Classroom
Assessment & Grading
that *Work*

Robert J. Marzano

CLASSROOM
ASSESSMENT & GRADING
that Work

Robert J. Marzano

All Web links in this book are correct as of the publication date below but may have become inactive or otherwise modified since that time. If you notice a deactivated or changed link, please e-mail books@ascd.org with the words "Link Update" in the subject line. In your message, please specify the Web link, the book title, and the page number on which the link appears.

This book is dedicated

to Cliff Nolte:

- Father-in-law

- Friend

- Teacher

CLASSROOM ASSESSMENT & GRADING *that* Work

LIST OF FIGURES

1

The Case for
Classroom Assessment

Improving the academic achievement of K–12 students has been a central concern of educators in the United States since at least the early 1890s, when leaders of industry, politicians, parents, and the society at large realized that an educated populus was the closest thing a country could have to a guarantee of a bright future (Ravitch, 1983). Since that time, a wide array of educational innovations have been tried, all of which were designed to enhance student achievement. Educators have experimented with such things as changing the schedule, decreasing the student-to-teacher ratio, increasing the availability and use of technology, and so on. All of these innovations have merit. However, not even the best has demonstrated the impact on student achievement of the most intuitively important variable in the educational system—the classroom teacher.

Virtually every study that has examined the role of the classroom teacher in the process of educating students has come to the same straightforward conclusion: an effective teacher enhances student learning more than any other aspect of schooling that can be controlled. To illustrate, after analyzing test scores of more than 60,000 students across grades 3 through 5, researchers S. Paul Wright, Sandra Horn, and William Sanders (1997) made the following observation:

> The results of this study will document that the most important factor affecting student learning is the teacher. In addition, the results show wide variation in effectiveness among teachers. The immediate and clear implication of this finding is that seemingly more can be done to improve education by improving the effectiveness of teachers than by any other single factor. Effective teachers appear to be effective with students of all achievement levels, regardless of the level of heterogeneity in their classrooms. If the teacher is ineffective, students under the teacher's tutelage will

show inadequate progress academically regardless of how similar or different they are regarding their academic achievement. (p. 63)

Other studies have corroborated the conclusions of Wright, Horn, and Sanders (for a review of other studies, see Nye, Konstantopoulos, & Hedges, 2004). Kati Haycock (1998) dramatizes the effect of a classroom teacher by comparing what one can expect from a student spending a year with the "most effective teacher" and the "least effective teacher" (for a discussion of how "most effective" and "least effective" teachers are defined, see Technical Note 1.1). Haycock explains that the most effective teacher produces an achievement gain of 52 percentile points in student achievement, whereas the least effective teacher produces a gain of only 14 percentile points—a difference of 38 percentile points. This finding is made even more dramatic when one realizes that it has been estimated that students gain about 6 percentile points in academic achievement simply from growing one year older and gleaning new knowledge and skill from daily life (Cahen & Davis, 1987; Hattie, 1992). The ineffective teacher adds little more than life experience.

Given the potentially strong and positive effect of a classroom teacher, a logical question is, what do highly effective teachers do? Again, many answers have been proposed, most of which focus on lists of instructional and management strategies (see Hattie, 1992; Marzano, Marzano, & Pickering, 2003; Marzano, Pickering, & Pollock, 2001). These lists emphasize the use of strategies such as well-designed practice activities, comparison activities, communicating learning goals, and using pictures, graphs, and pictographs to represent knowledge. Although it is certainly true that "high-yield" instructional strategies and classroom management strategies are a critical part of effective teaching, this book is about one aspect of teaching that is frequently overlooked in discussions of ways to enhance student achievement: classroom assessment.

To the surprise of some educators, major reviews of the research on the effects of classroom assessment indicate that it might be one of the most powerful weapons in a teacher's arsenal. To illustrate, as a result of a synthesis of more than 250 studies, Paul Black and Dylan Wiliam (1998) describe the impact of effective classroom assessment in the following way:

The research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be quite considerable, and as noted earlier, amongst the largest ever reported for educational interventions. As an illustration of just how big these gains are, an effect size of 0.7 [see Technical Note 1.2 for a description of an effect size], if it could be achieved on a nationwide scale, would be equivalent to raising the mathematics attainment score of an "average"

country like England, New Zealand or the United States into the "top five" after the Pacific rim countries of Singapore, Korea, Japan and Hong Kong. (p. 61)

It is important to note that Black and Wiliam's (1998) comments address formative as opposed to summative assessments. This distinction is addressed in the next section. To get a sense of Black and Wiliam's conclusions, consider Figure 1.1 (see Technical Note 1.3 for a description of how Figure 1.1 was derived). The upper part of Figure 1.1 depicts a teacher who begins at the 50th percentile in terms of her skill at using classroom assessments and a student in her class who begins at the 50th percentile in terms of his achievement. Over time the teacher increases her effectiveness at using classroom assessment to the 84th percentile. Given Black and Wiliam's findings, one would predict that the student's achievement would increase to the 63rd percentile. The lower part of Figure 1.1 represents an even more dramatic scenario. If the teacher increases from the 50th to the 99th percentile in terms of skill at using classroom assessments, one would predict the student's achievement to increase to the 78th percentile.

At face value, the findings reported in Figure 1.1 are remarkable—classroom assessment can have a dramatic influence on student achievement. Given these findings, one might be tempted to conclude that assessing students more will automatically increase their learning. Such a conclusion would be wrong. Like most things in education, classroom assessment enhances student achievement under certain conditions only. Fortunately, the research provides some guidance regarding those conditions.

A Brief Review of the Research on Classroom Assessment

Scholars have conducted many reviews of the research on classroom assessment. Some of the more comprehensive reviews are those by Natriello (1987); Fuchs and Fuchs (1986); Crooks (1988); Bangert-Drowns, Kulik, and Kulik (1991); Bangert-Drowns, Kulik, Kulik, and Morgan (1991); Kluger and DeNisi (1996); and Black and Wiliam (1998). The reviews lead to many conclusions that provide insights into effective classroom assessment; however, four generalizations are particularly germane to this book:

• Feedback from classroom assessments should give students a clear picture of their progress on learning goals and how they might improve.
• Feedback on classroom assessments should encourage students to improve.
• Classroom assessment should be formative in nature.
• Formative classroom assessments should be frequent.

FIGURE 1.1
Effect of Teacher's Increased Skill in Classroom Assessment on Student Achievement

Predicted increase in student achievement when teacher's skill in classroom assessment increases from 50th to 84th percentile.
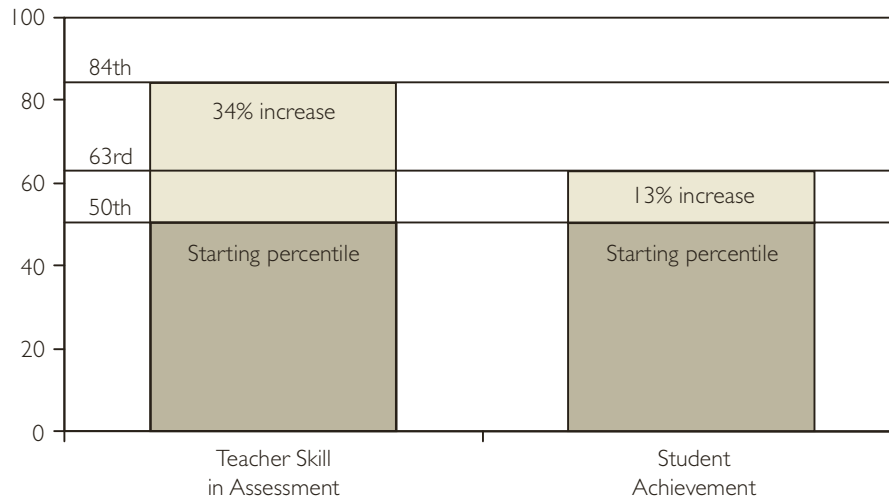
Predicted increase in student achievement when teacher's skill in classroom assessment increases from 50th to 99th percentile.
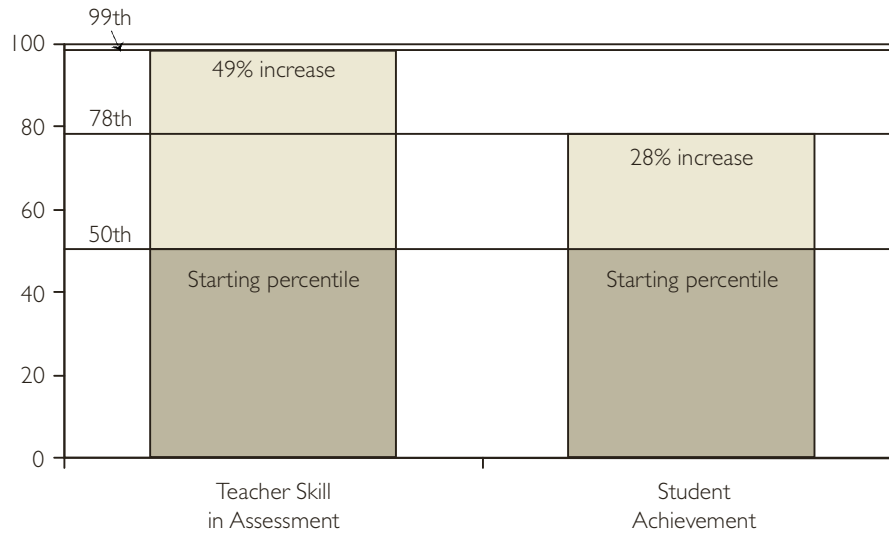
Source	Characteristics of Feedback from Classroom Assessment	Number of Studies*	Effect Size	Percentile Gain or Loss in Student Achievement
FIGURE 1.2				
Findings on the Effects of Different Types of Feedback				
Bangert-Drowns, Kulik, Kulik, & Morgan (1991)	Right/wrong	6	–.08	–3
	Provide correct answer	39	.22	8.5
	Criteria understood by students vs. not understood	30	.41	16
	Explain	9	.53	20
	Repeat until correct	4	.53	20
Fuchs & Fuchs (1986)	Displaying results graphically	89	.70	26
	Evaluation (interpretation) by rule	49	.91	32

*Indicates the number of studies that were examined by the researchers to compute an effect size. See Technical Note 1.2 for discussion of an effect size.

Providing a Clear Picture of Progress and How to Improve

At a basic level, classroom assessment is a form of feedback to students regarding their progress, and it stands to reason that feedback will enhance learning. Indeed, as a result of reviewing almost 8,000 studies, researcher John Hattie (1992) made the following comment: "The most powerful single modification that enhances achievement is feedback. The simplest prescription for improving education must be 'dollops of feedback' " (p. 9).

As compelling as Hattie's comments are, all forms of feedback are not equally effective. In fact, some forms of feedback might work against learning. To illustrate, consider the research findings depicted in Figure 1.2. The figure presents findings from two major meta-analytic studies—one by Robert Bangert-Drowns, Chen-Lin Kulik, James Kulik, and Mary Teresa Morgan (1991), which reviewed 40 studies on classroom assessment; and one by Lynn Fuchs and Douglas Fuchs (1986), which reviewed 21 studies of assessment. The findings from these two synthesis studies as depicted in Figure 1.2 help one understand this first principle of effective classroom assessment.

Consider the first five rows of Figure 1.2, from the Bangert-Drowns, Kulik, Kulik, and Morgan synthesis. Row 1 indicates that when students receive feedback on a classroom assessment that simply tells them whether their answers are correct or incorrect, learning is negatively influenced. This finding is illustrated by the loss of 3 percentile points. However, when students are provided with the correct

answer, learning is influenced in a positive direction. This practice is associated with a gain of 8.5 percentile points in student achievement, as shown in Row 2.

Row 3 of Figure 1.2 addresses whether students are clear about the criteria used to judge their responses. Clarity regarding scoring criteria is associated with a gain of 16 percentile points in student achievement. Row 4 reports a particularly interesting finding—providing students with explanations as to why their responses are correct or incorrect is associated with a gain of 20 percentile points in student achievement. Finally, Row 5 indicates that asking students to continue responding to an assessment until they correctly answer the items is associated with a gain of 20 percentile points.

Rows 6 and 7 of Figure 1.2 are from the Fuchs and Fuchs (1986) study. Row 6 shows the effect of graphically displaying student results. As we shall see in subsequent chapters, displaying assessment results graphically can go a long way to helping students take control of their own learning. However, this practice can also help teachers more accurately judge students' levels of understanding and skill, and it is associated with a gain of 26 percentile points in student achievement. Presumably, seeing a graphic representation of students' scores provides teachers with a more precise and specific frame of reference for making decisions about next instructional steps. Finally, Row 7 addresses the manner in which the teacher interprets assessment results. If the interpretation is done by a set of "rules," student achievement is enhanced by 32 percentile points. In Chapter 3 we will consider this issue in depth. Briefly, though, this finding implies that teachers within a school or a district should have rigorous and uniform ways of interpreting the results of classroom assessments.

Encouraging Students to Improve

One perplexing finding from the research literature is that the manner in which feedback is communicated to students greatly affects whether it has a positive or a negative effect on student achievement. This was one of the major conclusions of a meta-analysis conducted by Avraham Kluger and Angelo DeNisi (1996). After analyzing 607 experimental/control comparisons involving some 23,000 students, the researchers found that in 33 percent of the studies they examined, feedback had a negative impact on achievement. One causal factor they identified for this paradoxical effect is whether feedback encourages or discourages students. To illustrate, Kluger and DeNisi found that when assessment feedback is discouraging to students, it has an effect size of *negative* .14. This translates into a *decrease* in student achievement of 5.5 percentile points (see Technical Note 1.2 for a discussion of effect sizes).

Of course, the critical question that emerges from this finding is, what constitutes encouraging versus discouraging feedback? Kluger and DeNisi warn that this question has no simple answers, but the research provides some strong guidance. To understand the dynamics of encouraging versus discouraging feedback, we must consider two foundational aspects of motivation theory—drive theory and attribution theory.

Drive theory postulates that much of human motivation can be explained as a function of two competing forces, or drives—the striving for success and the fear of failure (Atkinson, 1957, 1964, 1987; Atkinson & Raynor, 1974). Over time, people develop tendencies toward one drive or the other—to be either *success oriented* or *failure avoidant*. When these tendencies become habituated, they translate into strong expectations regarding new tasks—particularly tasks that are challenging to a student.

Success-oriented students tend to be encouraged by challenges because they anticipate the positive feelings that accompany success. Failure-avoidant students tend to be discouraged by challenges because they anticipate the negative feelings that accompany failure. In fact, failure-avoidant students might use self-handicapping strategies that ensure they fail for reasons other than lack of ability. These strategies include procrastination (Rothblum, Solomon, & Murakami, 1986; Solomon & Rothblum, 1984), setting unattainable goals so that failure is ensured (Snyder, 1984), and admitting to small weaknesses or handicaps to establish an excuse for failing—establishing an "academic wooden leg" (Covington, 1992; Covington, Omelich, & Schwarzer, 1986).

Attribution theory provides another perspective on encouraging versus discouraging feedback. It postulates that the manner in which students explain or attribute failure and success encourages or discourages them (Weiner, 1972, 1974; Weiner et al., 1971). In general, individuals attribute their success to four causes: ability, effort, luck, and task difficulty. Of these, the attribution of effort provides the most encouragement. As Martin Covington (1992) explains:

> One of the most important features of attribution theory is its focus on the role of effort in achievement. This emphasis is justified for several reasons. For one thing, if students believe their failures occur for a lack of trying, then they are more likely to remain optimistic about succeeding in the future. For another thing, trying hard is known to increase pride in success and to offset feelings of guilt at having failed. And, perhaps most important of all, the emphasis on the role of effort in achievement is justified because it is widely believed that student effort is modifiable through the actions of teachers. (p. 16)

A fairly straightforward relationship exists between attribution theory and drive theory. Specifically, students who tend to be success oriented also tend to

believe in the effort attribution. They perceive that working hard will bring them success. Thus they have a way to succeed, even when faced with challenging tasks. One of the more encouraging aspects of attribution theory is that students who do not believe their efforts produce success can learn over time that they do. Martin Seligman (Seligman, 1975; Seligman, Maier, & Greer, 1968) postulates that students can even cultivate an "explanatory style" that is effort oriented, if they have enough direct experience that effort produces success. Seligman refers to this dynamic as "learned optimism."

Drive theory and attribution theory provide plausible explanations as to why assessment feedback might be encouraging to some students and discouraging to others. Assume that a student has done poorly on an assessment. If the student is failure avoidant, the negative outcome will strengthen the student's belief that he cannot succeed at challenging tasks and the negative feelings associated with such tasks. This combination will most likely discourage the student. However, if the student is success oriented, poor performance on the test will not be as discouraging because the student has a strategy for improvement—to work harder.

In short, drive theory tells us that classroom assessment that is encouraging must have two characteristics. First, teachers must provide students with a way to interpret even low scores in a manner that does not imply failure. If not, failure-avoidant students will continually be discouraged when they do not receive high scores. Second, teachers must provide students with evidence that effort on their part results in higher scores.

The Value of Formative Versus Summative Assessments

The terms *formative* and *summative* are frequently used in discussions of educational assessments. Actually, the concepts of formative and summative assessment when first developed had little to do with classroom assessment or even with learning.

The distinction between formative and summative assessment was first popularized by Michael Scriven in 1967 as part of an American Educational Research Association monograph series on evaluation. Scriven's original point was that a distinction should be made between programs that are being formulated versus programs that have evolved to their final state. Consequently, evaluation takes on different characteristics and is interpreted differently in formative versus summative situations. This distinction was soon applied to the assessment of students. Specifically, *formative assessment* was defined as occurring while knowledge is being learned. *Summative assessment* was defined as occurring at the end of a learning episode—for example, at the end of a course (see McMillan, 2000). More formally,

Peter Airasian (1994) defines formative assessments as those that "are interactive and used primarily to form or alter an ongoing process or activity. In contrast, assessments that come at the end of a process or activity, when it is difficult to alter or rectify what has already occurred, are called summative assessments" (pp.135–136).

Although the terms *formative* and *summative* have both been widely used in literature on classroom assessment, formative assessment has received more attention in the research literature. Specifically, formative classroom assessment has been the focus of almost every major attempt to synthesize the research on classroom assessment. Recall the finding from Black and Wiliam's (1998) synthesis of more than 250 studies that formative assessments, as opposed to summative ones, produce the more powerful effect on student learning. In his review of the research, Terrance Crooks (1988) reports that effect sizes for summative assessments are consistently lower than effect sizes for formative assessments. In short, it is formative assessment that has a strong research base supporting its impact on learning.

Unfortunately, within the research literature, formative assessment is not defined consistently. As Black and Wiliam (1998) note, "Formative assessment does not have a tightly defined and widely accepted meaning" (p. 7). For the purposes of this book, I use the definition offered by Black and Wiliam that formative assessment encompasses "all those activities undertaken by teachers and/or by students which provide information to be used as feedback to modify the teaching and learning activities in which they engage" (pp. 7–8). This definition casts a wide net in terms of both types of activities that qualify as assessments and the timing of those activities. By definition, then, formative classroom assessment can and should begin immediately within a learning episode and span its entire duration. Additionally, formative classroom assessment can take a wide variety of formats, both formal (e.g., paper-and-pencil quiz) and informal (e.g., a discussion with a student).

The Importance of Frequency

One of the strongest findings from the research is that the frequency of assessments is related to student academic achievement. This finding was dramatically demonstrated in the meta-analysis by Robert Bangert-Drowns, James Kulik, and Chen-Lin Kulik (1991). They analyzed findings from 29 studies on the frequency of assessments. Their findings are depicted in Figure 1.3.

To interpret the figure, assume that we are examining the learning of a particular student taking a 15-week course. (For a discussion of how this figure was constructed, see Technical Note 1.4.) The figure depicts the increase in learning

FIGURE 1.3
Gain Associated with Number of Assessments over 15 Weeks

Number of Assessments	Effect Size	Percentile-Point Gain
0	0	0
1	.34	13.5
5	.53	20.0
10	.60	22.5
15	.66	24.5
20	.71	26.0
25	.78	28.5
30	.80	29.0

Note: Effect sizes from data reported by Bangert-Drowns, Kulik, and Kulik (1991).

that one might expect when differing numbers of formative assessments are used during that 15-week session. If the teacher uses 5 assessments, a gain in student achievement of 20 percentile points is expected; if the teacher administers 25 assessments, a gain in student achievement of 28.5 percentile points is expected; and so on. Bangert-Drowns, Kulik, and Kulik (1991) comment on a number of aspects of this finding. First, they emphasize the relatively strong effect of a single assessment—13.5 percentile points, as depicted in Figure 1.3. Second, they highlight the fact that the frequency effect of assessment tapers off over time. As shown in Figure 1.3, the effect of assessment jumps dramatically from 0 to 10 assessments and then tends to level off. The recommendation from Bangert-Drowns, Kulik, and Kulik is not that teachers should use 30 assessments over a 15-week period but that teachers should systematically use classroom assessments as a form of feedback.

Fuchs and Fuchs (1986) reported this same phenomenon in their meta-analysis of 21 controlled studies. They reported that providing two assessments per week results in an effect size of .85, or a percentile gain of 30 points. Although there is no set number of assessments that should be administered during a unit of instruction or a grading period, the message from the research is clear: systematic use of classroom assessments—weekly or even more frequently—can have a strong positive effect on student achievement.

Summary and Conclusions

Research supports the conclusion that formative classroom assessment is one of the most powerful tools a classroom teacher might use. Formative assessments are defined as any activity that provides sound feedback on student learning. Characteristics of sound feedback include that it should be frequent, give students a clear picture of their progress and how they might improve, and provide encouragement. In the remaining chapters, these principles are used to design a comprehensive system of effective classroom assessment.