ROBERT J. MARZANO

Formative Assessment & Standards-Based Grading

THE CLASSROOM STRATEGIES SERIES

555 North Morton Street
Bloomington, IN 47404

888.849.0851
FAX: 866.801.1447

email: info@marzanoresearch.com
marzanoresearch.com

Visit **marzanoresearch.com/classroomstrategies** to download reproducibles in this book.

Printed in the United States of America

Library of Congress Control Number: 2009908602

ISBN: 978-0-9822592-2-1 (paperback)

Vice President of Production: Gretchen Knapp

Managing Production Editor: Caroline Wise

Senior Production Editor: Risë Koben

Proofreader: Sarah Payne-Mills

Text Designer: Amy Shock

Cover Designer: Pamela Rude

MARZANO RESEARCH DEVELOPMENT TEAM

Staff Writer

Lindsay A. Carleton

Marzano Research Associates

Chris Briggs-Hale

Jane K. Doty Fischer

Maria C. Foseid

Mark P. Foseid

Tammy Heflebower

Edie Holcomb

Sharon Kramer

David Livingston

Beatrice McGarvey

Diane E. Paynter

Debra Pickering

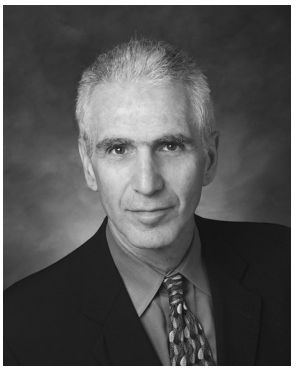
Salle Quackenboss

Tom Roy

CONTENTS

Italicized entries indicate reproducible pages.

ABOUT THE AUTHOR



Dr. Robert J. Marzano is the cofounder and CEO of Marzano Research in Denver, Colorado. Throughout his forty years in the field of education, he has become a speaker, trainer, and author of more than thirty books and 150 articles on topics such as instruction, assessment, writing and implementing standards, cognition, effective leadership, and school intervention. His books include: *The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*, *Making Standards Useful in the Classroom*, *District Leadership That Works: Striking the Right Balance*, *Designing and Teaching Learning Goals and Objectives*, and *On Excellence in Teaching*. His practical translations of the most current research and theory into classroom strategies are internationally known and widely practiced by both teachers and administrators. He received a bachelor's degree from Iona College in New York, a master's degree from Seattle University, and a doctorate from the University of Washington.

ABOUT MARZANO RESEARCH

Marzano Research is a joint venture between Solution Tree and Dr. Robert J. Marzano. Marzano Research combines Dr. Marzano's forty years of educational research with continuous action research in all major areas of schooling in order to provide effective and accessible instructional strategies, leadership strategies, and classroom assessment strategies that are always at the forefront of best practice. By providing such an all-inclusive research-into-practice resource center, Marzano Research provides teachers and principals the tools they need to effect profound and immediate improvement in student achievement.

INTRODUCTION

Formative Assessment and Standards-Based Grading is the second in a series of books collectively referred to as *The Classroom Strategies Series.* The purpose of this series is to provide teachers as well as building and district administrators with an in-depth treatment of research-based instructional strategies that can be used in the classroom to enhance student achievement. Many of the strategies addressed in this series have been covered in other works such as *The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction* (Marzano, 2007), *Classroom Assessment and Grading That Work* (Marzano, 2006), and *Classroom Instruction That Works* (Marzano, Pickering, & Pollock, 2001). Although those works devoted a chapter or a part of a chapter to particular strategies, *The Classroom Strategies Series* devotes an entire book to an instructional strategy or set of related strategies.

Designing effective assessments is critical for any teacher. In order to make judgments about the status of a student or an entire class at any given point in time, teachers need as much accurate data as possible about an individual student's progress, or the progress of the class as a whole, to determine their next instructional steps. As straightforward as this might sound, designing assessments, using them purposefully, and incorporating them into a system of overall grading take insight and practice. *Formative Assessment and Standards-Based Grading* addresses the misconceptions about formative assessment and how it can be used in an overall grading scheme.

We begin with a brief but inclusive chapter that reviews the research and theory on formative assessment, instructional feedback, and grading. Although you might skip this chapter and move right into those that provide recommendations for classroom practice, you are strongly encouraged to examine the research and theory, as it is the foundation for the entire book. Indeed, a basic purpose of *Formative Assessment and Standards-Based Grading* and other books in *The Classroom Strategies Series* is to present the most useful instructional strategies that are based on the strongest research and theory available.

Because research and theory can provide only a general direction for classroom practice, *Formative Assessment and Standards-Based Grading* (and each book in the series) goes one step further to translate that research into applications for the classroom. Specifically, it addresses misconceptions about formative assessment, provides formatively based classroom assessment strategies, and discusses in depth how those strategies can effect change in overall grading systems on both small and large scales. It is important to note, however, that individual teachers, schools, and districts must make necessary adaptations to meet the unique needs of their students.

How to Use This Book

Formative Assessment and Standards-Based Grading can be used as a self-study text that provides an in-depth understanding of how to design and interpret assessments and use those assessments to develop meaningful grades. As you progress through the chapters, you will encounter exercises. It is important to complete these exercises and then compare your answers with those in the back of the text. Such interaction provides a review of the content and allows you to examine how clearly you understand it.

Teams of teachers or an entire faculty that wishes to examine the topics of assessment and grading in depth may also use *Formative Assessment and Standards-Based Grading*. When this is the case, teacher teams should do the exercises independently and then compare their answers in small-group and large-group settings.

Chapter 1

RESEARCH AND THEORY

Assessment and grading are two of the most talked about and sometimes misunderstood aspects of K–12 education. *Formative Assessment and Standards-Based Grading* seeks to bring some clarity to one particular type of assessment—formative—and explore through recommendations how it interacts with traditional and nontraditional grading practices. In this chapter, we review the research and theory that underpin these recommendations. We begin by discussing feedback, the practice in which both assessment and grading have their roots.

Feedback

The topic of feedback and its effect on student achievement is of great interest to researchers and practitioners. In fact, studies on the relationship between the two are plentiful and span about three decades. In an effort to operationally define feedback, researchers John Hattie and Helen Timperley (2007) explained that its purpose is "to reduce discrepancies between current understandings and performance and a goal" (p. 86). Researcher Valerie Shute (2008) said feedback is "information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning" (p. 154).

Feedback can be given formally or informally in group or one-on-one settings. It can take a variety of forms. As the preceding definitions illustrate, its most important and dominant characteristic is that it informs the student, the teacher, and all other interested parties about how to best enhance student learning.

Table 1.1 (page 4) presents the results from a variety of studies on feedback. The first column lists the major studies that have been conducted since 1976. The last three columns are related. Critical to understanding exactly how they are related are the concepts of meta-analysis and effect size (ES). Appendix B (page 153) explains the concepts of meta-analysis and effect size in some depth. Briefly though, *meta-analysis* is a research technique for quantitatively synthesizing a series of studies on the same topic. For example, as table 1.1 indicates, Kluger and DeNisi (1996) synthesized findings from 607 studies on the effects of feedback interventions. Typically, meta-analytic studies report their findings in terms of average ESs (see the ES column in table 1.1). In the Kluger and DeNisi meta-analysis, the average ES is 0.41. An *effect size* tells you how many standard deviations larger (or smaller) the average score for a group of students who were exposed to a given strategy (in this case, feedback) is than the average score for a group of students who were not exposed to a given strategy (in this case, no

feedback). In short, an ES tells you how powerful a strategy is; the larger the ES, the more the strategy increases student learning.

Table 1.1 Research Results for Feedback

Synthesis Study	Focus	Number of Effect Sizes (ESs)	Average ES	Percentile Gain
Bloom, 1976	General effects of feedback	8	1.47	43
Lysakowski & Walberg, 1981[a]	General effects of feedback	39	1.15	37
Lysakowski & Walberg, 1982	General effects of feedback	94	0.97	33
Yeany & Miller, 1983	Diagnostic feedback in science	49	0.55	21
Moin, 1986[b]	General effects of feedback	Not reported	0.29	11
Haller, Child, & Walberg, 1988[c]	General effects of feedback	115	0.71	26
Tenenbaum & Goldring, 1989	General effects of feedback	16	0.66	25
Bangert-Drowns, Kulik, Kulik, & Morgan, 1991	General effects of feedback	58	0.26	10
Kumar, 1991[d]	General effects of feedback	5	1.35	41
Azevedo & Bernard, 1995[e]	Immediate feedback in computer-based instruction	22	0.80	29
Kluger & DeNisi, 1996	Effects of feedback interventions	607	0.41	16
Walberg, 1999	General effects of feedback	20	0.94	33
Hattie, 1999[b]	General effects of feedback	5,755	0.95	33
Haas, 2005	General effects of feedback	19	0.55	21

[a] Reported in Fraser, Walberg, Welch, & Hattie, 1987.

[b] Reported in Hattie & Timperley, 2007.

[c] Feedback was embedded in general metacognitive strategies.

[d] The dependent variable was engagement.

[e] Reported in Hattie, 2009.

ESs are typically small numbers. However, small ESs can translate into big percentile gains. For example, the average ES of 0.41 calculated by Kluger and DeNisi (1996) translates into a 16 percentile point gain (see appendix B, page 153, for a detailed description of ESs and a chart that translates ES numbers into percentile gains). Another way of saying this is that a student at the 50th percentile in a

class where feedback was not provided (an average student in that class) would be predicted to rise to the 66th percentile if he or she were provided with feedback.

Hattie and Timperley (2007) synthesized the most current and comprehensive research in feedback and summarized findings from twelve previous meta-analyses, incorporating 196 studies and 6,972 ESs. They calculated an overall average ES of 0.79 for feedback (translating to a 29 percentile point gain). As shown by Hattie (2009), this is twice the average ES of typical educational innovations. One study by Stuart Yeh (2008) revealed that students who received feedback completed more work with greater accuracy than students who did not receive feedback. Furthermore, when feedback was withdrawn from students who were receiving it, rates of accuracy and completion dropped.

Interestingly, though the evidence for the effectiveness of feedback has been quite strong, it has also been highly variable. For example, in their analyzing of more than six hundred experimental/control studies, Kluger and DeNisi (1996) found that in 38 percent of the studies they examined, feedback had a negative effect on student achievement. This, of course, raises the critically important questions, What are the characteristics of feedback that produce positive effects on student achievement, and what are the characteristics of feedback that produce negative effects? In partial answer to this question, Kluger and DeNisi found that negative feedback has an ES of negative 0.14. This translates into a predicted decrease in student achievement of 6 percentile points. In general, negative feedback is that which does not let students know how they can get better.

Hattie and Timperley (2007) calculated small ESs for feedback containing little task-focused information (punishment = 0.20; praise = 0.14) but large ESs for feedback that focused on information (cues = 1.10, reinforcement = 0.94). They argued that feedback regarding the task, the process, and self-regulation is often effective, whereas feedback regarding the self (often delivered as praise) typically does not enhance learning and achievement. Operationally, this means that feedback to students regarding how well a task is going (task), the process they are using to complete the task (process), or how well they are managing their own behavior (self-regulation) is often effective, but feedback that simply involves statements like "You're doing a good job" has little influence on student achievement. Hattie and Timperley's ultimate conclusion was:

> Learning can be enhanced to the degree that students share the challenging goals of learning, adopt self-assessment and evaluation strategies, and develop error detection procedures and heightened self-efficacy to tackle more challenging tasks leading to mastery and understanding of lessons. (p. 103)

Assessment

In K–12 classrooms, the most common form of feedback is an assessment. While the research and theory on feedback and assessment overlap to a great extent, in this section we consider the research and theory that is specific to assessment.

Research on Assessment

The research on the effects of assessments on student learning paints a positive picture. To illustrate, table 1.2 (page 6) provides a synthesis of a number of meta-analytic studies on the effects of assessment as reported by Hattie (2009).

Table 1.2 Meta-Analytic Studies on Assessment as Reported by Hattie (2009)

Synthesis Study	Focus	Number of Effect Sizes (ESs)	Average ES	Percentile Gain
Kulik, Kulik, & Bangert-Drowns, 1984	Frequency of assessment	19	0.42	16
Fuchs & Fuchs, 1986		34	0.28	11
Bangert-Drowns, Kulik, & Kulik, 1991		35	0.23	9
Gocmen, 2003		233	0.40	16
Kim, 2005[a]		644	0.39	15
		622	0.39	15
Lee, 2006		55	0.36	14
Hausknecht, Halpert, DiPaolo, & Gerrard, 2007		107	0.26	10
Menges & Brinko, 1986	General effects of assessment	31	0.44	17
Bangert-Drowns, Kulik, Kulik, & Morgan, 1991		58	0.26	10
Travlos & Pratt, 1995		17	0.71	26
Neubert, 1998		16	0.63	24
Swanson & Lussier, 2001		170	1.12	37
Witt, Wheeless, & Allen, 2006		81	1.15	37
Fuchs & Fuchs, 1986	Providing assessment feedback to teachers	21	0.70	26
Burns & Symington, 2002		57	1.10	36

[a] Two effect sizes are listed because of the differences in variables as reported by Hattie (2009). Readers should consult that study for more details.

Notice that table 1.2 is subdivided into three categories: frequency of assessment, general effects of assessment, and providing assessment feedback to teachers. The first category speaks to how frequently assessments are given. In general, student achievement benefits when assessments are given relatively frequently as opposed to infrequently. The study by Robert Bangert-Drowns, James Kulik, and Chen-Lin Kulik (1991) depicted in table 1.3 adds some interesting details to this generalization.

Note that in table 1.3, the effect of even one assessment in a fifteen-week period of time is substantial (0.34). Also note that there is a gradual increase in the size of the effect as the number of assessments increases. This trend should not be misconstrued as indicating that the more tests a teacher gives, the more students will achieve. As we shall see in subsequent chapters, a test is only one of many ways to obtain assessment data.

Table 1.3 Achievement Gain Associated With Number of Assessments Over Fifteen Weeks

Number of Assessments	Effect Size	Percentile Point Gain
0	0	0
1	0.34	13.5
5	0.53	20
10	0.60	22.5
15	0.66	24.5
20	0.71	26
25	0.78	28.5
30	0.82	29

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

The second category in table 1.2, general effects of assessment, is the broadest and incorporates a variety of perspectives on assessment. Again, many of the specific findings in these studies manifest as the recommendations in subsequent chapters. Here it suffices to note that in the aggregate, these studies attest to the fact that properly executed assessments can be an effective tool for enhancing student learning.

The third category in table 1.2 deals with providing assessment feedback to teachers. Lynn Fuchs and Douglas Fuchs (1986) found that providing teachers with graphic representations of student progress was associated with an ES of 0.70, which translates into a 26 percentile point gain. This is quite consistent with a set of studies conducted at Marzano Research in which teachers had students chart their progress on specific learning goals (Marzano Research, 2009). The results are depicted in table 1.4.

Table 1.4 Studies on Students Tracking Their Progress

Study	Effect Size	Percentile Gain
1	2.44	49
2	3.66	49
3	1.50	43
4	-0.39	-15
5	0.75	27
6	1.00	34
7	0.07	3
8	1.68	45
9	0.07	3
10	1.20	38
11	-0.32	-13
12	0.43	17
13	0.84	30
14	0.63	24
Average	0.92	32

Table 1.4 reports the results of fourteen studies conducted by K–12 teachers on the effects of tracking student progress. The average ES of these fourteen studies was 0.92, which translates into a 32 percentile point gain. Taking these findings at face value, one would conclude that learning is enhanced when students track their own progress.

Note that in studies 4 and 11, tracking student progress had a negative effect on student achievement (indicated by the negative ES). As is the case with all assessment (and instructional) strategies, this strategy does not work equally well in all situations. Effective assessment requires ascertaining the correct way to use a strategy. In subsequent chapters, we make recommendations as to the correct way to track student progress.

Formative Assessments

Formative assessment has become very popular in the last decade. It is typically contrasted with summative assessment in that *summative assessments* are employed at the end of an instructional episode while *formative assessments* are used while instruction is occurring. As Susan Brookhart (2004, p. 45) explained, "Formative assessment means information gathered and reported for use in the development of knowledge and skills, and summative assessment means information gathered and reported for use in judging the outcome of that development."

Formative assessments became popular after Paul Black and Dylan Wiliam (1998a) summarized the findings from more than 250 studies on formative assessment. They saw ESs in those studies that ranged from 0.4 to 0.7 and drew the following conclusion:

> 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, among the largest ever reported for educational interventions. As an illustration of just how big these gains are, an effect size of 0.7, 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)

In effect, Black and Wiliam were saying that an ES of 0.70 (the largest ES reported in the studies they summarized), when sustained for an entire nation, would dramatically enhance student achievement. Indeed, consulting the table in appendix B (page 155), we see that an ES of 0.70 is associated with a 26 percentile point gain in student achievement. The reporting of these findings captured the attention of U.S. educators.

The Black and Wiliam study is sometimes referenced as a meta-analysis of some 250 studies on formative assessment. As described in appendix B of this book, a meta-analysis is a quantitative synthesis of research in a specific area. When performing a meta-analysis, a researcher attempts to compute an average ES of a particular innovation (in this case, formative assessment) by examining all of the available studies. While Black and Wiliam certainly performed a rigorous analysis of the studies they

examined, they did not conduct a traditional meta-analysis. In fact, in a section of their article titled "No Meta-Analysis," they explain, "It might seem desirable, and indeed might be anticipated as conventional, for a review of this type to attempt a meta-analysis of the quantitative studies that have been reported" (1998a, p. 52). They go on to note, however, that the 250 studies they examined were simply too different to compute an average ES.

It is important to keep two things in mind when considering the practice of formative assessment. The first is that, by definition, formative assessment is intimately tied to the formal and informal processes in classrooms. Stated differently, it would be a contradiction in terms to use "off the shelf" formative assessment designed by test makers. James Popham (2006) has harshly criticized the unquestioning use of commercially prepared formative assessments. He noted:

> As news of Black and Wiliam's conclusions gradually spread into faculty lounges, test publishers suddenly began to relabel many of their tests as "formative." This name-switching sales ploy was spurred on by the growing perception among educators that formative assessments could improve their students' test scores and help schools dodge the many accountability bullets being aimed their way. (p. 86)

To paraphrase Popham (2006), externally developed assessments simply do not meet the defining characteristics of formative assessment. Lorrie Shepard (2006) made the same point:

> The research-based concept of formative assessment, closely grounded in classroom instructional processes, has been taken over—hijacked—by commercial test publishers and is used instead to refer to formal testing systems called "benchmark" or "interim assessment systems." (as cited in Popham, 2006, p. 86)

A similar criticism might be leveled at many district-made "benchmark" assessments in that they frequently violate many of the basic assumptions underlying good formative assessment. As James McMillan (2007) explained:

> These tests, which are typically provided by the district or commercial test publishers, are administered on a regular basis to compare student achievement to "benchmarks" that indicate where student performance should be in relation to what is needed to do well on end-of-year high stakes tests. . . . Although the term *benchmark* is often used interchangeably with *formative* in the commercial testing market, there are important differences. Benchmark assessments are formal, structured tests that typically do not provide the level of detail needed for appropriate instructional correctives. (pp. 2–3)

The second thing to keep in mind is that while there is a good deal of agreement about its potential as a tool to enhance student achievement, the specifics of formative assessment are somewhat elusive. In fact, most descriptions of formative assessment are very general in nature. To illustrate, in their original study, Black and Wiliam (1998a) noted that "formative assessment does not have a tightly defined

and widely accepted meaning" (p. 7). Dylan Wiliam and Siobhan Leahy (2007) described formative assessment as follows:

> The qualifier *formative* will refer not to an assessment or even to the purpose of an assessment, but rather to the function it actually serves. An assessment is formative to the extent that information from the assessment is fed back within the system and actually used to improve the performance of the system in some way (i.e., that the assessment *forms* the direction of improvement). (p. 31)

Rick Stiggins, Judith Arter, Jan Chappuis, and Stephen Chappuis (2006) described formative assessment as assessment *for* learning rather than assessment *of* learning:

> Assessments for learning happen while learning is still underway. These are the assessments that we conduct throughout teaching and learning to diagnose student needs, plan our next steps in instruction, provide students with feedback they can use to improve the quality of their work, and help students see and feel in control of their journey to success. . . . This is not about accountability—these are assessments *of* learning. This is about getting better. (p. 31)

Susan Brookhart and Anthony Nitko (2007) explained that "formative assessment is a loop: Students and teachers focus on a learning target, evaluate current student work against the target, act to move the work closer to the target, and repeat" (p. 116).

Along with these general descriptions, specifics regarding the practice of formative assessment have been offered. Unfortunately, there is no clear pattern of agreement regarding the specifics. For example, some advocates stress that formative assessments should not be recorded, whereas others believe they should. Some assert that formative assessments should not be considered when designing grades, where others see a place for them in determining a student's true final status (see O'Connor, 2002; Welsh & D'Agostino, 2009; Marzano, 2006). To a great extent, the purpose of this book is to articulate a well-crafted set of specifics regarding the practice of formative assessment.

Learning Progressions and Clear Goals

The development of learning progressions has become a prominent focus in the field of formative assessment. Margaret Heritage (2008) explained the link between learning progressions and formative assessment as follows:

> The purpose of formative assessment is to provide feedback to teachers and students during the course of learning about the gap between students' current and desired performance so that action can be taken to close the gap. To do this effectively, teachers need to have in mind a continuum of how learning develops in any particular knowledge domain so that they are able to locate students' current learning status and decide on pedagogical action to move students' learning forward. Learning progressions that clearly articulate a progression of learning in a domain can provide the big picture of what is to be learned, support instructional planning, and act as a touchstone for formative assessment. (p. 2)

One might think that learning progressions have already been articulated within the many state and national standards documents. This is not the case. Again, Heritage noted:

> Yet despite a plethora of standards and curricula, many teachers are unclear about how learning progresses in specific domains. This is an undesirable situation for teaching and learning, and one that particularly affects teachers' ability to engage in formative assessment. (p. 2)

The reason state and national standards are not good proxies for learning progressions is that they were not designed with learning progressions in mind. To illustrate, consider the following standard for grade 3 mathematics from the state of Washington (Washington Office of Superintendent of Public Instruction, 2008):

> Students will be able to round whole numbers through 10,000 to the nearest ten, hundred, and thousand. (p. 33)

This sample provides a fairly clear target of what students should know by grade 3, but it does not provide any guidance regarding the building blocks necessary to attain that goal. In contrast, Joan Herman and Kilchan Choi (2008, p. 7) provided a detailed picture of the nature of a learning progression relative to the concept of buoyancy. They identified the following levels (from highest to lowest) of understanding regarding the concept:

- Student knows that floating depends on having less density than the medium.
- Student knows that floating depends on having a small density.
- Student knows that floating depends on having a small mass and a large volume.
- Student knows that floating depends on having a small mass, or that floating depends on having a large volume.
- Student thinks that floating depends on having a small size, heft, or amount, or that it depends on being made out of a particular material.
- Student thinks that floating depends on being flat, hollow, filled with air, or having holes.

Obviously, with a well-articulated sequence of knowledge and skills like this, it is much easier to provide students with feedback as to their current status regarding a specific learning goal and what they must do to progress.

While one might characterize the work on learning progressions as relatively new and therefore relatively untested, it is related to a well-established and heavily researched area of curriculum design—learning goals. One might think of learning progressions as a series of related learning goals that culminate in the attainment of a more complex learning goal. Learning progressions can also be used to track student progress. The research on learning goals is quite extensive. Some of the more prominent studies are reported in table 1.5 (page 12).