

A SPECIAL SECTION ON ASSESSMENT

Assessment Crisis: The Absence Of Assessment *FOR* Learning

If we wish to maximize student achievement in the U.S., we must pay far greater attention to the improvement of classroom assessment, Mr. Stiggins warns. Both assessment of learning and assessment for learning are essential. But one is currently in place, and the other is not.

BY RICHARD J. STIGGINS

A real voyage of discovery consists not of seeking new landscapes but of seeing through new eyes.

— *Marcel Proust*

IF WE ARE finally to connect assessment to school improvement in meaningful ways, we must come to see assessment through new eyes. Our failure to find a potent connection has resulted in a deep and intensifying crisis in assessment in American education. Few elected officials are aware of this crisis, and almost no school officials know how to address it. Our current assessment systems are harming huge numbers of students for reasons that few understand. And that harm arises directly from our failure to balance our use of standardized tests and classroom assessments in the service of school improvement. When it comes to assessment, we have been trying to find answers to the wrong questions.

Politicians routinely ask, How can we use assessment as the basis for doling out rewards and punishments to increase teacher and student effort? They want to know how we can intensify the intimidation associated with annual testing so as to force greater achievement. How we answer these questions will certainly affect schools. But that impact will not always be posi-

RICHARD J. STIGGINS is president of Assessment Training Institute, Inc., Portland, Ore.

tive. Moreover, politicians who ask such questions typically look past a far more important pair of prior questions: How can we use assessment to help all our students *want* to learn? How can we help them feel *able* to learn? Without answers to these questions, there will be no school improvement. I explain why below.

School administrators in federal, state, and local education agencies contribute to our increasingly damaging assessment crisis when they merely bow to politi-

cians' beliefs and focus unwaveringly on the question of how to make our test scores go up. To be sure, accountability for student learning is important. I am not opposed to high-stakes testing to verify school quality — as long as the tests are of sound quality.¹ However, our concern for test scores must be preceded by a consideration of more fundamental questions: Are our current approaches to assessment improving student learning? Might other approaches to assessment have a greater impact? Can we design state and district assessment systems that have the effect of helping our students want to learn and feel able to learn?

Furthermore, the measurement community, of which I am a member, also has missed an essential point. For decades, our priorities have manifested the belief that our job is to discover ever more sophisticated and efficient ways of generating valid and reliable test scores. Again, to be sure, accurate scores are essential. But there remains an unasked prior question: How can we maximize the positive impact of our scores on learners? Put another way, How can we be sure that our assessment instruments, procedures, and scores serve to help learners want to learn and feel able to learn?

We are a nation obsessed with the belief that the path to school improvement is paved with better, more frequent, and more intense standardized testing. The problem is that such tests, ostensibly developed to “leave no student behind,” are in fact causing major segments of our student population to be left behind because the tests cause many to give up in hopelessness — just the opposite effect from that which politicians intended.

Student achievement suffers because these once-a-year tests are incapable of providing teachers with the moment-to-moment and day-to-day information about student achievement that they need to make crucial instructional decisions. Teachers must rely on classroom assessment to do this. The problem is that teachers are unable to gather or effectively use dependable information on student achievement each day because of the drain of resources for excessive standardized testing. There are no resources left to train teachers to create and conduct appropriate classroom assessments. For the same reasons, district and building administrators have not been trained to build assessment systems that balance standardized tests and classroom assessments. As a direct result of these chronic, long-standing problems, our classroom, building, district, state, and national assessment systems remain in constant crisis, and students suffer the consequences. All school practitioners know this, yet almost no politicians do.

We know how to build healthy assessment environments that can meet the information needs of all instructional decision makers, help students want to learn and feel able to learn, and thus support unprecedented increases in student achievement. But to achieve this goal, we must put in place the mechanisms that will make healthy assessment possible. Creating those mechanisms will require that we begin to see assessment through new eyes. The well-being of our students depends on our willingness to do so.

THE EVOLUTION OF OUR VISION OF EXCELLENCE IN ASSESSMENT

The evolution of assessment in the United States over the past five decades has led to the strongly held view that school improvement requires:

- the articulation of higher achievement *standards*,
- the transformation of those expectations into rigorous *assessments*, and
- the expectation of *accountability* on the part of educators for student achievement, as reflected in test scores.

Standards frame accepted or valued definitions of academic success. Accountability compels attention to these standards as educators plan and deliver instruction in the classroom. Assessment provides the evidence of success on the part of students, teachers, and the system.

To maximize the energy devoted to school improvement, we have “raised the bar” by setting world-class standards for student achievement, as opposed to minimum competencies. To further intensify the impact of our standards and assessments, policy makers often attach the promise of rewards for schools that produce high scores and sanctions for schools that do not.

In this context, we rely on high-stakes assessments *of learning* to inform our decisions about accountability. These tests tell us how much students have learned, whether standards are being met, and whether educators have done the job they were hired to do.

Such assessments of learning have been the norm throughout the U.S. for decades. We began with standardized college admissions tests in the early decades of the last century, and this use of testing continues essentially unchanged today. But these tests are not used merely for college admission. For decades, we have ranked states according to average SAT scores.

Meanwhile, in response to demands for accountability in public schools in the 1960s, we launched dis-

trictwide standardized testing programs that also remain in place today. In the 1970s, we began the broad implementation of statewide testing programs, and these programs have spread throughout the land. Also in the 1970s and extending into the 1980s, we added a national assessment program that continues to this day. During the 1990s, we became deeply involved and invested in international assessment programs. Across the nation, across the various levels of schooling, and over the decades, we have invested billions of dollars to ensure the accuracy of the scores on these assessments of learning. Now in 2002, President Bush has signed a school reform measure that requires standardized testing of every pupil in the U.S. in mathematics and reading every year in grades 3 through 8, once again revealing our faith in assessment as a tool for school improvement.

In the context of school improvement, we have seen assessment merely as an index of the success of our efforts. It is testimony to our societal belief in the power of standardized tests that we would permit so many levels of testing to remain in place, all at the same time and at very high cost. Clearly, over the decades, we have believed that by checking achievement status and reporting the results to the public we can apply the pressure needed to intensify — and thus speed — school improvement. At the same time, we have believed that providing policy makers and practicing educators with test results can inform the critically important school improvement decisions that are made at district, state, and federal levels.

THE FLAW IN THE VISION

The assessment environment described above is a direct manifestation of a set of societal beliefs about what role assessment ought to play in American schools. Over the decades, we have succeeded in carrying these beliefs to unfortunate extremes.

For example, we have believed that assessment should serve two purposes: inform decisions and motivate learning. With respect to the former, we have built our assessment systems around the belief that the most important decisions are made by those program planners and policy makers whose actions affect the broadest range of classrooms and students. The broader the reach of the decision makers (across an entire school district or state), the more weight we have given to meeting their information needs first. This is the foundation of our strong belief in the power of standardized tests. These

are the tests that provide comparable data that can be aggregated across schools, districts, and states to inform far-reaching programmatic decisions.

With respect to the use of assessment to motivate, we all grew up in classrooms in which our teachers believed that the way to maximize learning was to maximize anxiety, and assessment has always been the great intimidator. Because of their own very successful experiences in ascending to positions of leadership and authority, most policy makers and school leaders share the world view that, “when the going gets tough, the tough get going.” They learned that the way to succeed when confronted with a tougher challenge is to redouble your efforts — work harder and work smarter. If you do so, you win. And so, they contend, the way to cause students to learn more — and thus the way to improve schools — is to confront them with a tougher challenge. This will cause them to redouble their efforts, they will learn more, their test scores will go up, and the schools will become more effective. We can motivate students to greater effort, they believe, by “setting higher academic standards,” “raising the bar,” and implementing more high-stakes testing. This is the foundation of our belief in the power of accountability-oriented standardized tests to drive school improvement.

In point of fact, when some students are confronted with the tougher challenge of high-stakes testing, they do redouble their efforts, and they do learn more than they would have without the added incentive. Please note, however, that I said this is true for “some students.”

Another huge segment of our student population, when confronted with an even tougher challenge than the one that it has already been failing at, will not redouble its efforts — a point that most people are missing. These students will see both the new high standards and the demand for higher test scores as unattainable for them, and they will *give up in hopelessness*.

Many political and school leaders have never experienced the painful, embarrassing, and discouraging trauma of chronic and public academic failure. As a result, they have no way of anticipating or understanding how their high-stakes testing program, whether local or statewide, could lead to even greater failure for large numbers of students. But tapping the intimidation power of standardized tests for public accountability has an effect on the success of this segment of the student population that is exactly the opposite of what we intend.

Thus it is folly to build our assessment environments

on the assumption that standardized testing will have the same effect on all students. It will not. Some students approach the tests with a strong personal academic history and an expectation of success. Others approach them with a personal history and expectation of very painful failure. Some come to slay the dragon, while others expect to be devoured by it. As a result, high-stakes assessment will enhance the learning of some while discouraging others and causing them to give up. Yet, as they attempt to weave assessment into the school improvement equation, federal, state, and local policy makers seem unable to understand or to accommodate this difference.

A MORE POWERFUL VISION

There is another way in which assessment can contribute to the development of effective schools that has been largely ignored in the evolution of the standards, assessment, and accountability movement described above. We can also use assessments *for learning*.² If assessments *of learning* provide evidence of achievement for public reporting, then assessments *for learning* serve to help students learn more. The crucial distinction is between assessment to determine the status of learning and assessment to promote greater learning.

Assessments *of* and *for* learning are both important. Since we in the U.S. already have many assessments *of learning* in place, if we are to balance the two, we must make a much stronger investment in assessment *for learning*. We can realize unprecedented gains in achievement if we turn the current day-to-day classroom assessment process into a more powerful tool for learning. We know that schools will be held accountable for raising test scores. Now we must provide teachers with the assessment tools needed to do the job.

It is tempting to equate the idea of assessment *for learning* with our more common term, “formative assessment.” But they are not the same. Assessment *for learning* is about far more than testing more frequently or providing teachers with evidence so that they can revise instruction, although these steps are part of it. In addition, we now understand that assessment *for learning* must involve students in the process.

When they assess *for learning*, teachers use the classroom assessment process and the continuous flow of information about student achievement that it provides in order to advance, not merely check on, student learning. They do this by:

- understanding and articulating *in advance of teaching* the achievement targets that their students are to hit;
- informing their students about those learning goals, *in terms that students understand*, from the very beginning of the teaching and learning process;
- becoming assessment literate and thus able to transform their expectations into assessment exercises and scoring procedures that *accurately reflect student achievement*;
- using classroom assessments to *build students' confidence* in themselves as learners and help them take responsibility for their own learning, so as to lay a foundation for lifelong learning;
- translating classroom assessment results into frequent *descriptive feedback* (versus judgmental feedback) for students, providing them with specific insights as to how to improve;
- continuously *adjusting instruction* based on the results of classroom assessments;
- engaging students in *regular self-assessment*, with standards held constant so that students can watch themselves grow over time and thus feel in charge of their own success; and
- actively involving students in *communicating* with their teacher and their families about their achievement status and improvement.

In short, the effect of assessment *for learning*, as it plays out in the classroom, is that students keep learning and remain confident that they can continue to learn at productive levels if they keep trying to learn. In other words, students don't give up in frustration or hopelessness.

ARE TEACHERS READY?

Few teachers are prepared to face the challenges of classroom assessment because they have not been given the opportunity to learn to do so. It is currently the case that only about a dozen states explicitly require competence in assessment as a condition to be licensed to teach. Moreover, there is no licensing examination in place at the state or federal level in the U.S. that verifies competence in assessment. Thus teacher preparation programs have taken little note of competence in assessment, and the vast majority of programs fail to provide the assessment literacy required to enable teachers to engage in assessment *for learning*. It has been so for decades.

Furthermore, lest we believe that teachers can turn to their principals for help, it is currently the case that

almost no states require competence in assessment as a condition to be licensed as a principal or school administrator at any level. Consequently, assessment training is almost nonexistent in administrator training programs. It has been so for decades.

Thus we remain a national faculty that is unschooled in the principles of sound assessment — whether assessment *of* or *for* learning. This fact has been a matter of record for decades. To date, as a nation, we have invested almost nothing in assessment *for learning*. Teachers rarely have the opportunity to learn how to use assessment as a teaching and learning tool. And our vigorous efforts to assess learning through our various layers of standardized tests cannot overcome the effects of this reality.

As a result of this state of affairs, we face the danger that student progress may be mismeasured, day to day, in classrooms across the nation. That means that all the critically important day-to-day instructional decisions made by students, teachers, and parents may be based on misinformation about student success. The result is the misdiagnosis of student needs, students' misunderstanding of their own ability to learn, miscommunication to parents and others about student progress, and virtually no effective assessment *for learning* in classrooms. The extremely harmful consequences for student learning are obvious.

RELEVANT POSITION STATEMENTS

The dire consequences of this assessment crisis and the urgent need for action have not gone unnoticed. For example, during the 1990s, virtually every professional association that had anything to do with teaching adopted standards of professional competence for teachers that include an assessment component.³ This group included the American Federation of Teachers (AFT), the National Education Association (NEA), the Council of Chief State School Officers, the National Board for Professional Teaching Standards, and the National Council on Measurement in Education (NCME).

The documents that were issued included a collaborative statement of assessment competencies for teachers developed by a joint committee representing AFT, NEA, and NCME.⁴ In addition to other standards, this joint statement expects teachers to be trained to choose and develop proper assessment methods; to administer, score, and interpret assessment results; to connect those results to specific decisions; to assign grades appropri-

ately; and to communicate effectively about student achievement. It is troubling to realize that these standards are more than a decade old and still have had little impact on the preparation of teachers and administrators.

In its 2001 report, the Committee on the Foundations of Assessment of the National Research Council advanced recommendations for the development of assessment in American schools that included the following:

Recommendation 9: Instruction in how students learn and how learning can be assessed should be a major component of teacher preservice and professional development programs. This training should be linked to actual experience in classrooms in assessing and interpreting the development of student competence. To ensure that this occurs, state and national standards for teacher licensure and program accreditation should include specific requirements focused on the proper integration of learning and assessment in teachers' educational experience.⁵

* * *

Recommendation 11: The balance of mandates and resources should be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment designed to assist learning.⁶

Similarly, the Commission on Instructionally Supportive Assessment convened by the American Association of School Administrators, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the NEA, and the National Middle School Association included the following in its list of nine requirements for state-mandated accountability tests:

Requirement 8: A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessment.⁷

We understand what teachers need to know and the proficiencies that they need to develop in order to be able to establish and maintain productive assessment environments. The challenge we face is to provide the opportunity for teachers to master those essential classroom assessment competencies. The depth of this challenge becomes clear when we realize that we must pro-

vide opportunities both for new teachers to gain these competencies before they enter the classroom and for experienced teachers who had no chance to master them during their training to gain them as well.

BALANCING ASSESSMENTS OF AND FOR LEARNING

Therefore, our national assessment priority should be to make certain that assessments both *of* and *for* learning are accurate in their depiction of student achievement and are used to benefit students. Since our standardized assessments *of learning* have been developed by professionals and are currently in place, they are poised to detect any improvements in the level or rate of student achievement.

But these tests provide information only once a year, and we must not delude ourselves into believing that they can serve all assessment purposes. They can reflect large-group increases or decreases in learning on an annual basis, and they can serve as gatekeepers for high-stakes decisions. They cannot inform the moment-to-moment, day-to-day, and week-to-week instructional decisions faced by students and teachers seeking to manage the learning process as it unfolds. They cannot diagnose student needs during learning, tell students what study tactics are or are not working, or keep parents informed about how to support the work of their children. These kinds of uses require assessments *for learning*. The critical question for school improvement is, What would happen to standardized test scores if we brought assessments *for learning* online as a full partner in support of student learning? Several published reviews of research reveal the startling and very encouraging answer.

In 1984 Benjamin Bloom provided a summary of research comparing standard whole-class instruction (the control condition) with two experimental interventions, a mastery learning environment and one-on-one tutoring of individual students. One hallmark of both experimental conditions was the extensive use of classroom assessment *for learning* as a key part of the instructional process. The analyses revealed differences ranging from one to two standard deviations in student achievement attributable to differences between experimental and control conditions.⁸

In their 1998 research review, Paul Black and Dylan Wiliam examined the research literature on assessment worldwide, asking if improved formative (i.e., classroom) assessments yield higher student achievement as reflected in summative assessments. If so, they

asked, what kinds of improvements in classroom assessment practice are likely to yield the greatest gains in achievement?

Black and Wiliam uncovered and then synthesized more than 250 articles that addressed these issues. Of these, several dozen directly addressed the question of the impact on student learning with sufficient scientific rigor and experimental control to permit firm conclusions. Upon pooling the information on the estimated effects of improved formative assessment on summative test scores, they reported unprecedented positive effects on student achievement. They reported effect sizes of one-half to a full standard deviation. Furthermore, Black and Wiliam reported that “improved formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement overall.”⁹ This result has direct implications for districts seeking to reduce achievement gaps between minorities and other students. Hypothetically, if assessment *for learning*, as described above, became standard practice only in classrooms of low-achieving, low-socioeconomic-status students, the achievement gaps that trouble us so deeply today would be erased. I know of no other school improvement innovation that can claim effects of this nature or size.

To fully appreciate the magnitude of the effect sizes cited above, readers need to understand that a gain of one standard deviation, applied to the middle of the test score distribution on commonly used standardized achievement tests, can yield average gains of more than 30 percentile points, two grade-equivalents, or 100 points on the SAT scale. Black and Wiliam report that gains of this magnitude, if applied to the most recent results of the Third International Mathematics and Science Study, would have raised a nation in the middle of the pack among the 42 participating countries (where the U.S. is ranked) to the top five.

This research reveals that these achievement gains are maximized in contexts where educators increase the accuracy of classroom assessments, provide students with frequent informative feedback (versus infrequent judgmental feedback), and involve students deeply in the classroom assessment, record keeping, and communication processes. In short, these gains are maximized where teachers apply the principles of assessment *for learning*.

Black and Wiliam conclude their summary of self-assessment by students as follows:

Thus self-assessment by pupils, far from being a

luxury, is in fact *an essential component of formative assessment*. When anyone is trying to learn, feedback about the effort has three elements: redefinition of the *desired goal*, evidence about *present position*, and some understanding of a *way to close the gap between the two*. All three must be understood to some degree by anyone before he or she can take action to improve learning.¹⁰ (Emphasis in original.)

ANTICIPATING THE BENEFITS OF BALANCE

Students benefit from assessment *for learning* in several critical ways. First, they become more confident learners because they get to watch themselves succeeding. This success permits them to take the risk of continuing to try to learn. The result is greater achievement for all students — especially low achievers, which helps reduce the achievement gap between middle-class and low-socioeconomic-status students. Furthermore, students come to understand what it means to be in charge of their own learning — to monitor their own success and make decisions that bring greater success. This is the foundation of lifelong learning.

Teachers benefit because their students become more motivated to learn. Furthermore, their instructional decisions are informed by more accurate information about student achievement. Teachers also benefit from the savings in time that result from their ability to develop and use classroom assessments more efficiently.

Parents benefit as well in seeing higher achievement and greater enthusiasm for learning in their children. They also come to understand that their children are learning to manage their own lifelong learning.

School administrators and instructional leaders benefit from the reality of meeting accountability standards and from the public recognition of doing so. Political officials benefit in the same way. When schools work more effectively, both political leaders and school leaders are recognized as contributing to that outcome.

In short, everyone wins. There are no losers. But the price that we must pay to achieve such benefits is an investment in teachers and their classroom assessment practices. We must initiate a program of professional development specifically designed to give teachers the expertise they need to assess *for learning*.

AN ACTION PLAN

If we wish to maximize student achievement in the

U.S., we must pay far greater attention to the improvement of classroom assessment. Both assessment *of learning* and assessment *for learning* are essential. One is in place; the other is not. Therefore, we must:

- match every dollar invested in instruments and procedures intended for assessment *of learning* at national, state, and local levels with another dollar devoted to the development of assessment *for learning*;
- launch a comprehensive, long-term professional development program at the national, state, and local levels to foster literacy in classroom assessment for teachers, allocating sufficient resources to provide them with the opportunity to learn and grow professionally;
- launch a similar professional development program in effective large-scale and classroom assessment for state, district, and building administrators, teaching them how to provide leadership in this area of professional practice;
- change teacher and administrator licensing standards in every state and in all national certification contexts to reflect an expectation of competence in assessment both *of* and *for* learning; and
- require all teacher and administrator preparation programs to ensure that graduates are assessment literate — in terms both of promoting and of documenting student learning.

Federal education officials, state policy makers, and local school leaders must allocate resources in equal proportions to ensure the accuracy and effective use of assessments both *of* and *for* learning. Only then can we reassure families that their children are free from the harm that results from the mismeasurement of their achievement in schools. Only then can we maximize students' confidence in themselves as learners. Only then can we raise achievement levels for all students and "leave no child behind."

1. For specific standards of quality, refer to Commission on Instructionally Supportive Assessment, *Building Tests to Support Instruction and Accountability* (Washington, D.C.: AASA, NAESP, NASSP, NEA, and NMSA, 2001).

2. This term was coined by Assessment Reform Group, *Assessment for Learning: Beyond the Black Box* (Cambridge: School of Education, Cambridge University, 1999).

3. See the special section on Quality Teaching for the 21st Century in the November 1996 *Phi Delta Kappan*, pp. 190-227.

4. American Federation of Teachers, National Council on Measurement in Education, and National Education Association, "Standards for Teacher Competence in Educational Assessment of Students," *Educational Measurement: Issues and Practice*, vol. 9, no. 4, 1990, pp. 30-32.

5. James W. Pellegrino, Naomi Chudowsky, and Robert Glaser, eds., *Knowing What Students Know: The Science and Design of Educational Assessment* (Washington, D.C.: National Academy Press, 2001), p. 14.

6. Ibid.
7. Commission on Instructionally Supportive Assessment, p. 25.
8. Benjamin Bloom, "The Search for Methods of Group Instruction as Effective as One-on-One Tutoring," *Educational Leadership*, May 1984, pp. 4-17.
9. Paul Black and Dylan Wiliam, "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan*, October 1998, p. 141. Their work is reported in more detail in idem, "Assessment and Classroom Learning," *Assessment in Education*, March 1998, pp. 7-74.
10. Black and Wiliam, "Inside the Black Box," p. 143. 