



Rick Stiggins

Dr. Rick Stiggins is founder and executive director of the ETS Assessment Training Institute in Portland, Oregon. Since 1992, the Institute's professional learning programs have helped teachers and school leaders understand how to use the assessment process and its results to benefit (not merely monitor) student learning. The hallmark of these programs has been their focus on the use of student-involved assessment to maximize their confidence, motivation, and learning. The flagship program, Classroom Assessment FDR Student Learning, provides a multi-media, collaborative, and hands-on way for teachers to become competent, confident classroom assessors. Dr. Stiggins has served on the faculties of Michigan State University, the University of Minnesota, and Lewis and Clark College, Portland, as well as on the research and development staffs of ACT and the Northwest Regional Educational Laboratory.

Dr. Stiggins believes that while classroom assessment certainly can and should serve accountability purposes by providing evidence for report-card grading, we can and should take advantage of the process to help us accomplish far more than that: Classroom assessment can serve to promote student success.

In this chapter, Dr. Stiggins argues that truly productive assessment cannot merely be about qualities of instruments and their resulting scores. Rather, it must also be about the impact of the score *on the learner*. He first describes what it means to assess accurately using the four keys to assessment quality: clear purpose, clear targets, accurate assessment, and effective communication. He then describes the effective use of the assessment process and its results to help students advance their learning with enthusiasm so they feel in control of their learning as they attain new levels of proficiency: assessment *for learning*.

For more information about Dr. Rick Stiggins, visit www.ets.org/ati.

Chapter 3

Assessment for Learning: An Essential Foundation of Productive Instruction

Rick Stiggins

Years ago during my graduate studies in educational measurement, I learned that my job was to assure the dependability (validity and reliability) of the scores resulting from the assessment process. My studies of classical and modern psychometrics focused my attention on the attributes of the resulting scores. I was taught how to design and develop instruments that consistently produced scores that accurately reflected the learning target.

In the past decade, I have experienced a profound transformation in my understanding of the keys to productive assessment. I continue to honor my professional heritage: I believe more strongly than ever that assessment results must be accurate in all contexts. Inaccurate data leads to counterproductive instructional decisions, and thus it is harmful to students. However, I have come to realize that truly productive assessment cannot merely be about qualities of instruments and the attributes of their resulting scores. Rather, it must also be about the impact of that score *on the learner*. In other words, I have come to see that students also read, interpret, and, most importantly,

act on the data we generate with our assessments about their achievement. They make crucial decisions based on those data. In fact, I have come to understand that the decisions they make as users of assessment results exert far greater influence on their success as learners than do the decisions made by the adults—the parents, teachers, administrators, and policymakers—around them.

This transformation in my thinking about the dynamics of assessment came neither from my work in measurement research nor from my knowledge of professional literature, but rather from my experiences as a father. As our daughter grew, I had a chance to see firsthand the emotional and learning benefits that came from using effective assessment to support her learning. Unfortunately, I also had a front-row seat from which to observe the immense and long-lasting harm that can be done when assessment is clumsy, inept, or used in counterproductive ways.

For these reasons, I have built this chapter around two basic assessment lessons: First, we must assess accurately. I will describe exactly what this means. Second, we must use the assessment process and its results productively: to keep students believing in themselves as capable learners who make sound decisions that will lead them to greater levels of achievement. I will provide specific guidelines for how to do this as well.

In professional learning communities, faculties team up to create assessments and to gather and interpret evidence of student learning in order to make program improvement decisions that will enhance that learning and make schools more effective (DuFour, 2005). Obviously, we all believe decisions that are based on sound evidence will support student learning, while decisions based on inaccurate evidence or no evidence at all can harm that learning. Therefore, sound assessment practices represent a key foundation of effective schools. In this chapter, we will explore the keys to quality common assessments within professional learning communities.

But professional learning communities cannot be merely about teachers making decisions based on common assessment data. Students must be partners in the community, too. Assessment must encourage and support them in their pursuit of excellence. Historically, we have done this by using the threat of a pending assessment in an attempt to drive students to study. This worked for some, but not for others. More recently, however, we have come to understand that students who believe they are capable learners experience greater success in school than students who have lost faith in themselves. In other words, we can replace the emotional dynamics of fear and vulnerability with those of academic self-efficacy and eagerness to learn as the driving emotions for academic success. But we can do this only if we expand our sense of professional learning communities to include our students as instructional decision-makers, and if we provide faculties with the opportunity to learn to use assessment in ways that involve students. I will show you how to tap these emotional dynamics of assessment to help all students succeed.

The Keys to Assessment Quality

We have established that accuracy is necessary, but not sufficient, for productive assessment. There are three key factors to consider in assuring assessment accuracy. To create quality assessments, educators must do the following:

- Start with a clear *purpose* for assessment—a sense of why we are assessing.
- Include a clear achievement *target*—a vision of what we need to assess.
- Design an assessment that accurately *reflects* the target and *satisfies* the purpose.
- *Communicate* results effectively to the intended user(s).

If we fail to attend to any of these features, we risk the quality of the data we gather, and therefore we risk harm to learners. Let us consider each in greater detail.

Key 1: Clear Purpose

If we assess, in part, to gather evidence to inform instructional decisions, then in any specific assessment context, the assessor needs to start the process with answers to these questions:

- What are the instructional decisions we hope to make?
- Who is making them?
- What information will be helpful?

The answers to these questions will vary profoundly across contexts, from the classroom to the institutional and policy-making levels.

As instruction unfolds daily in the classroom, the key question to be answered with data is this: “What comes next in the learning?” The decision-makers are students and teachers. The information these decision-makers need is continuous evidence of how each individual student is doing on her or his learning journey towards each standard. Both student and teacher must know where the learner is now, how that compares to ultimate learning success, and how to close the gap between the two. Students must not be wondering *if* they will succeed—only *when* they will succeed. Obviously, this is the domain of day-to-day classroom assessment.

At the program level of decision-making, on the other hand, we need to know which achievement standards students are and are not mastering—which of our programs are working and which need adjustment. Those who will plan and implement those adjustments are teachers (often working in teams), principals, and curriculum personnel. The evidence they need must reveal who is and is not meeting standards—in terms that are comparable across classrooms

so data can be aggregated at the program level to affirm effectiveness or reveal potential changes. This, then, is the purview of the interim, benchmark, short-cycle assessment—the common assessment in the professional learning community context.

Note that both classroom and program levels of decision-making are important, but they are different. The former focuses on each student’s journey to each standard, while the latter centers on group mastery of the standards themselves. Classroom assessment must be continuous, while program assessment can be periodic. The classroom level informs students and teachers as they make immediate instructional decisions, while the program level informs teacher teams and school leaders as they make intermittent program adjustments. Both must be done well for students to prosper.

Finally, there is the institutional or policy level of assessment use which, these days, requires that school and community leaders determine if enough students are meeting standards. They need evidence of student mastery of state standards that is comparable across classrooms and schools once each year to be accountable to the community and to allocate resources for program improvement. This, then, is the purview of our annual large-scale accountability tests.

When it comes to assessment results, the users and their information needs are different at each level. This is why assessors in any context must start with a clear sense of purpose: They must know who their assessment should help, and how it should help, in order to design it properly.

Key 2: Clear Targets

The second key to quality assessment is the clear, complete, and appropriate articulation of the achievement target(s) to be mastered. We cannot dependably assess that which we have not defined. These days, we start target definitions with state standards or local adaptations of those standards. When these standards are organized

in a logical manner to unfold properly within and across grade levels, they can be the focus of both interim/benchmark/common assessments and the annual accountability tests that help us know if students are on track to success.

However, our thinking about clear targets cannot stop here. We have yet to account for the classroom level, where we need to help students ascend through the levels of proficiency leading to mastery of each standard. Every standard must be deconstructed into the scaffolding that students must climb on their journey to success. These continuously unfolding classroom targets, then, become the focus of day-to-day formative assessments (assessments used to guide learning).

To accomplish this deconstruction, faculties can work in teams to organize standards and build the scaffolding for each standard. They can start by asking specific questions for each standard. Following is an example of this process for the standard, “Learn to write proficiently.” The specific questions are followed by sample answers (oversimplified here for the purpose of demonstration).

- What must our students *know and understand* when the time comes for them to demonstrate that they have met this standard? What are the foundations of knowledge that underpin success here? In our example, the student must master the vocabulary, syntactic structure, and mechanics needed to communicate ideas, and, in addition, must bring knowledge about the subject to his or her writing.

- What patterns of *reasoning*, if any, must students have mastered to be ready to demonstrate that they have met this standard? Must they be prepared to use their knowledge to reason productively in this case? In our example, the student must infer what words, sentence structures, and organizational principles will lead to the clearest communication of the ideas to be shared in the writing. He or she must also be able

to evaluate the quality of the writing as it is being composed using appropriate quality criteria and make adjustments as needed to improve communication.

- What *performance skills*, if any, does this standard assume our students will master on their journey to competence? Do we expect achievement-related behaviors? In our example, the student must have mastered either the penmanship or keyboarding skills needed to record ideas for others to see.
- What *products*, if any, must our students learn to create to be judged competent in terms of this standard? In our example, the student must be able to compose an actual sample of his or her own original text that satisfies standards of word choice, sentence fluency, and organization.

The answers to these questions determine the progression of learning that will help students climb the scaffolding to meet each standard. These, then, will need to be the focus of day-to-day classroom assessment to inform the question, “What comes next in the learning?”

Key 3: Accurate Assessment

Given the need to provide information to assessment users about student mastery of specific targets in various contexts, the assessor must be prepared to design and build or select quality assessments for many different kinds of assessment circumstances. While this does not include the institutional or policy level where state assessments rule, it can include the interim level with common assessments, and it certainly does include the classroom level. To create an accurate assessment, the developer must do the following:

- Select a proper assessment *method*, such as selected response, written response, performance assessment, or personal communication.

- Build each assessment out of quality *ingredients*, whether they are test items, performance or essay tasks, or scoring guides.
- Include enough items to *sample* appropriately student knowledge so as to gather enough evidence to lead to a confident conclusion about achievement without wasting time gathering too much.
- Anticipate all relevant sources of *bias* that can distort results, such as wording in items that gives members of one cultural subgroup an advantage over another (perhaps due to differences in language or social experience); these must be avoided or eliminated if they creep into the assessment.

Assessment developers who cannot follow these guidelines place students in harm's way with the possibility of mismeasuring their achievement. Now let us consider each of these items in greater detail.

Proper methods. There is a variety of assessment methods to choose from: selected response formats (multiple choice, true/false, matching), written response or essay, performance assessment (observation and judgment), and direct personal interaction with our students (talking to them). These methods are not interchangeable. We cannot have a “favorite” method that we use every time. The assessment method must fit the context.

In education today, we continue to rely far too heavily on multiple-choice tests. It is simply not possible to obtain a comprehensive portrait of student achievement with only a multiple-choice test. That method is both efficient and powerful in a very limited array of contexts. But like the other methods, it is a formula for disaster in the hands of an incompetent user or common-assessment team. If they are used in a context (for a purpose or with a target) where they cannot provide a quality portrait of achievement, the result will be a misrepresentation of student learning and counterproductive instructional decisions.

No assessment method is inherently superior to others in all contexts. Assessment methods must be chosen carefully to reflect the achievement target(s) in question and to provide the information needed by the intended user(s). (See Stiggins, Arter, Chappuis, and Chappuis [2004] for specific guidance on choosing proper assessment methods.)

Quality ingredients. Once a method is selected, the assessment must be built of quality items, tasks, exercises, and scoring schemes. This is crucial. Performance assessments must be constructed of high-quality performance tasks and scoring rubrics; multiple-choice tests must be constructed of high-quality questions. Developers must understand the difference. This too may require some time spent in professional learning, depending on the developers' assessment backgrounds.

Quality sample. A quality assessment gathers enough information to lead to a confident conclusion, but it does not gather too much information. So how much is enough evidence? How many test items? How many performance tasks or essay exercises? The context determines the answers to these questions. The higher the stakes, the more certain we must be, and the more data we must gather. The more complex the target, the more evidence we may need. The more evidence a task provides, the fewer tasks we need. Each assessment method brings with it a set of rules for use that maximize sampling efficiency. One must take advantage of professional learning opportunities to master those sampling rules.

Minimizing bias. Even if we select a proper method, use high-quality ingredients, and sample appropriately, there are many things that can go wrong and distort results. For example, emotional upsets can keep students from providing dependable information about their achievement, and distractions during the administration of an assessment can keep them from showing their true level of attainment. If students are not disposed to let us know how much they

have achieved, our assessments will not provide dependable evidence of their learning. Bias can also arise in the scoring process when scoring is subjective, as it is in essay or performance assessment. For example, bias has crept into the assessment when the scorer's knowledge of a student's ethnic background inappropriately influences judgments of proficiency. If we fail to eliminate these sources of bias, our evidence will be inaccurate.

Key 4: Effective Communication

Once the information needs of assessment users have been identified, achievement expectations are in place, and accurate assessments are being used, the foundation is laid for gathering good data. But note that all of this work is wasted if procedures are not also in place to deliver the assessment results into the user's hands in a timely and understandable form.

For communication to be effective in the context of formative assessment, it needs to inform the learner about how to do better the next time; that is, it must be descriptive rather than judgmental. Feedback is most helpful when it focuses on attributes of the student's work ("Change your writing in this way . . ."), not when it focuses on the student's learning ("Try harder"). It must provide sufficient detail to inform without overwhelming, and it must arrive in time to help the learner.

These characteristics of effective communication are satisfied most easily when all involved parties agree in advance about the important achievement expectations, or learning targets. A lack of understanding about what success looks like is a barrier to understanding the true meaning and implications of assessment results.

Effective communication also depends on the availability of accurate information about each student's achievement (see Key 3). Inaccurate or inaccessible information serves no one well, as it can lead to counterproductive instructional decisions. To ensure the

effective transmission of information from one person to another, both parties must understand the meaning of the symbols involved, whether they are grades, test scores, work samples, or verbal descriptions of achievement. For example, miscommunication will occur if it is assumed that a report-card grade indicates a student's level of achievement when, in fact, the teacher has woven achievement, effort, attitude, compliance with rules, attendance, and other nonachievement characteristics into the grade.

Whether administering daily classroom assessments, common assessments in a professional learning community, or state assessments for accountability, the foundations of productive assessment are the same: We need accurate information about student learning, effectively communicated. A quality assessment arises from a clear knowledge of the information needs of the intended user and a clear vision of the achievement target to be assessed. With these in hand, we can design an accurate assessment—one that relies on a proper assessment method, includes high-quality items or exercises, samples appropriately, and eliminates all relevant sources of bias.

Using Assessment to Encourage and Support Learning

In the opening to this chapter, I argued that quality assessment represents a necessary but insufficient condition for productive assessment. I hope it is clear now specifically what we mean by quality, and how to achieve it. The second active ingredient in productivity is the effective use of the assessment process and its results to help students advance their learning with enthusiasm and feel in control of their learning as they attain new levels of proficiency.

While classroom assessment certainly can and should serve accountability purposes by providing evidence for report-card grading, we can and should take advantage of the process to help us accomplish far more than that: It can serve to promote student success. Traditionally, we think of assessment as an index of student

attainment, but it also can serve as the cause of achievement. In our work at the ETS Assessment Training Institute (ETIs), we refer to these two purposes as *assessment of learning* and *assessment for learning*. With *assessment for learning*, all students can experience the ongoing joy and optimism that comes from expecting to succeed and living up to that expectation.

Connecting Assessment of and for Learning

Traditionally, we have used assessments to discover how much our students have learned up to a particular point in time. Evidence from these assessments, according to our effective schools models, must be fed to the adults in the system so they can make informed instructional decisions to help students. This certainly makes sense in terms of school improvement under certain conditions. At ETS Assessment Training Institute, we call this *assessment of learning*.

But what if we supplement it with *assessment for learning* by asking, “How can we use the assessment process to cause students to learn more; that is, to increase achievement in the future?” To accomplish this, we use assessment to inform students about themselves. If assessments *of learning* check to see if our students are meeting standards (state, district, or classroom), assessments *for learning* ask if our students are making progress toward meeting those standards (day to day in the classroom—during the learning). One is for accountability, while the other is used to support learning. Both are important, but they are different because they serve fundamentally different purposes. The key to our collective success as educators is to balance the two—to find the synergy between them.

Examples of assessments *of learning* arise from our accountability legacy: externally imposed standardized tests like college admissions tests, state assessments, district-wide tests, and so on. They also include classroom assessments used to assign report-card grades such as unit tests, final exams, and the like. These are assessments conducted after a unit or class is finished to determine if learning

has occurred. They inform multiple levels of accountability and instructional decisions and, therefore, are important.

Examples of assessments *for learning* are those that we use to diagnose student needs, support students’ practice, or help students watch themselves improving over time. In all cases, we seek to provide teachers and students with the kinds of information they need to make decisions that promote continued learning. Assessments *for learning* occur while the learning is still happening and throughout the learning process. So early in the learning, students’ scores will not be high. This is not failure—it simply represents where students are now in their ongoing journey to ultimate success.

For example, when an elementary student is beginning to write, his or her emergent writer proficiencies—such as word choice, sentence structure, and organization of ideas—may be lacking. If the faculty has teamed up to map the learning progressions that lead to proficiency in writing, and then discovers which learning strategies are working for the student, they can provide the student with descriptive feedback that will help him or her understand how to do better the next time. The careful management of this kind of assessment and feedback, within and across grade levels in ways that involve students as partners over the long haul, will result in competent writers.

The teacher’s role in assessment *of learning* is as it always has been: to administer accurate assessments and use sound grading practices. But in assessment *for learning*, this role changes. The teacher’s role in this case is to carry out the following sequence:

1. Become a confident, competent master of the standard our students are expected to master.
2. Deconstruct each standard into the enabling classroom achievement targets that form the scaffolding leading up to the standard.

3. Create a student-friendly version of those targets to share with students from the beginning of the learning.
4. Create high-quality classroom assessments that reflect those targets.
5. Use those assessments (in collaboration with students) to track improvement over time.

The student's role in assessment *of* learning is as it always has been: to study hard and strive for the highest scores and grades; that is, demonstrate competence. But in assessment *for* learning, the student's role is to strive to understand what success looks like and to use each assessment to try to understand how to do better the next time. In other words, students seek to understand what good writing looks like so they can assess where they are currently and then close the gap between the two.

This leads to a fundamental redefinition of the relationship between assessment and student motivation. Rather than relying on assessment as the source of information used to decide who gets rewarded and punished—that is, for sorting winners from losers—we use assessment as a road map to ultimate success, with signposts along the way for both students and their teachers. Success at making progress (at learning) becomes its own reward, promoting confidence and persistence. This changes the emotional dynamics of the assessment experience in immensely productive ways for all students, especially for those who have not yet met standards. Students become good writers not to earn a good grade, but because they believe they can, and it is that belief that motivates them.

The Foundations of Assessment for Learning

We know that the consistent application of principles of assessment *for* learning enhances student learning, and we understand why. The evidence that it works in promoting greater achievement has been summarized by Black and William (1998). Based on results of

dozens of studies conducted around the world, we know that students experience profound gains in their achievement with the largest gains accruing for perennial low achievers. The Black and William synthesis instructs us that the keys to maximizing these gains are to increase:

- The accuracy of classroom assessments
- Student access to descriptive (versus judgmental) feedback
- Student involvement in assessment, record-keeping, and communication

It should be self-evident that accurate assessments lead to better instructional decisions than do inaccurate assessments. Descriptive feedback provided to students who understand the target will power them toward success far more productively than will judgmental feedback in the form of grades, for example. This is especially true for struggling learners.

These practices enhance the learner's sense of control of his or her academic success or academic self-efficacy. Albert Bandura (1994/1998) has described the psychology that, in effect, can be interpreted to underpin assessment *for* learning as I am defining it:

A strong sense of efficacy enhances human accomplishment and personal well-being in many ways. People with high assurance in their capabilities approach difficult tasks as challenges to be mastered rather than as threats to be avoided. Such an efficacious outlook fosters intrinsic interest and deep engrossment in activities. They set themselves challenging goals and maintain strong commitment to them. They heighten and sustain their efforts in the face of failure. They quickly recover their sense of efficacy after failures or setbacks. They attribute failure to insufficient effort or deficient knowledge and skills which are acquirable. They approach threatening situations with assurance that they can exercise control over them. . . .

In contrast, people who doubt their capabilities shy away from difficult tasks which they view as personal threats. They have low aspirations and weak commitment to the goals they choose to pursue. When faced with difficult tasks, they dwell on their personal deficiencies, on the obstacles they will encounter, and all kinds of adverse outcomes rather than concentrate on how to perform successfully. They slacken their efforts and give up quickly in the face of difficulties. They are slow to recover their sense of efficacy following failure or setbacks. Because they view insufficient performance as deficient aptitude it does not require much failure for them to lose faith in their capabilities. (pp. 1–2)

These two paragraphs capture the essence of what differentiates the productive and counterproductive emotional dynamics of assessment. Productive assessment leaves students feeling in control and optimistic, even if they do not perform well. Counterproductive assessment robs students of that sense of control, resulting in a pervasive sense of hopelessness.

How do we help students remain optimistic about their potential success at learning? One way is to help them understand from the beginning of the learning what success will look like when they get there (Sadler, 1989). We do this by starting instruction with a student-friendly version of the learning targets. We accompany this with actual samples of student work that illustrate the continuum of achievement students will travel on their journey to success. This gives students a frame of reference from which to track their own progress.

Another way is to help students know where they are now in relation to that vision of excellence (Sadler, 1989). We do this by providing them with continuous access to descriptive feedback that shows them how to do better the next time. And we help them learn to self-assess so that, over time, they can learn to generate their own

descriptive feedback. When we do these things, we bring our students to a place where they can become partners with us, their teachers, in setting goals for what comes next in the learning. This builds a strong sense of academic self-efficacy.

Finally, we help our students become increasingly efficacious when we show them how to close the gap between where they are now and where we want them to be in their learning (Sadler, 1989). We do this when we help them learn to improve the quality of their work one key attribute at a time, when we help them learn to see and keep track of changes in their own capabilities, and when we help them reflect on the relationship between those improvements and their own actions.

All of these strategies manifest specific ways to use assessment for learning. Making these strategies operational in the classroom requires in-depth professional development (see Stiggins, Arter, Chappuis, & Chappuis, 2004 for further details). These strategies are most obviously and easily applied at the classroom level. However, they also represent productive options in the development and use of common assessments for program evaluation and improvement. Imagine a professional learning community in which students become members of teams to design, conduct, and interpret common assessments. Students can collaborate to deconstruct standards, transform classroom targets into student-friendly language, devise assessment exercises, score schemes and assessments, and interpret results with an eye toward their own self-interest. Obviously, they would need careful guidance from their teachers, each of whom must understand the principles of productive assessment from the outset. But I believe a vision of partnership at this level of assessment holds immense promise.

References

Bandura, A. (1998). Self-efficacy. In H. Friedman (Ed.), *Encyclopedia of mental health*. San Diego: Academic Press. (Reprinted from *Encyclopedia of*

human behavior, Vol. 4, pp. 71–81, by V. S. Ramachaudran, Ed., 1994, New York: Academic Press.)

Black, P., & William, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles and Practice*, 5(1), 7–74. Also summarized in Black, P., & William, D., (1998), Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.

DuFour, R. (2005). What is a professional learning community? In R.

DuFour, R. Eaker, & R. DuFour (Eds.), *On common ground: The power of professional learning communities* (pp. 31–43). Bloomington, IN: Solution Tree (formerly National Educational Service).

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.

Stiggins, R. J., Arter, J., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment FOR student learning: Doing it right—Using it well*. Portland, OR: ETS Assessment Training Institute.