for addressing some of the myths about grades and criticisms of grading.

Reflecting on . . . the Seven Perspectives

1. Grading is not essential for learning.
2. Grading is complicated.
3. Grading is subjective and emotional.
4. Grading is inescapable.
5. Grading has a limited research base.
6. Grading has an emerging consensus about best practice.
7. Grading that is faulty damages students—and teachers.

Without reading any further, what is *your* reaction to these seven perspectives? With what perspectives do you agree? Disagree? Which ones are you not sure about? Keep a record of your initial reaction as you read the rest of this section.

Perspective One: Grading Is Not Essential for Learning

Although many teachers appear from their actions to believe otherwise, "teachers do not need grades or reporting forms to teach well, and students can and do learn well without them" (Guskey, 1996, p. 16). Proof of this can be found in cocurricular activities, such as teams and clubs, and in interest courses, such as noncredit night school classes. In each of these situations, excellent teaching and superb learning take place—without grades. The problem in the school system is that, as soon as grades are introduced, teachers, parents, and students emphasize grades rather than learning. Teachers usually say this happens because grades motivate. Kohn (1993b) believes very strongly that grades should be abolished because they serve as extrinsic motivators and destroy positive, intrinsic motivation. Kagan (1994), however, suggests that "if a student is performing a behavior and enjoys it and happens to receive praise or recognition, the recognition will not necessarily erode intrinsic motivation" (p. 16.8). Brookhart (1994) offers another view, saying,

> Cognitive evaluation theory suggests that if students get feedback that helps them make progress, then motivation and control should increase. . . . Students will behave because their efforts will cause learning, and because enhancing perceived competence is motivating in and of itself. Students will perceive grades and other assessments which teachers use to provide informational feedback as more soundly based and reliable than grades and other assessments used to provide controlling feedback. (p. 296)

The issue of motivation and learning is of vital importance in this analysis of grading. It is important to acknowledge several facts:

- Teachers need to learn more about motivation so that they can use knowledge, rather than perception, to guide their practices.
- Students—and parents—have been taught to overvalue grades. Although it will not be easy, if teachers assess and grade better, both may learn to value grades more appropriately.
- Good grades may motivate, but poor grades have no motivational value. In fact, the only grades that motivate are those that are higher than a student expects or usually receives.
- Educators must emphasize that learners are responsible for learning. It is then clear that the learner must be motivated by the intrinsic interest and the worth of what is being learned, not by the carrot-and-stick approach that emphasizes gold stars and A's. Kohn (1993b) suggests that what matters is the three C's of motivation: content (things worth knowing), choice (autonomy in the classroom), and collaboration (learning together).

Further, Stiggins (1999) points out that

> students succeed academically only if they want to succeed and feel capable of doing so. If they lack either desire or confidence, they will not be successful. Therefore, the essential question is a dual one: how do we help our students want to learn and feel capable of learning (p. 192)?

Students' responsibility for their own learning can be achieved most effectively by consciously involving students in the assessment process. Students should be involved in designing or selecting assessment strategies, developing criteria, keeping records of their achievement, and communicating about their learning (Stiggins, 2001b).

Perspective Two: Grading Is Complicated

Much grading is done in a mechanistic way, using formulas and/or computer grading software to produce the final grade merely as the result of arithmetic calculations. Teachers and students, therefore, come to believe that grading is simple; in fact, it is extremely complicated. Grades are shorthand; they are symbols that represent student performance. To arrive at grades, hundreds of decisions must be made along the way; the final grade could be very different if any of those choices is made differently. In particular, the decisions that are made about how the numbers are "crunched," or manipulated, are critical. This issue is addressed in Chapters 6 and 9, with suggestions about how to manipulate numbers in ways that support student learning better than traditional grading practices.

Perspective Three: Grading Is Subjective and Emotional

> "The question is not whether it is subjective, but whether the scoring system is defensible and credible."
>
> —Wiggins, 2000

Rather than looking at the volume and complexity of the decisions about calculations, this perspective focuses on decisions about what is included in grades and the why of calculations. Because grades are usually the result of at least some numerical calculation, teachers often claim that grades are objective measures of student performance.

Kohn (1993b) counters this claim: "What grades offer is spurious precision, a subjective rating masquerading as an objective assessment" (p. 201). Grades are as much a matter of values as they are of science; all along the assessment trail, the teacher has made value judgments about what type of assessment to use, what to include in each assessment, how the assessment is scored, the actual scoring of the assessment, and why the scores are to be combined in a particular way to arrive at a final grade. Most of these value judgments are professional ones; these are the professional decisions that teachers are trained (and paid) to make. It should be acknowledged that grades are, for the most part, subjective, not objective, judgments.

Many other educators support this view. Walvoord and Anderson (1998) say that we should "recognize that there is no such thing as an absolutely objective evaluation based on an immutable standard" (p. 11). Marzano (2000) notes that "the current system based on points and percentages is inherently subjective" (p. 61). Davies (2000) says, "Teachers' professional lives might be more pleasant if evaluating and reporting could be tidy and objective; but they aren't. Evaluation is inherently subjective" (p. 68). Thus, we must acknowledge and not apologize for the subjective nature of grading. What we must strive for are defensible and credible decisions throughout the assessment process.

It should also be acknowledged that, although most teachers' decisions are based on professional judgment, some are based on emotion. Teaching is and, it is hoped, always will be an interpersonal activity. How we feel about the individuals and the groups being assessed sometimes affects our judgment. Again, the point here is not that this is wrong but that all involved need to acknowledge that giving and receiving grades is not a purely objective act—it has a significant emotional component. The subjective and emotional aspects of grades have implications for how grading is done; grading will contribute to more effective learning when this perspective is acknowledged rather than denied.

It is also important to note that "subjectivity becomes detrimental [only] when it translates to bias" (Guskey & Bailey, 2001, p. 330). This does not mean that fairness means treating all students the same. As Gathercoal (2004) points out in his superb book *Judicious Discipline,* "consistency in education is providing the professional specialization and skills needed to help each student believe success is possible" (p. 47). Gathercoal also notes that "the group must learn to trust that decisions regarding exceptions [to the rules] will be fair for all as professional judgments are made for individuals caught up in any number of diverse and often complex circumstances" (p. 46).

Perspective Four: Grading Is Inescapable

Willis (1993) lists the following criticisms of grades:

- Grades are symbols, but what they represent is unclear.
- Grades sort students rather than help them to succeed.
- Grades give little information about student strengths and weaknesses.
- Grades are arbitrary and subjective.
- Grades undermine new teaching practices.
- Grades demoralize students who learn more slowly.

Many educators believe that grades should be abolished. Although this might be desirable, especially for younger students, it simply is not going to happen in the foreseeable future in most educational jurisdictions. In fact, almost everywhere that schools or school systems have tried to remove grades from report cards, they have been faced with community reaction so strongly negative that educators have been forced to return to traditional grades. Olson (1995) gives a clear example of this in a blow-by-blow description of what happened in Cranston, Rhode Island, when a parent-teacher committee proposed a report card without traditional grades for elementary schools. The committee prepared for the change very thoroughly, including piloting the new report cards. However, when the new format was adopted, the uproar in the community forced the school system to return to the former reporting methods.

> "Grading practices are inherently subjective—a fact that constitutes not a denunciation of education but a truth that needs to be told."
>
> —Farr, 2000, p. 14

Wiggins (1996) states that

> trying to get rid of familiar letter grades . . . gets the matter backwards while leading to needless political battles. . . . Parents have reasons to be suspicious of educators who want to tinker with a 120-year-old system they think they understand—even if we know traditional grades are often of questionable worth. (p. 142)

Getting it backwards means that it is inappropriate to focus on trying to eliminate grades; it is more productive to make grades better. Wiggins goes on to say that "what critics of grading must understand [is] that the symbol is not the problem; the lack of stable and clear points of reference in using symbols is the problem" (p. 142). These concerns are addressed in Chapters 1 and 2.

Wiggins (1996) makes another basic point: "Grades or numbers, like all symbols, offer efficient ways of summarizing" (p. 142). Although traditional grades may be of questionable worth, they have a long history. It is not worth fighting against this history; rather, it is worth fighting to make grades meaningful and more supportive of learning. That is what this book is about. Needed are "thoughtfully designed grading and reporting systems that emphasize the formative and communicative aspects of grades [that] can maintain students' focus on important learning goals" (Guskey & Bailey, 2001, p. 20). Furthermore, "the harmful effects of grades can be eliminated by changes in grading systems that provide more chances for success, more guidance, feedback, re-instruction, and encouragement" (Haladyna, 1999, p. 12).

Perspective Five: Grading Has a Limited Research Base

"What a mass and mess it all was." This is how Middleton (quoted in Guskey, 1996, p. 13) described the literature on grading practices—in 1933! Writing in 1995, Reedy said, "Since the introduction of percentage grades in public high schools in the early 1900s, grading and grade reporting have recycled rather than evolved" (p. 47). That there has been no real change over a period of almost 100 years probably stems from the fact that relatively little pure research has been done on grading practices. As can be seen from examining the resources in References and Additional Resources at the back of this book, many journal articles and reports have been written on grading, but most of them, including this book, are summaries of previous work and the opinion(s) of the author(s) on how grading should be done. Logical and well explained as the articles and reports may be, they do not have the weight or authority provided by research. Teachers freely ignore the advice of authors, even those they acknowledge as experts. Stiggins, Frisbie, and Griswold (1989) identified 19 grading practices that measurement experts agreed were desirable. When they examined the actual practices of a group of teachers, they found that teachers ignored the expert advice for 11 of these grading practices. Stiggins et al. suggest three reasons for this situation: recommendations may be opinion or philosophical position rather than established fact, recommendations may be unrealistic in terms of actual classroom practice, and recommendations may be outside the knowledge or expertise base of teachers.

Frary, Gross, and Weber came to similar conclusions in their 1992 study and stated that "large proportions of teachers hold opinions and pursue practices contrary to what many measurement specialists would recommend" (p. 2).

Perspective Six: Grading Has an Emerging Consensus About Best Practice

The lack of a research base and the fact that most methods of grading have advantages and disadvantages means that there is no absolutely right way to grade. The private nature of grading and the dramatic inconsistency in approaches within departments in high schools and colleges and between classrooms in elementary schools means that educators have major problems to address.

This is especially so where grades are "high stakes," that is, when grades serve as more than communication with students and parents. Thus, when grades are the prime or major component of the decision-making process (e.g., for college admission), there needs to be greater consistency, at least within a school and, preferably, across a school district. Despite a lack of "pure" research, "our knowledge base on grading is quite extensive and offers us clear guidelines for better practice" (Guskey & Bailey, 2001, p. 145). There are, therefore, principles that can be agreed on that could lead to consistency across many, or even all, educational jurisdictions. These principles have been articulated as guidelines by authors such as Cooper (2007), Guskey and Bailey (2001), Marzano (2006), Tomlinson and McTighe (2006), Stiggins et al. (for Educational Testing Service Portland; 2004), and Wormeli (2006). Although

there are some differences among the guidelines recommended by these authors, a consensus about best practices, previously lacking, has become evident since about the year 2000. This can be seen by reading the above sources and in the example provided in Appendix C. It is important to note that these are recommendations mostly by classroom assessment specialists, not just educational measurement experts. Please note the similarities—and differences—among the recommendations of these classroom assessment experts.

As Marzano and his colleagues have pointed out in the "What Works" books, we know what to do. Specifically, Marzano notes, "If we follow the guidance offered from 35 years of research, we can enter an era of unprecedented effectiveness for the public practice of education" (p. 1). Both the consensus noted above and the fact that we know what to do is reflected in this book, the basic purpose of which is to provide principled and practical guidelines that all teachers can follow.

Perspective Seven: Grading That Is Faulty Damages Students and Teachers

The flush rose on Alan's face. His hands quivered. "It's not fair," he shouted. "I worked hard. I didn't deserve a B+. This will wreck my chances for Harvard."

Mr. Beaster stood silent. As Alan took a breath, Beaster interjected, "Alan," he began, "your grade . . ."

Alan glared. "It's not my grade. I worked for an A, I deserve an A. I need it. This is the last semester. The good colleges will look at my grades. If you don't give me an A, my class rank drops." Again Mr. Beaster tried to interrupt, but Alan kept on, nostrils flaring, his face now beet red. "You're cheating me," he screeched. "You're ruining my life. My father will kill me. There's no way this grade is OK. If you liked me you would give me an A. You're not fair."

"Alan," countered Beaster, "I'm not going to debate this grade with you. If you want to discuss it when you're calm, I'll be glad to."

"Bull. You'll never change it," Alan pouted as he turned to leave. "You teachers are all alike."

Carmela stared at the floor. Mrs. Martinez sat beside her. Carmela did not move. "Carmela, what am I going to do with you?" Mrs. Martinez asked. "Your grades are getting worse. You are a bright girl. You should be doing better. You are not a D student."

Carmela still did not move. "I do care," she thought, "but it's not so easy. It never has been easy. I've got more to think about than school. School doesn't help me make the dinner or watch my brothers and sisters at night—especially when there is no dinner. And even if I do study, I'm always getting a C or D. So why bother? I can do C or D without studying." (Bellanca, 1992, pp. 297–298)

These two stories illustrate some of the problems with traditional grading practices. Alan had no concept of what good work was or how his grades were calculated. He had developed the idea that school was only about grades, not learning, and that teachers "gave" good grades to students they liked rather than those who produced quality work. Carmela had different problems; there were too many other things in her life for her to be able to show her ability by

producing quality work on demand. Rather than becoming angry, as Alan did, she developed a sense of futility—whatever she did, she would get Cs or Ds, so there was no point in trying to improve.

Overemphasis on grades and faulty grading practices have detrimental effects on student achievement, motivation, and self-concept, as can be seen in these examples. Faulty grading also damages the interpersonal relationship on which good teaching and effective learning depend. This problem occurs at least partly because of teachers' dual roles as coach and judge. Unfortunately, these roles frequently conflict, and as a result, teacher-student relationships are damaged. Many of the problems illustrated by Alan's and Carmela's stories may be at least alleviated and possibly even eliminated if grading practices that support learning and student success are used.

These perspectives on grading contrast with traditional perspectives on grading. Traditional grading is normally seen as being essential for learning ("If I don't give them grades, they won't do the work.") and as straightforward and scientific ("The formula says . . . ; the calculator shows . . ."). If one followed the first three perspectives to their logical conclusion, a strong case could be made against grading; but the fourth perspective means that, as it is virtually impossible to do away with grades, it is necessary to find ways to make grades more meaningful. Here, *making grades more meaningful* means to develop grading practices that support learning and encourage student success. Teachers must not see grades as weapons of control but rather use grading as an exercise in professional judgment to enhance learning. If teachers acknowledge the seven perspectives in their dealings with parents, students, and other teachers, grades can become a positive rather than a punitive aspect of educational practice.

Reflecting on . . . the Perspectives

- Now that you have read about each of the perspectives, what do you think?
- With which perspectives do you now agree? Disagree?
- Which perspectives in the list are you now not sure about?
- How has your thinking changed from when you first read the list?

GRADING PRACTICES AND GUIDELINES

This section actively engages readers in analyzing grading practices. It begins with some factual data about grading practices. Readers then examine their own beliefs about grading and their own grading practices. Case studies provide opportunities to analyze grading practices and identify grading issues: the what, how, and why of grading. Readers might keep a list of the issues that they identify to compare with a list provided in Figure 0.19. Having identified grading issues, one looks for solutions. One solution is practical guidelines that teachers use in their classrooms and in their grade books. A set of eight such guidelines is introduced in this section and examined in detail in Chapters 1 through 8.

How Is Grading Done?

Robinson and Craver (1989, p. 26) reported the use of letter and percentage grades at various grade levels in the United States in 1988. Figure 0.6 shows the usage levels for the two most prevalent grading symbol systems: letters and percentages. The information in Figure 0.6 demonstrates that letter or percentage grades were given to 15 to 20 percent of kindergarten students, 55 to 70 percent of students in Grades 1–3, and 80 to 100 percent of students in Grades 4–12.

A more recent (1997) survey of high schools by the College Entrance Examination Board (1998) reveals that 91 percent of schools were using A–F letter grades or numeric grades and that 93 percent of schools were not considering changing the grading system in the following three years. We are, therefore, examining an educational practice that is a significant fact of life for most students, parents, and teachers in North America.

Figure 0.6 Reported Use of Grading Symbols at Different Grade Levels

Grade	Letter	Percentage
Kindergarten	14.8%	4.8%
Grades 1–3	55.4%	15.6%
Grades 4–6	79.2%	20.7%
Grades 7–9	81.9%	26.8%
Grades 10–12	80.2%	28.5%

*Percentages may total more than 100% because some districts may use more than one grading symbol system at a grade level.

SOURCE: Robinson, G. E., and Craver, J. (1989). *Assessing and Grading Student Achievement.* Arlington, VA: Educational Research Service. Used with permission.

How Do *You* Grade?

"[Grading] practices are not the result of careful thought or sound evidence. . . . Rather, they are used because teachers experienced these practices as students and, having little training or experience with other options, continue their use" (Guskey, 1996, p. 20). This statement may be unfair to some teachers, but it is certainly true for many teachers.

Reflecting on . . . Your Grading Practices

- What are the principles on which your grading practices are based?
- What are your actual grading practices? Do you just crunch numbers?
- What were or are the main influences on your grading principles and practices?
- How do your grading principles and practices compare with those of other teachers in your school?

The best ways to evaluate grading practices (and the principles behind them) are to analyze why we grade, the values on which our grading practices are based, and the sets of marks and/or grades, and to identify the issues that arise from such analyses. Following are 10 case studies that give us the opportunity to analyze grading practices and discover grading issues.

Case Study 1: The Impact of Grades

Case Study 1 begins with reading a short article written by a Grade 11 student. She is obviously a veteran of the grading "wars" and has some very interesting things to say.

AUTHOR'S NOTE: In this article, Kathleen uses the terms *marks* and *grades* interchangeably. This treatment is slightly different from the definitions on page 2.

What's in a Mark?: Report Cards a Chance to Ask Yourself Tough Questions

Edmonton Journal, Friday, February 16, 2007, p. D4
Kathleen Elhatton-Lake

The report card: a source of dread for most of us.

It seems that no mark can be too high, even if you get 100 per cent, you still want 101 per cent. Parents, universities, teachers and students themselves are only propagating the idea that a mark can reflect honesty, perseverance, social skills and, above all, an ability to learn.

Universities need to look at a measure of learning. That's understandable, and teachers need a way of knowing how much their students know.

But here's my question: when did school start becoming less about learning and more about the mark?

Everyone knows the keeners, the people who will beg, plead and kill for a per cent. I'm one of them.

Lately however, I've begun to wonder whether I'm going to school to learn or to get a grade?

A grade can get you scholarships, acceptance into universities and praise from your parents. Yet aren't you supposed to be learning, not cramming information into your head the night before and forgetting it the day after? This is becoming a problem.

I recently received what I thought was an awful mark on an English essay, a mark I have to say I fully deserved. After getting upset and worrying about how this would change my average, I looked at the critique. After going over it, I discovered I learned a lot, probably more than I have from any other essay.

I could have been mad at my teacher, but instead I'm thinking that I am actually glad she's teaching me. So many teachers bend to pressure and give "easy" marks. Or students transfer out of so-called hard classes because their marks are not as high as they would like.

(Continued)

(Continued)

Where do you think you will learn more: a class that pushes you to do your best and earn every mark you receive, or a class where you can coast? Although high marks obtained the easy way may bring short-term gain, there are no long-term payoffs.

Either you hit university and your mark plummets or, if the pattern continues, you again choose the easy route and enter your chosen profession incompetent and a potential danger.

So when you pick up your report card, or when it comes home in the mail, take a deep breath. Open it and think about how you worked this term. Did you slack off or did you put in the hours studying? Do you know your stuff or are you just really good at guessing on that scantron?

In the end, it's what you learned that really counts, not the grade. A mark is supposed to be a reflection of learning and if your reflection is distorted, then you should probably try to fix it. I think we can all be sure that a number cannot reflect everything learned in the hours in a classroom.

As for me, well, I think I'm going to have to spend a bit more time on my English essays.

SOURCE: Kathleen Elhatton-Lake was in Grade 11 at Archbishop MacDonald High School when she wrote this article. She is now (2009) a student at the University of Alberta. Used with permission.

Reflecting on . . . Case Study 1

- What is your reaction to Kathleen's article?
- What issues does it raise for you in relation to your school/district grading policies?

Case Study 2: Interim Report Card Grade

Case Study 2 considers the impact of a zero mark on a grade and the possible impact on a student of grade reporting very early in a course/year.

The marks in Figure 0.7 were given to a student in a senior science class on an interim report card (after four weeks of 76-minute classes) in a school with a semester block schedule.

This case study dramatically illustrates the effect of assigning a zero for a missed test. The student has six marks of 90 percent or higher, two marks in the 80s, and no mark lower than 62.5 percent, but the interim grade was lower than all except the lowest mark! A grade like this could have a devastating effect on students, causing them to give up. This student is achieving well, but the grade suggests otherwise—because of a missed test.

Figure 0.7 *Scores for a Student in a Senior Science Class*

Task	Mark/Total Possible	Percentage
Tests (50%)		
Symbols	16/20	80
Matter	0/68 (absent)	0
Reactions	35/50	70
Daily Work (25%)		
Assignment	10/10	100
Homework	9/10	90
Homework	9/10	90
Atom Quiz	9/10	90
Moles Quiz	5/8	62.5
Homework	9/10	90
Lab Work (25%)		
MP/BP	18/20	90
Superation	20/24	83.3
Reactions	7/10	70
Periodicity Check	10/10	100

Reflecting on . . . Case Study 2

- What grade would you give the student? Why?
- The actual grade the student received was 68.8 percent. What is your reaction to this grade? Was this grade a fair reflection of the student's overall achievement?
- If the zero were not included, the grade would be 81.6 percent. Would this be a fairer reflection of the student's overall achievement?
- What grading issues arise from this case study?

Case Study 3: Chris Brown's Science Class

Case Study 3 considers the marks and grades of a teacher using a very traditional approach to grading. The student marks have been arranged so that, for most students, there are some obvious problems with their performance and/or the way it is graded.

The marks and grades in Figure 0.8 are for Chris Brown's science class in Ontario. If you are not a science teacher, put the appropriate items for your subject in place of the lab reports, care of equipment, and so forth. Note carefully the information that is shown below the grade book extract regarding the miscellaneous items, the way absence is dealt with, and the grading scale.

Enter to the right of the chart the letter grade each student would get using the grading scale in use in your district/school.

In the Ontario class, there are one A, one B, four Cs, and a D—but did they go to the right students? Marg got a D, but on her achievement alone, she probably deserved an A. Lorna got an A but had only a 60 percent average on tests and exams. Is she a weak student who is a teacher's pet—one who receives good marks on the things she can get help on—or is she a very capable student who suffers from severe test anxiety? Kay and Peter have the same grade, but Kay is getting high 80s at the end, whereas Peter is receiving failing marks—is this fair? These are just some of the considerations that arise from an analysis of this case study.

Figure 0.8 Scores in Chris Brown's Science Class

Name	Lab Reports										Total	Tests/Exams			Total	Miscellaneous*						Final Total	Final Grade		Your District
out of	*10*	*10*	*10*	*10*	*10*	*10*	*10*	*10*	*10*	*10*	*100*	*50*	*50*	*100*	*200*	*20*	*20*	*20*	*20*	*20*	*100*	*400*	%	Letter	
Robin	6	6	6	6	5	6	6	7	6	6	60	33	39	81	153	15	15	12	0	10	52	265	66	C	
Kay	2	3	5	5	6	6	7	8	9	10	61	11	29	86	126	15	13	18	10	10	66	253	63	C	
Marg	10	10	A	10	10	10	A	10	A	A	60	50	A	100	150	0	0	0	0	15	15	225	56	D	
Dennis	9	8	9	8	9	10	9	10	8	9	89	24	24	49	97	20	17	17	20	20	94	280	70	B	
Peter	10	10	9	9	8	8	7	7	6	5	79	45	36	32	113	20	10	15	10	5	60	252	63	C	
Lorna	10	10	10	10	10	10	10	10	10	10	100	32	29	59	120	20	20	20	20	20	100	320	80	A	
John	8	8	8	7	9	9	8	9	10	8	84	32	30	57	119	20	8	7	0	5	40	243	61	C	

A = Absent = 0 (for Lab Reports and Tests/Exams)

*Miscellaneous

1-Attendance; 2-Care of Equipment; 3-Attitude/Participation; 4-Notebook; 5-Reading Reports (4x5 marks)

Letter Grade Legend (in Ontario)

A = 80%–100%; B = 70%–79%; C = 60%–69%; D = 50%–59%; F = 0%–49%

SOURCE: Adapted from Todd Rogers, University of Alberta. Used with permission.

Reflecting on . . . Case Study 3

- Do the grades awarded fairly reflect the results from which they were derived for each student?
- If you answered yes . . . For which students? Why?
- If you answered no . . . For which students? Why?
- What grading issues arise from this case study?

Case Study 4: Hiring a Student

Very often, secondary report cards give little more information than the student's grade and a three- or four-word comment. Case Study 4 provides an opportunity to analyze how grades are calculated and whether grades provide meaningful information to potential employers, the students themselves, and their parents.

Figure 0.9 Scenarios for Determining Grades in Auto Mechanics

AUTO MECHANICS

	Student #1 71%		Student #2 52%
	WEIGHTS		
Scenario	*Practical*	*Theory*	*Grade*
A	25%	75%	
Student 1	0/25	71/75	71%
Student 2	25/25	27/75	52%
B	50%	50%	
Student 1	0/50	47/50	47%
Student 2	50/50	18/50	68%
C	75%	25%	
Student 1	0/75	24/25	24%
Student 2	75/75	9/25	84%

Figure 0.9 may appear to present an extreme example, but there have been—and probably still are—many classrooms where situations close to this exist. This case study illustrates the critical connection between teacher's intent and how grades are actually calculated.

Reflecting on . . . Case Study 4

- To which student would you give a job at the local auto repair shop based on the information from Scenario A?
- Study the additional information in Scenarios B and C. Which student would get the job now?
- What grading issues arise from this case study?

Case Study 5: Anita's Grade?

Case Study 5 provides many numbers and, therefore, many possibilities for how grades are calculated—number crunching again!

The teacher of this class bases grades only on unit tests but believes in multiple assessment opportunities, when it is feasible. Thus, on Test 2, there are questions on Unit 1 and Unit 2; on Test 3, there are questions on Units 1, 2, and 3; and on Test 4, there are questions on all four units. This gives students four opportunities to demonstrate their knowledge and skill on Unit 1, three opportunities on Unit 2, two opportunities on Unit 3, and only one opportunity on Unit 4. This approach yields many numbers for Anita, as shown in Figure 0.10.

Figure 0.10 Test Scores for Anita

Unit	Test Score (percentage)			
1	50/100 (50)	30/50 (60)	30/40 (75)	23/25 (92)
2		30/50 (60)	23/33 (70)	21/25 (84)
3			20/30 (67)	19/25 (76)
4				17/25 (72)
Test Average	50%	60%	71%	81%

Using traditional approaches, there are at least three alternatives for calculating the final grade for Anita:

Alternative A: Use the average mark on each test:

$$(50 + 60 + 71 + 81)/4 = 66\%$$

Alternative B: Use the final mark on each unit; that is, the marks for each unit on Test 4:

$$(92 + 84 + 76 + 72)/4 = 81\%$$

Alternative C: Use the mark for the first test on each unit:

$$(50 + 60 + 67 + 72)/4 = 62\%$$

As you can see, these three approaches result in a final grade for Anita that ranges from 62 to 81 percent, a variation of almost 20 percent.

One would hope that the teacher would use Alternative B, because this option provides multiple opportunities and supports the teacher's intent. However, many teachers would use Alternative A, and some would use Alternative C, even though it completely negates the multiple assessment opportunities.

Reflecting on . . . Case Study 5

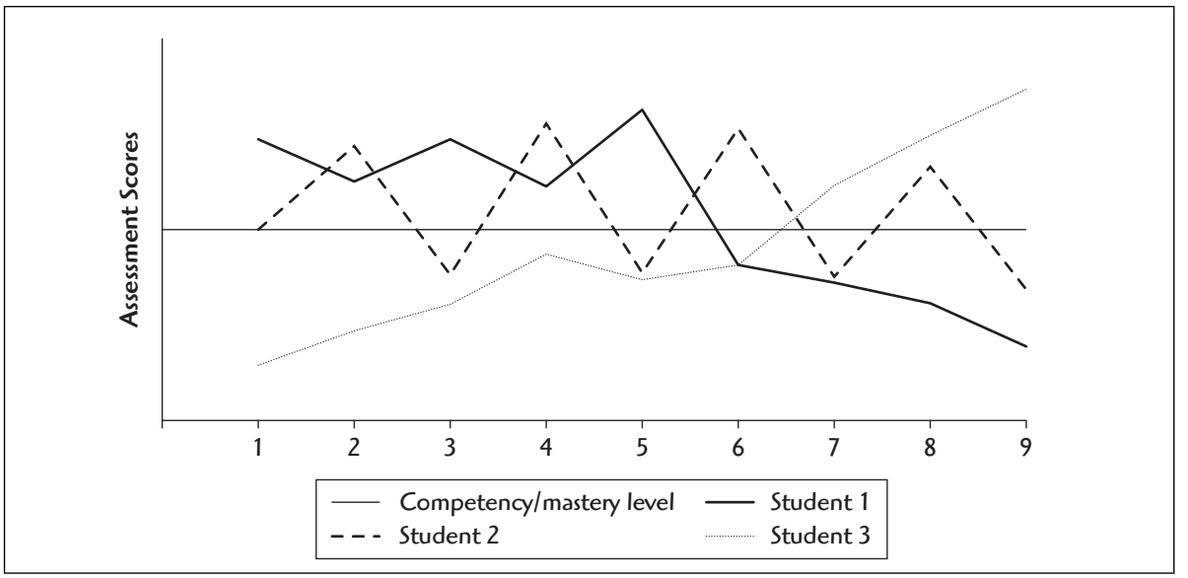
- Which grade would you give to Anita? Would you use Alternative A, B, or C or something else? Why?
- What grading issues arise from this case study?

Case Study 6: All or Some

Another aspect of number crunching is presented in this case study.

Imagine you are going to go skydiving. Presumably, you want a parachute that has a very good chance of opening properly. The skydiving company has provided you with the assessment scores of three students from a recent parachute-packing course. These three are the only people they employ to pack parachutes, so you have to take a parachute packed by one of them—unless you want to jump without a parachute! Please note the competency/mastery level for each assessment, as shown in Figure 0.11, and carefully consider which student you want to pack your parachute.

Figure 0.11 Parachute-Packing Test Scores



SOURCE: Originally developed by Michael Burger. Adapted from A. Davies. 2000. *Making Classroom Assessment Work, Second Edition.* Courtenay, BC, Canada: Connections Publishing.

Reflecting on . . . Case Study 6

- Which student will you choose to pack your parachute? Why?
- If these were scores in a typical teacher's grade book, which students would pass? Which students would fail?
- Is there any discrepancy between your answers to the above two questions? If so, why does this discrepancy occur?

Case Study 7: Grading Scales

What does A mean? What does F mean? For more than 50 years, as a student and as a teacher in Australia and in Canada, I have known that an A has been any grade more than 80 percent and an F has been less than 50 percent. Anything different is very hard for me to comprehend. The familiar becomes the norm—but is it right? Case Study 7 demonstrates that letter grades, honors, and pass/fail mean very different things in different educational jurisdictions.

Figure 0.12 shows grading scales used in North America at five different places. You may use the last row to enter the grading scale used in your district/school.

Figure 0.12 Grading Scales

	Symbol Conversion				
Source	*A*	*B*	*C*	*D*	*F*
Ontario	80–100%	70–79%	60–69%	50–59%	<50%
Ruth Evans*	90–100%	80–89%	70–79%	60–69%	<60%
Rick Werkheiser*	93–100%	85–92%	78–84%	70–77%	<70%
Pam Painter*	95–100%	85–94%	75–84%	65–74%	<65%
R. L. Canady**	95–100%	88–94%	81–87%	75–80%	<75%
Your District	_____	_____	_____	_____	_____

*From the Internet, *The School House Teachers' Lounge* (Nebraska)
**Canady, R. L., and P. R. Hotchkiss. 1989. *Phi Delta Kappan* September, 68–71.

SOURCE: Robert Lynn Canady and Phyllis R. Hotchkiss, "It's a Good Score: Just a Bad Grade," September 1989, pp. 68–71. *Phi Delta Kappan.* Reproduced with permission.

An A can mean anything from 80 to 95 percent, and a failing grade can be anywhere between 49 and 74 percent. What do these variations mean? For example, is a 49 percent in Ontario the same as 74 percent in the district identified by Canady and Hotchkiss (1989) as having the highest grade equivalents? There is no way of knowing this without comparing marked student work from both jurisdictions, but the wide variation makes one wonder about the meaning of grades.

Reflecting on . . . Case Study 7

- What is your reaction to the wide variation in grading scales?
- What letter grade would Anita (Case Study 5) get if she were in each of these school jurisdictions?
- What grading issues arise from this case study?

Case Study 8: Grading Plans

Case Study 8 looks at the recipes teachers use to "cook up" their grades. This case study lets teachers examine how their recipe, or plan, compares with those of their colleagues.

In most traditional grading situations (see Figure 1.1 in Chapter 1), teachers have a recipe for the ingredients in their grades. These usually include some assessment methods and some student behaviors. In addition to the components of grades, such plans usually include some indication of the relative importance of each component by giving it a (percentage) weight.

There is clearly no right answer or perfect grading plan, but for those who teach the same grade or course(s) in the same school and, ideally, in the same school district, it would not be unreasonable to expect that there *would* be some basic similarities or that discernible patterns would exist across their grading plans. If no similarities exist, serious professional discussion about how grading is carried out is needed.

Reflecting on . . . Case Study 8

See Figure 0.13, Grading Inventory, to identify the grading "recipe" you use. Ask colleagues to share their inventories with you. (If your categories do not appear on Figure 0.13, enter your categories at the bottom.)

- What similarities or differences exist between your inventory and your colleagues' inventories?
- Why do the differences exist? Should the differences exist?
- What grading issues arise from this case study?

Case Study 9: Grading Practices That Inhibit Learning

Canady and Hotchkiss (1989) identify 12 grading practices that inhibit learning (see Figure 0.14). Although the article was written almost 20 years ago, many of these are quite common practices that some—maybe even most—teachers would consider acceptable and normal. The fact that Canady and Hotchkiss labeled them as practices that inhibit learning requires teachers to analyze carefully their own grading practices.

The grading practices in numbers 2, 3, 4, 6, and 9 in Figure 0.14 were all part of my practices when I was a classroom teacher. Most teachers will probably admit that they use at least one third of the practices listed at least some of the time. The grading guidelines presented in this book, when fully implemented, eliminate most of these learning-inhibiting practices.

Reflecting on . . . Case Study 9

- Does Figure 0.14 reflect any practices you used in the past?
- What grading issues arise from this case study?

Case Study 10: Standards-Based Grading—What Grade Would Vivian Get?

In standards-based systems, instead of grading each subject by points and percentages based on assessments, as illustrated in Case Study 3, teachers collect evidence for each standard or cluster of learning goals. In the example in Figure 0.15 (page 36), the strands in the language arts curriculum in Florida have been used. Vivian's scores on 10 assessments in the first grading period have been recorded by strand using a scale where 4 means "excels," 3 means "competent," 2 means "approaching or partially competent," and 1 means "well below competency."

Figure 0.13 Grading Inventory

Items Included in Grades	Percentages Allocated			
	Self	Teacher #1	Teacher #2	Teacher #3
Exams				
Tests				
Projects • individual • group				
Demonstrations/Oral Presentations				
Written Assignments • small writing tasks • writing folders or portfolios • essays				
Class Participation and Effort • whole-class discussions • group discussions • homework • notebook • attendance, punctuality				
Peer Assessment				
Self-Assessment				
Additional Categories				

Figure 0.14 Grading Practices That Inhibit Learning

1. **Inconsistent grading scales**	The same performance results in different grades, in different schools or classes.
2. **Worshipping averages**	All of the math to calculate an average is used, even when "the average" is not consistent with what the teacher knows about the student's learning.
3. **Using zeros indiscriminately**	Giving zeros for incomplete work has a devasting effect on averages and often zeros are not even related to learning or achievement but to nonacademic factors like behavior, respect, punctuality, etc.
4. **Following the pattern of assign, test, grade, and teach**	Students are often told to read material and prepare for a test. The real discussion and teaching then takes place—after the test. It is far more logical to teach before testing, but we continue to an alarming extent to follow the pattern of assign, test, grade, and teach.
5. **Failing to match testing to teaching**	Too many teachers rely on trick questions, new formats, and unfamiliar material. If students are expected to perform skills and produce information for a grade, these should be part of the instruction.
6. **Ambushing students**	Pop quizzes are more likely to teach students how to cheat on a test than to result in learning. Such tests are often control vehicles designed to get even, not to aid understanding.
7. **Suggesting that success is unlikely**	Students are not likely to strive for targets that they already know are unattainable to them.
8. **Practicing "gotcha" teaching**	A nearly foolproof way to inhibit student learning is to keep the outcomes and expectations of their classes secret. Tests become ways of finding out how well students have read their teacher's mind.
9. **Grading first efforts**	Learning is not a "one-shot" deal. When the products of learning are complex and sophisticated, students need a lot of teaching, practice, and feedback before the product is evaluated.
10. **Penalizing students for taking risks**	Taking risks is not often rewarded in school. Students need encouragement and support, not low marks, while they try new or more demanding work.
11. **Failing to recognize measurement error**	Very often grades are reported as objective statistics without attention to weighting factors or the reliability of the scores. In most cases, a composite score may be only a rough estimate of student learning, and sometimes it can be very inaccurate.
12. **Establishing inconsistent grading criteria**	Criteria for grading in schools and classes often change from day to day, grading period to grading period, and class to class. This lack of consensus makes it difficult for students to understand the rules.

SOURCE: Robert Lynn Canady and Phyllis R. Hotchkiss, "It's a Good Score: Just a Bad Grade," September 1989, pp. 68–71. *Phi Delta Kappan.* Reproduced with permission.

Figure 0.15 Vivian's First-Quarter Assessment Scores

	Assessment 1	Assessment 2	Assessment 3	Assessment 4	Assessment 5	Assessment 6	Assessment 7	Assessment 8	Assessment 9	Assessment 10	Grade
Language Arts											
Reading	1		2	2			2			2	
Writing		1		2		3		3		3	
Listening, Speaking, and Viewing					4				4	4	
Language		1	2n			3		3		3	
Literature							1				

Summarizing Vivian's first-quarter achievement with such an array of data and such variability in the scores is obviously difficult. If her grade(s) were determined by calculating averages, her overall average would be 2.3, while the strand averages are reading 1.8; writing 2.4; listening, speaking, and viewing 4.0; language 2.4; and literature 1.0. If we did this but reported by letter grades, we would need a scale to convert the numbers to letters, so a decision would have to made about what would be an appropriate scale.

An alternative to the numerical, mechanical approach above would be to look for the most consistent level of achievement while taking particular notice of more recent achievement. If this approach were used, her overall grade would probably be 3, while the strand grades would be 2 for reading; 3 for writing; 4 for listening, speaking, and viewing; 3 for language; and not enough information for literature. Hopefully these grades would be left as numbers with a clear explanation of what each number meant, but as with the averaging example above, if reporting by letter grades were required, a conversion chart would be needed.

- What grade should Vivian get for language arts?
- Should Vivian get a grade for reading; writing; listening, speaking, and viewing; language; and literature?
- If yes, what should be the grade for each strand?
- Should Vivian get an overall grade for language arts or just grades for each strand or standard?

GRADING ISSUES

The case studies have identified the issues listed below:

- Basis for grades: Standards (learning goals) or assessment methods?
- Performance standards: What is good? How good is good enough?
- Ingredients: Achievement, ability, effort, attitude, behavior?
- Sources of information: Methods, purposes?
- What evidence: All or more recent evidence?
- Number crunching: Calculation method?
- Assessment quality and record keeping
- Student understanding and involvement

Although this list is general, I believe it includes all the major grading issues. (An expanded version of this list with the specific concerns that arise out of each issue can be found at the end of this introduction in Figure 0.19.)

Reflecting on . . . Grading Issues

- How does the list of grading issues above compare with your list?
- Which issues that you identified are included?
- Which issues that you identified are not included?

Basis for Grades

Traditionally, grades have been based on assessment methods, but in standards-based systems, it is questionable whether this is the appropriate link. For grades to reflect standards directly and not just by chance, grades must be based directly on the standards. If there are a limited number of standards (no more than seven or eight), grades can be based on the standards themselves. In most standards documents, however, there are so many standards that they need to be organized in some way to provide a manageable amount of information. Another important consideration is that the basis for grades is usually also used as the base for reporting—traditionally a single grade for each subject. Basing grades on standards also gives us the most appropriate base for reporting in standards-based systems—a grade for each standard or learning goal.

Performance Standards

For grades to have any real meaning, they must be based on clear performance standards. This requires some point of reference or comparison: norm, criterion, or self-referenced. Traditionally, grades have been norm-referenced; that is, they were based on comparing the individual with a group. This frequently involved the use of the bell curve or some modification of the curve.

With the introduction of state and local standards, grades should be based on these standards and so become criterion-referenced. Even where there are no published standards, teachers use a criterion-referenced approach when they provide their students with rubrics—scoring scales that clearly indicate the criteria for quality work. Classroom teachers determine most criterion-referenced standards, however, so variability from teacher to teacher is still a major issue. The concerns arising from this problem are discussed in Chapter 2.

Self-referencing, which compares students with their own previous performance, can also provide valuable information.

Ingredients

Teachers have included and mixed many ingredients to arrive at grades. Student characteristics often used in the mix are achievement, ability, effort, attitude, behavior, participation, and attendance. These ingredients are included because grades serve so many purposes. The result is that grades frequently become almost meaningless for their main purpose—communication. This is clearly illustrated in Rick's Mysterious Falling Grade, a case that begins Chapter 3.

To provide effective communication, grades must be clearly understood by the message senders (teachers and schools) and by the message receivers (students, parents, college admissions officers, employers, etc.). "To develop this shared understanding, there must be a consistent and limited basis for what is included in grades; instead of including everything, we must limit the variables or valued attributes that are included in grades" (O'Connor, 1995, p. 94).

Frisbie and Waltman (1992) identify a large set of evaluation variables, which include everything (or almost everything) students do in the classroom and the school. This large set of evaluation variables is reduced to a smaller subset of reporting variables. The size of this subset depends on the type of reporting to parents done by each school district. Care should be taken to

ensure that the most highly valued variables are included. The last step is to select a subset of the reporting variables as the grading variables. The grading variables should be the "status indicators at the end of the learning experience" (Frisbie & Waltman, 1992, p. 38).

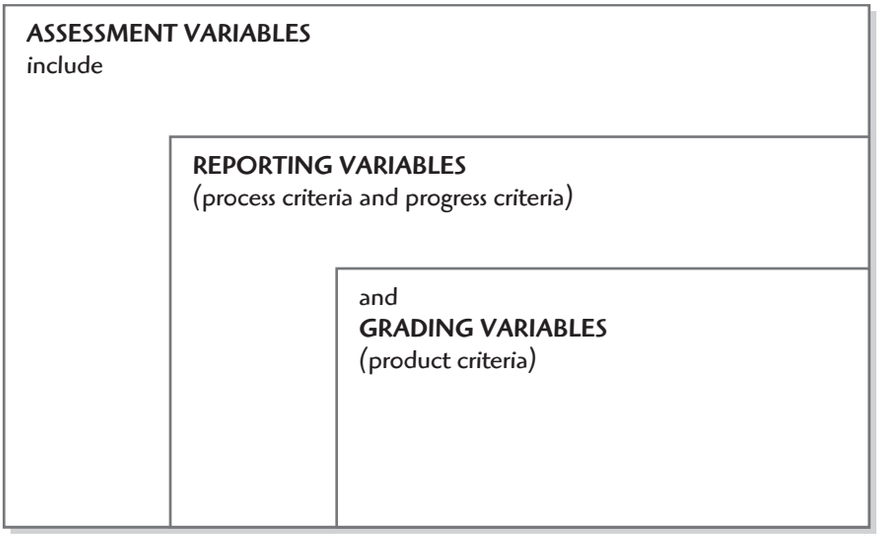
Guskey (1994) provides another approach to identifying the ingredients in grades. He identifies progress criteria (for improvement scoring or learning gain); process criteria (for work habits, attendance, participation, effort, and so forth); and product criteria (for final exams, overall assessments, or other culminating demonstrations of learning).

Frisbie and Waltman's (1992) and Guskey's (1994) concepts are combined in Figure 0.16. Figure 0.16 shows that in Frisbie and Waltman's terms, Guskey's process and progress criteria are the reporting variables, and the product criteria are the grading variables. This combination identifies variables that are separated for grading and reporting purposes. The interaction shown in Figure 0.16, however, is rather simplistic, as some process variables may be assessed over time as part of stated learning goals and, therefore, legitimately may be considered as grading variables. Figure 0.17 illustrates this more complex and more realistic identification of grading and reporting variables. By definition, in standards-based systems, the content standards now define achievement and should be the only grading variables.

Sources of Information

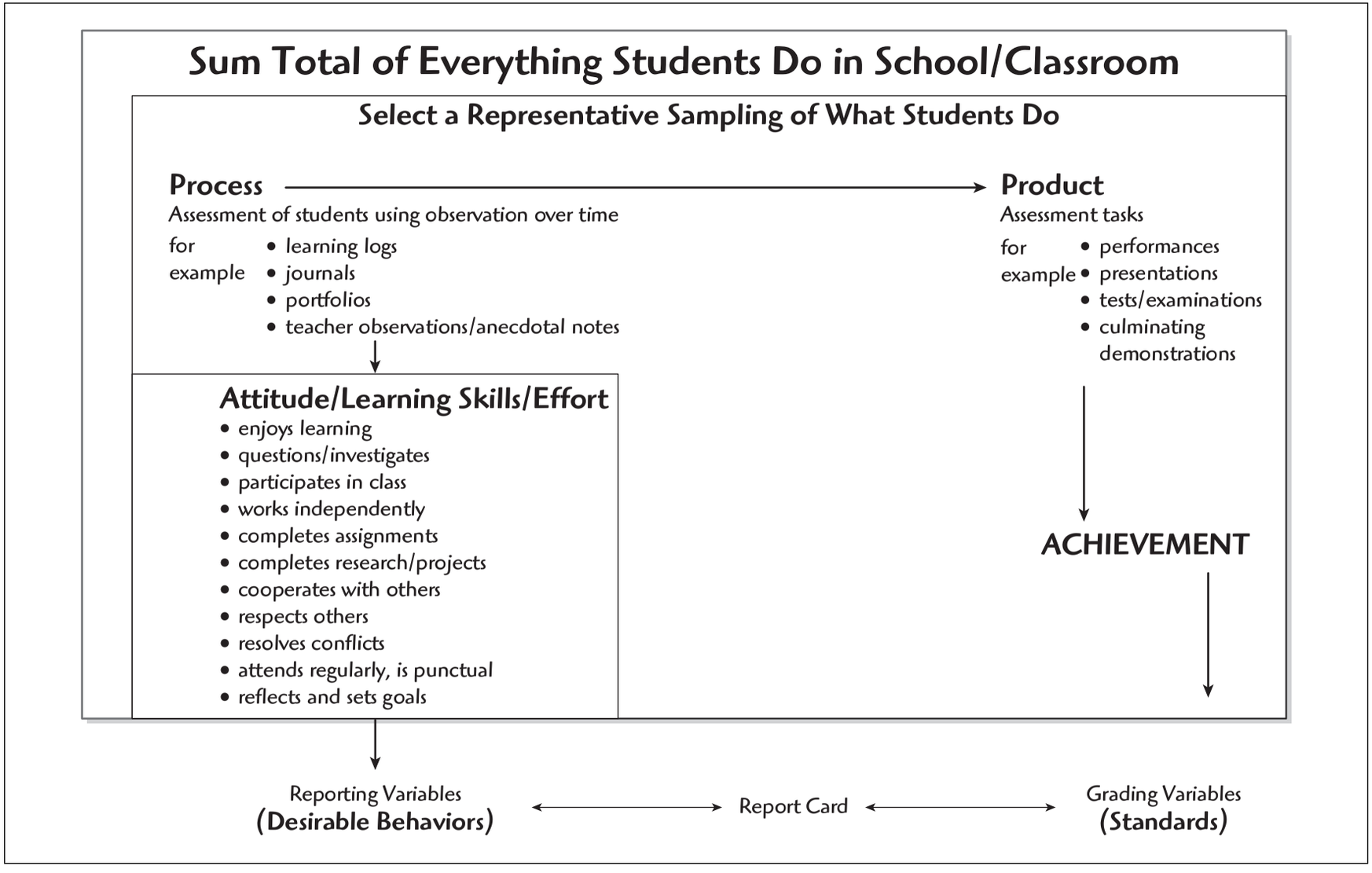
Teachers have many possible sources of information about student achievement. Teachers use a wide variety of assessment methods, but not all sources of information need be included in grades. They decide which sources of information to include based on the reliability and validity of the data and the purpose of the assessment. Teachers make these decisions consciously and carefully.

Figure 0.16 Reporting and Grading Variables



SOURCE: Adapted with permission from K. O'Connor, "Guidelines for Grading That Support Learning and Student Success." *NASSP Bulletin,* (May 1995): 91–101, National Association of Secondary School Principals.

Figure 0.17 Sum Total of Everything Students Do in School/Classroom



SOURCE: Adapted with permission from the work of K. O'Connor and Damian Cooper, Assessment Consultant.

What Evidence—All or Some?

Teachers tend to include everything that they score in student grades. The issue to consider is whether all these data are necessary or appropriate. The amount of data needed is only that which enables confidence that any further information will confirm the previous judgment. Focus should be on the most consistent level of performance, especially toward the end of any learning (grading) period, because this is the information that tells whether the learning goals have been met.

Number Crunching

The case studies demonstrate that numbers can be "crunched" in a variety of ways. They also demonstrate that, depending on the distribution of student scores and the method chosen, students can receive very different grades from the same set of scores. Thus, the methods chosen and a number of other aspects of number crunching, including the weighting of the various components, are very important considerations in the determination of grades.

Assessment Quality and Record Keeping

Because there are many ingredients in grades, even if only achievement information is used, teachers must ensure that the evidence comes from assessments that meet standards of quality. If, for example, assessment is not matched appropriately to teaching, student achievement will be measured incorrectly, and the evidence used to determine grades will be inaccurate.

Record keeping is also important. The complexity of learning goals requires that teachers base grades on complete and accurately tabulated records—on paper, on a computer, or both. It is not justifiable for data that go into a grade to come off the top of a teacher's head at the end of the grading period.

Student Understanding and Involvement

Frequently, students do not understand how the grades they receive are determined. This occurs because either the grading procedures are not discussed with them or the procedures are too complicated to be understood. The issue is how teachers may best ensure that students understand their grades. If grades are to serve learning, students must understand and be involved in the whole assessment process.

GUIDELINES FOR GRADING

Grading issues can be addressed in a variety of ways. To avoid the misuse and misinterpretation of grades, a set of grading guidelines that address the practical concerns of teachers is needed.

Trumbull (2000) notes that a virtual revolution in assessment practices took place in the 1990s but that grading practices evolved only slowly. Traditional grading practices need to change so that grading aligns with standards and supports current assessment and evaluation philosophy and practices.

The grading guidelines in Figure 0.18 were developed with these principles in mind. Some of them require radical changes in teacher practices, especially at the

high school and college levels, and in school and district policies. The guidelines are organized in approximate order of importance to the implementation of standards and to the support of student learning and success (although for this latter purpose, Guideline 8 is critical). The order also relates to where most change from traditional grading practices is needed—relatively few teachers using traditional approaches to grading use Guidelines 1 through 6, whereas many (maybe most) teachers already follow Guidelines 7 and 8. Each guideline stands on its own, but they interconnect significantly and, together, make a coherent group.

The specific relationships between the grading issues identified and the guidelines are shown in Figure 0.19. Each issue relates primarily to one guideline. Some of the specific concerns that arise out of each issue also are listed.

This set of grading guidelines has been modified considerably from those proposed by Gronlund and Linn (1990), but it is important to acknowledge that their list was the starting point. The guidelines are intended to provide practical guidance to teachers as they decide how to grade students' achievement—and they can actually be used by teachers in their grade books or in setting up their computer grading programs. Guidelines also should have school and/or district policy status, so that students and parents can understand the grading practices used in their classrooms and so that they can expect grading practices to be consistent among all teachers in each school. Currently, teachers are "all over the book"; these guidelines should at least get teachers in the same chapter and, eventually, on the same page!

In Chapters 1 through 8, each guideline is examined individually in detail.

Figure 0.18 *Guidelines for Grading in Standards-Based Systems*

To Support Learning
To Encourage Student Success

1. Relate grading procedures to learning goals (i.e., standards).
2. Use criterion-referenced performance standards as reference points to determine grades.
3. Limit the valued attributes included in grades to individual achievement.
4. Sample student performance—do not include all scores in grades.
5. Grade in pencil—keep records so they can be updated easily.
6. Crunch numbers carefully—if at all.
7. Use quality assessment(s) and properly recorded evidence of achievement.
8. Discuss and involve students in assessment, including grading, throughout the teaching/learning process.

A more detailed version of these guidelines can be found in Appendix B: Guidelines for Grading in Standards-Based Systems.

Reflecting on . . . the Guidelines

- What is *your* initial reaction to each of the guidelines for grading in Figure 0.18? Why?
- Think in terms of what is positive, what is a concern, and what is just interesting. List your reflections for later reference.

Figure 0.19 Relationships Between Grading Guidelines and Issues/Concerns

Guideline	Issue(s)	Concern(s)
1	***Basis for grades*** Assessment Methods or Learning Goals	which groupings—standards, strands?
2	***Performance standards*** Norm or Criterion Referenced	how good is good enough? levels or %? to curve or not to curve (bell, that is)?
3	***Ingredients*** Achievement, Behavior(s)	learning skills/work habits/effort late assignments/extra credit group grades/marks
4	***Sources of information*** Formative, Summative Methods	how much data? tests? quizzes? homework? variety: paper and pencil, performance assessment, personal communication
5	***Changing grades***	recent or all information second- or multiple-opportunity assessment
6	***Number crunching*** Mean, Median, Mode	method of calculation how many points on the scale? role of professional judgment effect of zeros/missed work
7	***Quality*** ***Record keeping***	e.g., fairness—time on tests management/tracking system(s)
8	***Student understanding***	clear criteria how much student involvement?