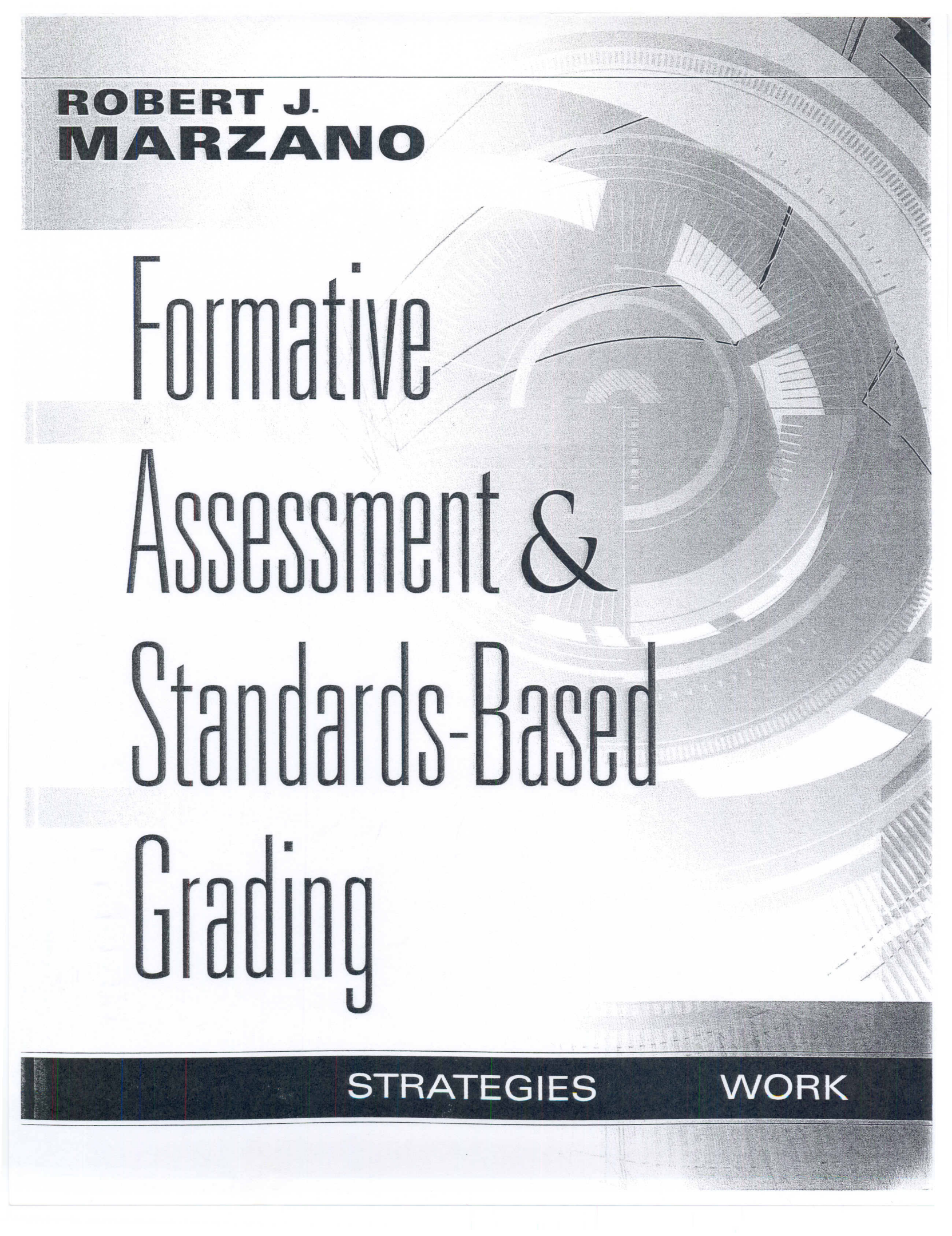




**ROBERT J.
MARZANO**

Formative Assessment & Standards-Based Grading

STRATEGIES

WORK

performance assessments administered in high school biology courses. They focused their review on English language learners, noting that “the language demand of content assessments may introduce construct-irrelevant components into the testing process for EL students” (p. 3). Specifically, they found that the students with a stronger grasp of the English language would perform better on the tests even though they might not have had any better understanding of the science content. The same concept holds true for standardized tests in content-specific areas. They noted that “the language demand of a content assessment is a potential threat to the validity of the assessment” (p. 3).

Grading

At the classroom level, any discussion of assessment ultimately ends up in a discussion of grading. As its title indicates, *Formative Assessment and Standards-Based Grading* focuses on grading as well as on formative assessment. Not only are teachers responsible for evaluating a student’s level of knowledge or skill at one point in time through classroom assessments, they are also responsible for translating all of the information from assessments into an overall evaluation of a student’s performance over some fixed period of time (usually a quarter, trimester, or semester). This overall evaluation is in the form of some type of overall grade commonly referred to as an “omnibus grade.” Unfortunately, grades add a whole new layer of error to the assessment process.

Brookhart (2004) discussed the difficulties associated with grading:

Grades have been used to serve three general purposes simultaneously: ranking (for sorting students into those eligible for higher education and those not eligible); reporting results (accounting to parents the degree to which students learned the lessons prescribed for them); and contributing to learning (providing feedback and motivating students). (p. 23)

While all three purposes are valid, they provide very different perspectives on student achievement.

Since the teachers in many schools and districts have not agreed on any one grading philosophy, they are forced to design their own systems. To illustrate, consider the following grading criteria Thomas Guskey (2009, p. 17) listed as elements frequently included in teachers’ grading practices:

- Major exams or compositions
- Class quizzes
- Reports or projects
- Student portfolios
- Exhibits of students’ work
- Laboratory projects
- Students’ notebooks or journals
- Classroom observations
- Oral presentations
- Homework completion
- Homework quality

- Class participation
- Work habits and neatness
- Effort
- Attendance
- Punctuality of assignments
- Class behavior or attitude
- Progress made

He made the point that because of their different philosophies, different teachers rely on different combinations of these elements to construct an overall grade. For example, one teacher might include major exams, quizzes, class participation, and punctuality of assignments in his or her grading policy while another teacher teaching exactly the same course might include major exams, reports, effort, and attendance. Consequently, the grading schemes for the same course taught by two different teachers might be so different that grades are not comparable from teacher to teacher. In effect, individual teachers' grades are interpretable only in the context of the grading scheme constructed by that specific teacher.

Norm-Referenced Grading

One approach to grading that has been used over the years is to report how students are performing in relation to one another. This might be called *norm-referenced grading*. In his doctoral dissertation, Kenneth Haponstall (2009) pointed out that this might have been the impetus for what was referred to as the "grading system" in the mid-nineteenth century, whereby students were grouped by level of knowledge and skill as well as by age so that teachers might provide more focused instruction to these homogeneous groups. He explained that James Baldwin (1884, in Haponstall, 2009) saw problems with the system even then, pointing out that no standard criteria about how students were "graded" or by whom had been established. The decision was subjective and left to anyone from the superintendent to the school secretary or a member of the board of education.

While few, if any, grading schemes currently in place use a strict norm-referenced approach, vestiges of it can be found in the practices of class rankings and grading on a curve.

Class Rankings

Class rankings are related to the concept of norm-referenced grading. Haponstall pointed out that "with districts using differing measures, including grade weighting for advanced placement classes, grade improvement for special education classes, [and] credit recovery for failed courses, there seems to be no standard method for schools to demonstrate those students who are showing academic excellence" (p. 22).

Lawrence Cross and Robert Frary (1999) noted that even though "grading is a hodgepodge of attitude, effort, and achievement at the middle and high school levels, colleges not only accept the grade point and class ranking in determining enrollment, but many are starting to use these measures exclusively" (as cited in Haponstall, 2009, p. 22). The admissions policies of many colleges exacerbate the practice of class ranking. To illustrate, David Lang (2007) pointed out that states such as California, Florida, and Texas guarantee a certain top percentage of each graduating class admission to a state school. This renders class ranks a high-stakes endeavor, particularly for those ranked too low for a guarantee of admission to a state school (as cited in Haponstall, 2009).

Grading on the Curve

When a teacher grades on the curve, he or she gives the highest grade to the student who performed best on an assessment and then gives every other student a grade by ranking his or her performance accordingly. This system essentially grades students in relation to one another. Thus, it has a basis in norm-referencing. Proponents for grading on the curve maintain that it is fair and equitable because most classes will have a normal distribution of achievement scores in any given subject area (for a discussion, see Brookhart, 2004).

Thomas Guskey (2009), however, maintained that “grading ‘on the curve’ communicates nothing about what students have learned or are able to do” (p. 11). Instead of telling teachers what a student has learned, it simply reports how much or how little he or she learned in relation to his or her fellow students. He also pointed to research by Benjamin Bloom (Bloom, 1976; Bloom, Madaus, & Hastings, 1981) indicating that student achievement does not necessarily follow a normal distribution when teachers exhibit a high level of instructional acumen. Grading students only in relation to one another, therefore, may provide information about a student’s rank in class, but it does not speak to the student’s academic achievement.

Self-Referenced Grading

Self-referenced grading occurs in relation to one’s own past performances. Proponents say that it may reduce competition in classrooms and serve to motivate students (for a discussion, see Brookhart & Nitko, 2007). On first glance, this kind of grading seems to make intuitive sense: the reference point for each student is his or her personal growth and the extent of active engagement in his or her own learning. But Brookhart and Nitko pointed out that this form of grading tends to be used primarily with low-ability students, and while heavily weighting factors such as effort, behavior, attitude, and participation might seem positive, this emphasis is one of the major criticisms of this form of grading. Mixing nonacademic competencies with academic competencies contaminates the meaning of a grade.

Standards-Based Grading

Grading that references student achievement to specific topics within each subject area is growing in popularity. This is called standards-based grading, and many consider this method to be the most appropriate method of grading (for a discussion, see Brookhart & Nitko, 2007, p. 219). Where there is interest in this system, however, there is also quite a bit of poor practice on top of considerable confusion about its defining characteristics.

As described in Marzano (2006), the origins of standards-based reporting can be traced to the concept of a performance standard. The term was popularized in a 1993 report commonly referred to as the Malcom Report in deference to Shirley M. Malcom, chair of the planning group. The report defined a “performance standard” as “how good is good enough” (National Education Goals Panel, 1993, pp. ii–iii). Since then, a popular practice has been to define student performance in terms of four categories: advanced, proficient, basic, and below basic. The scheme has its roots in the work of the National Assessment of Educational Progress. As Popham (2003) noted:

Increasingly, U.S. educators are building performance standards along the lines of the descriptive categories used in the National Assessment of Educational Progress (NAEP), a test administered periodically under the auspices of the federal government. NAEP results permit students’ performances in participating states to be compared. . . . Since 1990, NAEP results

have been described in four performance categories: *advanced*, *proficient*, *basic*, and *below basic*. Most of the 50 states now use those four categories or labels quite similar to them. (p. 39)

The actual practice of standards-based reporting requires the identification of what we have referred to as reporting topics or measurement topics (Marzano, 2006; Marzano & Haystead, 2008). For example, consider the following common measurement topics for language arts at the fourth grade:

Reading

Word recognition and vocabulary

Reading comprehension

Literary analysis

Writing

Spelling

Language mechanics and conventions

Research and technology

Evaluation and revision

Listening and Speaking

Listening comprehension

Analysis and evaluation of oral media

Speaking applications

Here, ten measurement topics are organized under three categories (or strands, as some districts call them): reading, writing, and listening and speaking. For reporting purposes, each student would receive a score of advanced, proficient, basic, or below basic on each of the ten measurement topics. Typically, some type of rubric or scale that describes these levels is constructed for each measurement topic (we discuss this in depth in chapters 3, 5, and 6).

While this system seems like good practice, without giving teachers guidance and support on how to collect and interpret the assessment data with which scores like advanced, proficient, basic, and below basic are assigned, standards-based reporting can be highly inaccurate. Indeed, at the writing of this book, no major study (that we are aware of) has demonstrated that simply grading in a standards-based manner enhances student achievement. However, as the previous discussion illustrates, a fairly strong case can be made that student achievement will be positively affected if standards-based reporting is rooted in a clear-cut system of formative assessments.

Another problem that plagues standards-based reporting is the lack of distinction between *standards-referenced* systems and *standards-based* systems. Grant Wiggins (1993, 1996) was perhaps the first modern-day educator to highlight the differences between a standards-based system and a standards-referenced system. In a standards-based system, a student does not move to the next level until he or she can demonstrate competence at the current level. In a standards-referenced system, a student's status is reported (or referenced) relative to the performance standard for each area of knowledge and skill on the report card; however, even if the student does not meet the performance standard for each topic, he or she moves to the next level. Thus, the vast majority of schools and

districts that claim to have standards-based systems in fact have standards-referenced systems. As we shall see in chapter 6, both systems are viable, but they are quite different in their underlying philosophies. Understanding the distinctions between standards-based and standards-referenced systems helps schools and districts design a grading system that meets their needs.

Translating Research Into Classroom Practice

In subsequent chapters, we draw from the research and theory in this chapter and from sources such as *Classroom Assessment and Grading That Work* (Marzano, 2006) and *Designing and Teaching Learning Goals and Objectives* (Marzano, 2009) to discuss how formative assessment can be effectively implemented in the classroom. We also outline a system of grading that, when used uniformly and consistently, can yield much more valid and reliable information than that provided by traditional grading systems.

As mentioned in the introduction, as you progress through the remaining chapters, you will encounter exercises that ask you to examine the content presented. Some of these exercises ask you to answer specific questions. Answer these questions and check your answers with those provided in the back of the book. Other exercises are more open-ended and ask you to generate applications of what you have read.

In addition to keeping track of progress on each learning goal, students keep track of the number of goals in which they have demonstrated “mastery.” Typically, a score of 3.0 on the scale is considered mastery of the content. The expectation is that by the end of the year, all students will have a score of 3.0 or higher (scores 3.5 or 4.0) on each learning goal. Thus, the continual-improvement approach has a mastery orientation. Each student keeps track of his or her individual progress across all learning goals throughout the year with the intent of achieving a score of 3.0 on each one. Table 5.4 depicts a record-keeping form students might use (a reproducible version of this form is found at the end of this chapter on page 103).

Table 5.4 Continual Progress Report

	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Goal 1: Students will use word roots and prefixes and suffixes to spell and decode unknown words.									
Goal 2: Students will use context clues to decode the meaning of unknown words.									
Goal 3: Students will use common punctuation marks such as commas, semicolons, colons, quotations marks, apostrophes, and parentheses correctly.									
Goal 4: Students will describe slight differences in the meanings of related words.									
Goal 5: Students will describe the meaning and effect of metaphors, similes, and hyperboles.									
Goal 6: Students will appropriately use commonly misused verbs (lay/lie).									
Goal 7: Students will describe both explicit and implicit themes.									
Goal 8: Students will describe complex causal relationships.									
Goal 9: Students will describe complex chronologies.									
Goal 10: Students will make and defend predictions.									

Continued on next page →

Name:	John Mark										
Address:	123 Some Street										
City:	Anytown, CO 80000										
Grade Level:	4										
Homeroom:	Ms. Smith										
Language Arts	2.46	C		Participation		3.40		A			
Mathematics	2.50	B		Work Completion		2.90		B			
Science	2.20	C		Behavior		3.40		A			
Social Studies	3.10	A		Working in Groups		2.70		B			
Art	3.00	A									
				0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Language Arts											
<i>Reading:</i>											
Word Recognition and Vocabulary	2.5										
Reading for Main Idea	1.5										
Literary Analysis	2.0										
<i>Writing:</i>											
Language Conventions	3.5										
Organization and Focus	2.5										
Research and Technology	1.0										
Evaluation and Revision	2.5										
Writing Applications	3.0										
<i>Listening and Speaking:</i>											
Comprehension	3.0										
Organization and Delivery	3.0										
Analysis and Evaluation of Oral Media	2.5										
Speaking Applications	2.5										
<i>Life Skills:</i>											
Participation	4.0										
Work Completion	3.5										
Behavior	3.5										
Working in Groups	3.0										
Average for Language Arts	2.46										
Mathematics											
Number Systems	3.5										
Estimation	3.0										

Figure 6.1 Sample report card depicting a standards-referenced approach.

Continued on next page →

Addition/Subtraction	2.5	
Multiplication/Division	2.5	
Ratio/Proportion/Percent	1.0	
<i>Life Skills:</i>		
Participation	4.0	
Work Completion	2.0	
Behavior	3.5	
Working in Groups	2.0	
Average for Mathematics	2.50	
Science		
Matter and Energy	2.0	
Forces of Nature	2.5	
Diversity of Life	1.5	
Human Identity	3.5	
Interdependence of Life	1.5	
<i>Life Skills:</i>		
Participation	3.0	
Work Completion	1.5	
Behavior	2.5	
Working in Groups	1.0	
Average for Science	2.20	
Social Studies		
The Influence of Culture	3.5	
Current Events	3.0	
Personal Responsibility	4.0	
Government Representation	3.5	
Human and Civil Rights	1.5	
<i>Life Skills:</i>		
Participation	3.5	
Work Completion	3.5	
Behavior	3.5	
Working in Groups	4.0	

Average for Social Studies	3.10								
Art									
Purposes of Art	3.5								
Art Skills	3.0								
Art and Culture	2.5								
<i>Life Skills:</i>									
Participation	2.5								
Work Completion	4.0								
Behavior	4.0								
Working in Groups	3.5								
Average for Art	3.00								

Copyright 2007 Marzano & Associates. All rights reserved. Reprinted from *District Leadership That Works* (Marzano & Waters, 2009, pp. 49–51).

Notice that the report card in figure 6.1 contains measurement topics for life skills. Specifically, each subject area reports scores for the following life skills: participation, work completion, behavior, and working in groups. Although we have not addressed scales for these types of skills, they are easily constructed using the generic scale presented in chapter 3. To illustrate, consider table 6.5 (page 118).

In table 6.5, specific behaviors have been identified that illustrate score 2.0, 3.0, and 4.0 competence for the life skill of work completion at grades 6 through 8. For a listing of other scales for life skills, see Marzano and Haystead (2008).

At the top of the report card in figure 6.1 are traditional omnibus letter grades; these were generated using the schemes described in the first part of this chapter. Summative score (and knowledge gain) bar graphs are reported for each student in the life skills in each subject area. However, the omnibus grade for each subject area does not include these life-skills scores. Specifically, figure 6.1 shows that in language arts, the student received summative scores of 4.0, 3.5, 3.5, and 3.0 for the life skills of participation, work completion, behavior, and working in groups, respectively. These life-skills scores were not included when computing the average scale score of 2.46 for language arts (the average for language arts is reported next to the letter grade for language arts). That average for language arts is based solely on the measurement topics that deal with language arts content. To compute the life-skills grades listed at the top right of the report card, the life-skills scores were averaged across subject areas. For example, the student's average score of 3.40 for the life skill of participation (see the top right section of figure 6.1) was computed using the student's participation scores across the subject areas.

A report card like that in figure 6.1 could be accompanied by a traditional transcript that lists courses taken, credits earned (in the case of high school), and an overall grade point average (GPA). As mentioned previously, this report card (and variations of it) is referred to as "standards-referenced." In a standards-referenced system, students do not have to demonstrate proficiency in each measurement topic to move on to another grade level. Requiring students to demonstrate competence in all content is the domain of a standards-based system.