

CHAPTER	Formative Assessment
1	Why, What, and Whether
<p>[Chapter Preview]</p> <ul style="list-style-type: none"> • Why formative assessment is garnering so much attention • What formative assessment is and what it isn't • The case for using formative assessment 	

FORMATIVE ASSESSMENT WORKS! THIS PHRASE, OR SOME paraphrased version of it, is voiced with increasing frequency in many parts of the world. But *why* are more and more of today's educators touting the instructional virtues of formative assessment?

Most observers credit British researchers Paul Black and Dylan Wiliam with kicking off today's worldwide interest in formative assessment. In 1998, Black and Wiliam published two important works: an article in the journal *Phi Delta Kappan*, and an extensive review of empirical research studies focused on classroom assessment. In their *Kappan* article, Black and Wiliam

(1998b) argue that formative assessment, properly employed in the classroom, will help students learn what is being taught to a substantially better degree. They support this argument with evidence from their research review (1998a), a meta-analysis in which they conclude that student gains in learning triggered by formative assessment are "amongst the largest ever reported for educational interventions" (p. 61).

Once educators realized there was ample evidence that formative assessment really was an effective way to improve student learning, it wasn't long before they began investigating the implications and asking the all too obvious follow-up question: If formative assessment could improve student learning in the classroom, couldn't it *also* improve student test scores on external accountability tests? Considering that so many educators are now figuratively drowning in an ocean of accountability, it's not surprising to see formative assessment cast in the role of life preserver. If it is true that drowning people will grasp at straws in an effort to stay afloat, it is surely as true that they will grasp even more eagerly at "research-proven" straws.

This is not to suggest that all advocates of formative assessment see it primarily as a strategy to fend off pervasive accountability pressure. Many believe that formative assessment will simply help educators do a better job of teaching. These educators might even point out that Black and Wiliam's research synthesis focuses primarily on the classroom dividends of formative assessment and devotes little attention to its potential role in raising external test scores. And many more, if not most, formative assessment proponents choose neither camp exclusively. They believe it can have a positive effect on both students' in-class learning and students' subsequent performance on accountability tests.

All right, that's the *why* underlying today's ever-expanding interest in formative assessment. Now it's time to take a close look at the *what*: What formative assessment actually is and what it isn't.

What Is Formative Assessment?

Because there has been so much attention lavished on formative assessment lately, most of today's teachers and administrators have at least a rough idea of what it is. If you asked them to explain it, they might tell you it involves testing students in the midst of an ongoing instructional sequence and then using the test results to improve instruction. By and large, this explanation is correct.

But a "by and large correct" explanation just isn't good enough when it comes to formative assessment. As you'll see in later pages, formative assessment is a potentially transformative instructional tool that, if clearly understood and adroitly employed, can benefit both educators and their students. Mushy, "by and large correct" understandings of formative assessment will rarely allow the fullness of this assessment-based process to flower. That's why I'm now asking you to join me as I dig for a while into the innards of formative assessment.

Historical and Etymological Underpinnings

There is no single officially sanctified and universally accepted definition of formative assessment. Educators have drawn our use of the term *formative* from Michael Scriven's (1967) groundbreaking essay about educational evaluation, in which he contrasts summative evaluation with formative evaluation. According to Scriven, if the quality of an early-version educational program is evaluated while the program is still malleable—capable of being improved because of an evaluation's results—this constitutes *formative* evaluation. In contrast, when a mature, final-version educational program is evaluated in order to make a decision about its continuation or termination, this constitutes *summative* evaluation.

Scriven's insightful split of two program-evaluation roles was widely and rapidly accepted by educational evaluators. Although

a handful of early writers, notably Bloom (1969), attempted to transplant the formative/summative *evaluation* distinction directly onto *assessment*, few educators were interested in investigating this idea further because it seemed to possess few practical implications for the day-to-day world of schooling.

In fact, it was only during the past decade or two that educators began to discuss whether a distinction between the formative and summative roles of assessment could benefit teachers' instructional decisions. When meaningful interest in this assessment difference finally blossomed, the essence of Scriven's original distinction between the two roles of educational evaluation was retained. That is, we continue to see formative assessment as a way to improve the caliber of still-underway instructional activities and summative assessment as a way to determine the effectiveness of already-completed instructional activities.

With these origins understood, it's time to press on toward the definition of formative assessment that we'll use in this book.

A Carefully Considered Definition

The Council of Chief State School Officers (CCSSO) is a major U.S. organization composed of individuals who head the educational system of each state. These officials, typically referred to as "state superintendents" or "state commissioners," are either elected by popular vote or appointed by governors or state boards of education. The "chiefs" have enormous educational influence in their respective states, and in 2006, when CCSSO launched a major initiative focused on a more balanced use of educational assessment and a heightened emphasis on formative assessment, it was a significant policy shift likely to have long-lasting influence on practices in U.S. public schools.

A central activity in the CCSSO assessment initiative was the creation of a new consortium focused specifically on formative assessment. A CCSSO consortium is composed of key department of education personnel from those states that wish to participate. Each of these groups is referred to as a State Collaborative on Assessment and Student Standards (SCASS), and a new SCASS

dealing exclusively with formative assessment, known as Formative Assessment for Students and Teachers—or FAST SCASS, if you're in a hurry—was formed in mid-2006.

FAST SCASS held its inaugural meeting in Austin, Texas, in October 2006, with about 60 participants representing roughly 25 states. The chief mission of the four-day meeting was to reach consensus on a definition of formative assessment, with the ultimate aim of shaping the way U.S. educators understand this practice. Prominent among the concerns of the FAST SCASS members was that the definition reflect the latest research findings regarding assessment practices found to improve the quality of students' learning. Remember this point. It is important.

After considering a variety of earlier definitions, and after numerous foreseeable rounds of participants' wordsmithing, the FAST SCASS group adopted the following definition:

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.

Let's look at the key features of the FAST SCASS definition:

- Formative assessment is a *process*, not any particular test.
- It is used not just by teachers but by *both teachers and students*.
- Formative assessment takes place *during instruction*.
- It provides *assessment-based feedback* to teachers and students.
- The function of this feedback is to help teachers and students make *adjustments* that will improve students' achievement of intended curricular aims.

I took part in those October 2006 deliberations, and I was relatively pleased with the group's final definition and delighted that it was adopted without dissent. I sincerely hope the FAST SCASS definition will be widely used. But we're not going to use that definition in *this* book. Frankly, I believe that in our effort to scrupulously reflect the research findings available to us and

satisfy the definitional preferences of all the meeting participants, our FAST SCASS group produced a definition that is verbally cumbersome. Although I have no quarrel with what the definition says, it just doesn't say it succinctly enough.

A More Succinct and Useful Definition

What educators really need is a definition of formative assessment that helps them instantly recognize what's most important about this approach. Thus, with apologies to my FAST SCASS colleagues, I present *my* definition of formative assessment, the one we'll be using in this book:

Formative assessment is a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics.

As is typical of progeny, the above conceptualization of formative assessment shares much with the FAST SCASS definition from whence it came:

- Again, formative assessment is not a test but a process—a *planned* process involving a number of different activities.
- One of those activities is the *use of assessments*, both formal and informal, to elicit *evidence regarding students' status*: the degree to which a particular student has mastered a particular skill or body of knowledge.
- Based on this evidence, *teachers adjust* their ongoing instructional activities or *students adjust* the procedures they're currently using to try to learn whatever they're trying to learn.

Phrasing it more tersely still:

Formative assessment is a planned process in which teachers or students use assessment-based evidence to adjust what they're currently doing.

Now, let's take a slightly deeper look at each of the key attributes of this conception of formative assessment.

A planned process. Formative assessment involves a series of carefully considered, distinguishable acts on the part of teachers or students or both. Some of those acts involve educational assessments, but the assessments play a role in the process—they are not the process itself. An educator who refers to “a formative test” has not quite grasped the concept, because there’s no such thing. There are tests that can be used as part of the multistep, formative assessment process, but each of those tests is only a part of the process.

If you accept the distinction between the formative and summative use of test results, then you will recognize that students’ results on a particular test might be used for either a summative or a formative purpose. It is not the nature of the test that earns the label formative or summative but the use to which that test’s results will be put. If the purpose of Test *X* is to provide teachers and students with the evidence they need to make any warranted adjustments, then Test *X* is playing a role in the formative assessment process.

Assessment-elicited evidence. The adjustment decisions teachers and students make during the formative assessment process must be based not on whim but on evidence of the students’ current level of mastery with respect to certain skills or bodies of knowledge. Accordingly, the assessment procedures designed to generate this evidence are an indispensable element of the process. Although teachers may certainly employ paper-and-pencil tests for this purpose, they can also obtain the evidence they need via a wide variety of less traditional and much less formal assessment plays, many of which I will describe later in this book.

Teachers’ instructional adjustments. Formative assessment’s *raison d’être* is to improve students’ learning. One of the most obvious ways to do this is for teachers to improve how they’re teaching. Accordingly, one component of the formative assessment process is for teachers to adjust their ongoing instructional activities. Relying on assessment-based evidence of students’ current status, such as test results showing that students are weak in their mastery of a particular cognitive skill, a

teacher might decide to provide additional or different instruction related to this skill.

It's worth stressing that because the formative assessment process deals with *ongoing* instruction, any teacher-made modifications in instructional activities must focus on students' mastery of the curricular aims currently being pursued. It's not a matter of looking at test data and deciding to try a new approach next time; it's a matter of doing something different (or differently) now.

Students' learning tactic adjustments. Within the formative assessment process, students also take a look at assessment evidence and, if need be, make changes in how they're trying to learn. Consider, for example, a high school student who is working toward becoming a better public speaker by practicing a particular speech many times before a mirror. That repeated, solo-mirror practice is the student's learning tactic; based on assessment evidence, this tactic may or may not need adjustment.

I want to wrap up this definition overview by circling back to something very important. One of the most difficult tasks for educators who accept this conception of formative assessment is to grasp the overarching idea that it is a process rather than a test. You may have noted that in many of the preceding paragraphs I have referred to the "formative assessment process." That triple-word phrasing was a ploy to drive the point home. From here on, whenever you see the phrase *formative assessment*, I trust you'll know that it refers to a multistep process and not to a particular assessment tool.

Why We Need Definitional Clarity

Why have I been making such a fuss about a definition of formative assessment? Is it just so you won't talk past your colleagues when you discuss formative assessment with them? No, there's a more important reason, and it stems from what we do know about certain applications of educational assessment and what we don't know about others.

There are certain educational assessment practices that empirical evidence has shown to have a positive influence on student learning. There are other educational assessment practices that research has not (or has not yet) shown to have this effect. Educators need to be able to distinguish between the former and the latter. Why? So that they will be forearmed against commercial test-development companies that are eager to hitch a profit-making ride on the enthusiasm for formative assessment and, thus, will label as "formative assessment" practices that *are not actually consonant with the body of research that validates formative assessment* and, therefore, may not deliver the instructional benefits accompanying appropriately implemented formative assessment.

To illustrate how easily educators might be led astray, we need only consider the number of testing companies that are distributing as "formative assessments" products typically referred to as *interim* or *benchmark* tests. An interim or benchmark test is one that is administered periodically (perhaps once every two or three months) to measure students' status with respect to mastery of important curricular outcomes. An example of such outcomes might be 15 state-identified mathematical skills assessed each May by a statewide accountability test. A commercial test vendor might develop three different forms of an "interim test" to assess these skills. With each form containing 75 items (5 items per state-identified math skill), the test is designed to provide an indication of a student's mastery status with respect to all 15 skills. The test vendor's marketing materials might suggest that teachers administer the forms at various intervals during the school year: perhaps Form 1 in the fall, Form 2 just after winter break, and Form 3 in the spring, one month before the date of the statewide accountability test. The test vendor would send teachers the results of the interim tests, thus providing 3 snapshots of their students' current status in regard to the 15 state-sanctioned mathematics skills that will be measured "for keeps" in May.

This sort of periodic assessment may be a good thing to do. It may help teachers do a better instructional job with their

students by, for example, helping teachers accurately predict which of their students are likely to have difficulty with the May math accountability test and go on to single-out those students for particular attention. But there is currently no research evidence supporting the hypothesis that this kind of periodic assessment is educationally beneficial. Thus, describing interim or benchmark tests as “formative” in the sense that they are in accord with research evidence such as that synthesized by Black and William (1998a) is a fundamental misrepresentation. A test vendor touting them as such is being disingenuous or, as we used to say, is *lying*.

It is not only profit-preoccupied commercial vendors who mislabel their assessments as formative. In more than a few states and school districts, educators have created periodically administered assessments that they characterize as “formative” and go on to disseminate with great enthusiasm. One suspects that these educators, in a well-meant effort to assist classroom teachers, have tried to develop periodic assessments that will be instructionally beneficial. And perhaps those assessments *will* help teachers. But again, *there is no evidence that these district-developed or state-developed assessments boost student achievement.* Characterizing these periodic tests as “formative” is simply inaccurate. I’m not suggesting that district- or state-level educators are affixing the label to their tests for nefarious purposes; they’re simply being imprecise about their conceptualization of what formative assessment truly is.

And *that’s* why I have lavished so much attention on a defensible definition of formative assessment—one reflecting the key attributes of formative assessment that empirical research evidence has shown to improve student learning. The future may yet yield evidence that alternative assessment approaches are similarly effective. But for the time being, the closer your “formative assessment” practices and procedures match my definition, the more likely it is they will deliver the instructional benefits you desire. It’s the difference between *very probably* getting a return on your investment and *maybe just possibly* getting one. If I were a teacher, I know the choice I would make.

Is Formative Assessment Always Classroom Assessment?

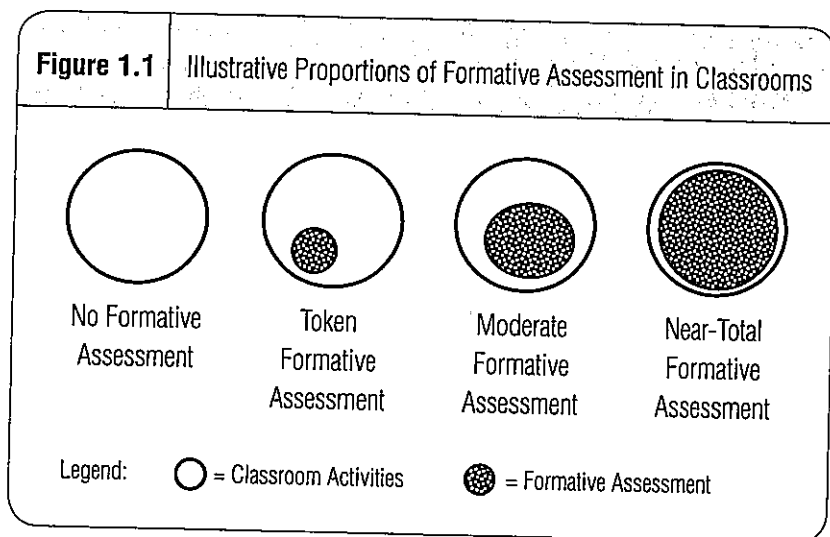
We've just about concluded our definitional digging, but there's one more important issue that still needs to be addressed. Put simply, as formative assessment is defined in this book, does it necessarily need to be *classroom* assessment? Is it possible, say, for *administrators* in states, districts, or schools to install and operate the sort of assessment-rooted process envisioned in the definition we've been dissecting, or must that kind of process spring exclusively from the activities of a particular teacher in a particular classroom?

Well, we know for sure that all *classroom* assessment need not be *formative* assessment. That's because teachers could administer frequent tests with the exclusive purpose of assigning grades and not try at all to use data from such tests to improve instruction. But the question under consideration is a little different: Must formative assessment, by its very nature, take place only at the classroom level?

You can arrive at your own judgment about this issue, but let me get my own answer on the table so you'll at least recognize my preference. Although it might be technically possible to encounter versions of formative assessment that have been externally imposed on classrooms rather than emerging from those classrooms themselves, this would be really rare. It should be. Formative assessment that really pays off for students will, I believe, be *classroom* formative assessment. *agreed*

Remember, for formative assessment (as we define it) to exist at all, it must lead to instructional adjustment decisions by teachers or learning tactic adjustment decisions by students, and these adjustments will affect activities or efforts already in progress. The decisions to adjust or not to adjust, and the decisions about the nature of any adjustments (the *what* and the *how*) need to be made on the spot or almost on the spot—when there's still instructional and learning time available. It's because of this limited time frame that when I think of formative assessment, I always precede it with an invisible modifier. Formative assessment, to me, is *classroom* formative assessment.

Figure 1.1 shows illustrative proportions of formative assessment that might be seen in many classrooms. The classrooms at the far left are those in which all assessment is grading focused rather than improvement focused. These are classrooms where there is plenty of classroom assessment but zero formative assessment. The classrooms at the far right are those in which almost all classroom assessment, both formal and informal, has a formative improvement mission. The middle two circles represent token or moderate uses of formative assessments as part of a teacher's classroom assessment activities. The two circles at the extreme right are not completely coterminous (one of my favorite words, meaning "having identical borders") because, in most settings, there will be some required minimum of classroom assessment activities that must be devoted to the determination of students' grades.



Whether to Use Formative Assessment

We've looked briefly at the *why* (why there's so much fuss about formative assessment these days), and we've also taken a pretty thorough look at the *what* (what the meaning of formative assessment is). Now we need to consider the *whether*, namely, whether

teachers ought to employ formative assessment in their own classrooms.

For educators, now is when this book gets personal. We are no longer talking in the abstract about definitions or the reasons other folks have been focusing on formative assessment. If you're a teacher, the time has come to start wrestling with the vital question of whether formative assessment is something you ought to do, for your own sake and for the sake of your students. If you're an administrator, it's time to begin tussling with the issue of whether you will advocate formative assessment among those teachers with whom you work.

There are essentially two ways educators arrive at decisions about whether to alter the way they approach instruction: (1) by accepting a data-free argument that a new approach is worthwhile and (2) by accepting the research-based evidence of a new approach's efficacy. Let's consider each.

A Data-Free Argument

Many educators choose to implement or advocate new instructional approaches simply because they've been persuaded by the well-articulated positions of advocates. Given that there is a ton of artistry in teaching, trying out a new method because someone speaks highly of it is a perfectly reasonable way for teachers to tinker with their instructional approaches. You show me a teacher who hasn't tried out an instructional tactic based on someone else's recommendation, and I'll show you a teacher who is rooted in rigidity. If someone puts forward what seems to be a smart way of improving instruction, it often makes sense to take that idea out for a spin in your own classes.

If you're looking for an advocate of formative assessment, clearly you've come to the right place. Given the topic of this book, it shouldn't surprise you that I am one. I believe that *all* teachers should use this process because, when it's properly implemented, formative assessment will improve how well students learn. Moreover, because evidence of improved student learning is likely to be regarded as evidence of a teacher's own instructional effectiveness, formative assessment has the happy

effect of helping successful teachers be accurately regarded as successful.

Here's the thrust of my evidence-free argument in favor of formative assessment. First off, teachers function in order to help kids learn, but few teachers will always create perfect instructional designs the first time around. Most discover that even their best-laid instructional plans usually need to be massaged, sometimes meaningfully, in order to work optimally. And if this is so, then why not make such massaging as potent as possible by finding out about students' current status during an ongoing instructional process? It makes eminent sense for teachers to alter their instructional procedures only after getting an accurate fix on their students' current achievement status. It makes eminent sense for teachers to provide students with the assessment-based evidence those students will need when deciding how well their learning tactics are working. In short, it makes eminent sense to rely on assessment-based evidence about students' skills and knowledge before figuring out how to help those students learn better.

Instruction should not be a Ouija-boardlike game in which teachers guess about what to do next. Educating kids is far too important for that sort of approach. Rather, instructing students should be a carefully conceived enterprise in which decisions about what to do next are predicated on the best available information. And the best available information about what to do next almost always flows from a determination about what students currently know and can do. Formative assessment supplies the evidence a teacher needs in order to make any necessary instructional adjustments. Formative assessment supplies the evidence students need in order to make any necessary adjustments in how they are trying to learn something. Formative assessment, then, can help both teachers teach better and learners learn better.

So, even if there were not a shred of empirical evidence to support the worth of formative assessment in the classroom, I would still be advocating its use because it makes so darn much sense! Thus, the first reason I hope you will adopt formative

assessment in your own classroom or advocate its adoption in other classrooms is because formative assessment constitutes the key cornerstone of clearheaded instructional thinking. Formative assessment represents evidence-based instructional decision making. If you want to become more instructionally effective, and if you want your students to achieve more, then formative assessment should be for you.

Adjusting
instructional
based
on
student
learning

An Evidence-Based Case

Of course, teachers' other approach to coming up with improved ways to teach their students is to adopt *research-proven* instructional procedures. Because the advocacy of formative assessment you'll find in this book rests chiefly on this second approach, it is important for you to understand both the strengths and shortcomings of "research-proven" instructional strategies.

Having research evidence on your side is wonderful; in the midst of a disagreement, all you need to do is choose the right moment to roll out your research evidence and, if the evidence is sufficiently compelling, you'll be running up the victory flag in a matter of minutes. Unfortunately, when professionals in the field of education assert that a procedure is "research proven," we typically mean something different from what professionals in many other fields mean when they invoke this description. Let's consider why.

Proof and probability. Educational research revolves around human beings, most of whom are *little* human beings. And human beings, regardless of their size, are complicated creatures. As a consequence, even if we rely on results of research studies simply reeking of methodological rigor, *rarely* can we say that "if a teacher takes Action X, then the teacher's students will unquestionably end up achieving Outcome Z." Education because it involves human beings, is just too complex. Every instructional situation involves a *particular* teacher whose unique background makes that teacher profoundly different from a teacher born on the very same day in the same hospital and raised just down the street by a pair of equally loving parents. Teachers themselves

Be wary
of generic
research

are a widely divergent variable, whether it is because of DNA differences or because of their idiosyncratic experiences. *Particular* teachers also function in settings where there are *particular* principals, *particular* colleagues, and, most important of all, *particular* kids. Teaching, in short, is such a particularistic endeavor that, with remarkably few exceptions, the best we can ever derive from educational research investigations is *probability-based instructional guidance*.

What this means is that even if several education research studies arrive at essentially identical findings, the most apt conclusion a teacher can draw from such findings will be something along these lines: "If I engage in this research-supported instructional action, it is *likely* that my students will learn what I want them to learn." Even the most compelling of "research-proven" conclusions typically can't go much beyond something such as, "If you do this thing, it is *probable* that you'll get the result you want."

Yes, the patent particularism of real-world instruction always requires us to regard "research proven" as providing probabilistic guidance rather than as supplying us with definitive truths. However, let's not knock probabilistic guidance. If professionals—be they surgeons, attorneys, or teachers—can increase their odds of achieving a desired result even slightly, those professionals should always opt for the improved-probability course of action. Just ask any professional gambler about the dividends of having the odds "ever so slightly" on your side.

Examining the evidence. Clearly, we need to look at the evidence supporting the dividends of formative assessment. Just how compelling is that evidence? If you are a teacher, is it sufficiently compelling for you to personally install formative assessment as an integral part of your own approach to instruction? If you are an administrator, is it sufficiently compelling for you to personally urge your teacher-colleagues to incorporate formative assessment as an integral part of their own classroom instruction?

As indicated earlier, much of the current interest in formative assessment, especially in the United States and Canada, was

spurred by the work of two British researchers, Paul Black and Dylan Wiliam. Their *Kappan* essay (1998b) draws heavily on their own extensive *meta-analysis* of classroom formative assessment published earlier the same year in a special issue of the journal *Assessment in Education* (1998a). In that synthesis of research studies, Black and Wiliam conclude that "the research reported here shows conclusively that formative assessment does improve learning" (1998a, p. 49).

These days, one often encounters advocates of certain instructional interventions who support their advocacy by invoking meta-analyses as though a meta-analysis report constitutes some sort of sanctified scripture. Meta-analyses are not holy writ. They differ dramatically in quality simply because analysis is an endeavor requiring judgment. The work of a meta-analyst requires many along-the-way judgments. If all of those judgments are defensible, then we can place reasonable confidence in the conclusions of the meta-analysis. If some of those judgments are indefensible, then confidence quickly crumbles.

Thus, you should not routinely defer to the array of numerical effect sizes that often seems to conclude a meta-analysis in a quantitative (hence "credible") manner. Those numbers bubble forth from a judgmental stew of the meta-analyst's own making. *The confidence you should place in the conclusions of a meta-analysis should be completely dependent on the caliber of the meta-analysis itself.* And this brings us directly to the specific meta-analysis that has galvanized educators' interest in formative assessment.

The Black and Wiliam meta-analysis. "Assessment and Classroom Learning" (Black & Wiliam, 1998a) was published in the journal *Assessment in Education* along with comments on the analysis by leading educational experts from several nations. Let's consider, then, the essential features of the Black and Wiliam meta-analysis.

They based it on nine years' worth of research reports and used two earlier extensive reviews (Crooks, 1988; Natriello, 1987) as starting points. Focusing on formative classroom assessment defined essentially in the same manner as it has been defined in

META-ANALYSIS: MAKING SENSE OF THE MYRIAD

Serious educational research has been under way in the United States for more than a century, and the volume of that research seems to be expanding at an almost exponential rate. Consider that in 1958, when I made my first trip to the annual meeting of the American Educational Research Association (AERA), the total attendance was about 300. Approximately 200 research papers were submitted for presentation, and 100 papers were presented. Fifty years later, in 2007, almost 16,000 researchers attended the AERA Annual Meeting, more than 10,000 research papers were submitted, and about 3,000 papers were presented. More research is a good thing, of course, but beyond a certain point, the volume of it can be overwhelming. Back in 1958, I might have been able to read *all* the 100 research reports presented at that year's AERA meeting over the span of a weekend. There is no way I or anyone could have read every report presented at the 2007 meeting over a weekend—not even a three-day weekend! If a relevant research report falls outside the range of an interested educator's reading-volume capacity, does it still make a sound?

It does now. Insightful researchers, recognizing that there are too few three-day weekends and far too many research reports to make sense of, provided a solution: *meta-analysis*. The leader of this effort was Gene Glass of Arizona State University, who not only pioneered the method of meta-analysis but provided readily understood quantitative tactics for coalescing the results of dissimilar investigations. The Greek prefix, *meta*, means “beyond,” signifying that meta-analysts aim to carry out analyses *beyond* the individual study results under review. Skilled meta-analysts can synthesize the results from a variety of studies so that, despite study-to-study differences, we can still derive generalizable conclusions from a welter of seemingly divergent investigations. With good reason, meta-analysis has become a powerful tool to help educators comprehend the huge array of research now available.

this book, Black and Wiliam were able to identify 681 publications that appeared potentially relevant to their review. From those 681 contenders, they read approximately 250 reports in full and used them in the final meta-analysis. You can find a detailed description of the procedures employed in the identification and scrutiny of possible publications in the report of the Black and Wiliam meta-analysis (1998a). The studies they considered were diverse in nature: carried out by researchers in several nations, involving students ranging from 5-year-olds to undergraduates, and focused on a variety of subject fields. Although classroom formative assessment was an explicit focus of many of the studies included in the meta-analysis, in some instances the use of formative assessment was only ancillary to a particular study's chief investigative focus.

Based on their meta-analysis, Black and Wiliam report typical effect sizes of between 0.4 and 0.7 in favor of students taught in classrooms where formative assessment was employed. Such effect sizes, the two meta-analysts observe, "are larger than most of those found for educational interventions" (1998b, p. 141). To help their readers understand the meaning of such effect sizes, Black and Wiliam point out that an effect-size gain of 0.7 in a recent 41-nation international study of mathematics would have raised the rank of a middle-of-the-pack country to one of the top-5 nations. A particularly important finding in the bulk of the meta-analyzed studies is that "improved formative assessment helps low achievers more than other students—and so reduces the range of achievement while raising achievement overall" (Black & Wiliam, 1998b, p. 141).

In their lengthy report of this meta-analysis, Black and Wiliam are also particularly attentive to the considerable variety of studies with which they were working. Nonetheless, their overall conclusion is strikingly supportive of the contribution that formative assessment can make to the instructional process:

The consistent feature across the variety of these examples is that they all show that attention to formative assessment can lead to significant learning gains. Although there is no guarantee that it will do

Huge effect size for meta-analysis

critical for low level kids

so irrespective of the particular approach adopted, we have not come across any report of negative effects following an enhancement of formative practice. (1998a, p. 17)

The absence of reported negative effects linked to formative assessment is worth noting, for in many meta-analyses of instructional interventions, the meta-analysts discover a number of results representing either "no significant differences" or differences actually favoring the "untreated" students.

Black and Wiliam emphasize that they have not offered "optimal modes" to follow in the installation of classroom formative assessment. Happily, as these two researchers point out,

the range of conditions and contexts under which studies have shown that gains can be achieved must indicate that the principles that underlie achievement of substantial improvements in learning are robust. Significant gains can be achieved by many different routes, and initiatives here are not likely to fail through neglect of delicate and subtle features. (1998a, p. 61)

This is reassuring. However, Black and Wiliam wrap up their extensive review with a powerful caution to practitioners, namely, that "the changes in classroom practice are central rather than marginal, and have to be incorporated by each teacher into his or her practice in his or her own way" (1998a, p. 62).

When considering the persuasiveness of a meta-analysis, the confidence we should ascribe to it must be derived from the quality with which that analysis was carried out. Most reviewers of the Black and Wiliam meta-analysis conclude that both the caliber of methodological rigor employed and the high quality of the judgments used throughout render this important review worthy of our confidence. The overwhelmingly positive results found in almost a decade's worth of diverse empirical investigations lends powerful support to the notion that "formative assessment works!" Accordingly, if you're a teacher, you ought to make it work for your students' well-being. And if you're an administrator who supports teachers' efforts, you ought to advocate formative assessment's adoption.

The Effectiveness Issue

Teachers, especially seasoned ones, know from bitter experience that there's a decisive difference between (1) an instructional procedure and (2) an *effective* instructional procedure. So far in this chapter, we've considered *why* there's so much interest in formative assessment, *what* formative assessment is, and *whether* teachers ought to employ formative assessment in their own classrooms.

You'll notice, however, that we have not focused on distinguishing between formative assessment that's flawed and formative assessment that's fabulous. Please be patient. Later on, we'll get to the important issue of how to maximize formative assessment's quality, but to tackle the quality issue sensibly, it will be necessary to do so in a particular way. If you had an opportunity to read the preface to this book (and if you actually did read it), you saw that we'll be looking at four distinctive levels of formative assessment. Well, it simply makes more sense to deal with *per-level* quality factors rather than pretending one set of quality considerations will apply to all four strikingly different levels.

If at this point you have acquired a reasonably solid understanding of what formative assessment is, I promise, by the end of the book you'll have an equally solid understanding of how to estimate the likelihood that a given implementation of formative assessment will work well.

SUPER-SUCCINCT SUMMARY

- Formative assessment is a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics.
- Because formative assessment has been shown to improve students' in-class learning, many educators have adopted it in the hope that it will also raise their students' performances on accountability tests.
- The expanded use of formative assessment is supported not only by instructional logic but also by the conclusions of a well-conceived and skillfully implemented meta-analysis by Paul Black and Dylan Wiliam.