From Formative Assessment to Assessment FOR Learning:
A Path to Success in Standards-Based Schools

As the mission of schools changes from ranking students to ensuring that all learn to specified standards, Mr. Stiggins argues that the purpose and form of assessments must change as well.

BY RICK STIGGINS

SOCIETY HAS SEEN fit to redefine the role of its schools. No longer are they to be places that merely sort and rank students according to their achievement. Now, they are to be places where all students become competent, where all students meet prespecified standards and so are not left behind. With increasing intensity, policy makers are turning to assessment as the power tool that will compel schools to fulfill this new role. If we look closely at the union of this redefined mission and the growing reliance on assessment, we can find a surprising and immensely powerful way to use assessment in the development of effective schools.

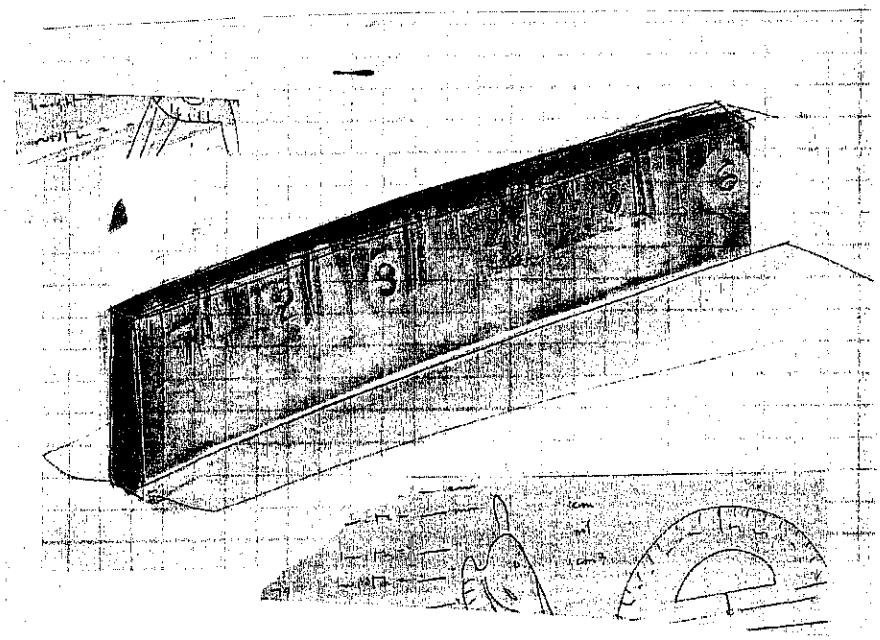
Traditionally, schools have used assessment — the pending final exam, the unannounced pop quiz, and the threat of low or failing report card grades — to motivate students. To maximize learning, our teachers believed, maximize anxiety. Assessment has served as the great intimidator. Pressure to get high test scores and good grades, it was believed, would motivate greater effort and thus more learning.

The recent change in the mission of schools has clouded this traditional view of the relationship between assessment and motivation. To see how and why, we must explore our assessment legacy and its motivational intricacies. As you will see, through that retrospective, we will discover a far more productive way for assessment to help students succeed.

RICK STIGGINS is the founder of Assessment Training Institute, Inc., Portland, Ore. The institute website can be accessed at www.assessmentinst.com. *©2005, Richard J. Stiggins.*

THE OLD MISSION AND ITS LEGACY

Today's adults grew up in schools designed to sort us into the various segments of our social and economic system. The amount of time available to learn was fixed: one year per grade. The amount learned by the end of that time was free to vary: some of us learned a great deal; some, very little. As we advanced through the grades, those who had learned a great deal in previous grades continued to build on those foundations. Those who had failed to master the early prerequisites within the allotted time failed to learn that which followed. After 12 or 13 years of cumulative treatment of this kind, we were, in effect, spread along an achievement continuum that was ultimately reflected in each student's rank in class upon graduation.



From the very earliest grades, some students learned a great deal very quickly and consistently scored high on assessments. The emotional effect of this was to help them to see themselves as capable learners, and so these students became increasingly confident in school. That confidence gave them the inner emotional strength to take the risk of striving for more success because they believed that success was within their reach. Driven forward by this optimism, these students continued to try hard, and that effort continued to result in success for them. They became the academic and emotional winners. Notice that the trigger for their emotional strength and their learning success was their perception of their success on formal and informal assessments.

But there were other students who didn't fare so well. They scored very low on tests, beginning in the earliest grades. The emotional effect was to cause them to question their own capabilities as learners. They began to lose confidence, which, in turn, deprived them of the emotional reserves needed to continue to take risks. Public failure was embarrassing, and it seemed better not to try and thus to save face. As their motivation waned, of course, their performance plummeted. These students embarked on what they believed to be an irreversible slide toward inevitable failure and lost hope. Once again, the emotional trigger for their decision not to try was *their perception* of their performance on assessments.

Consider the reality — indeed, the paradox — of the schools in which we were reared. If some students worked hard and learned a lot, that was a positive result, and they would finish high in the rank order. But if some students gave up in hopeless failure, that was an acceptable result, too, because they would occupy places very low in the rank order. Their achievement results fed into the implicit mission of schools: the greater the spread of achievement among students, the more it reinforced the rank order. This is why, if some students gave up and stopped trying (even dropped out of school), that was regarded as the student's problem, not the teacher's or the school's. The school's responsibility was to provide the opportunity to learn. If students didn't take advantage of the opportunity, that was fine within the system.

Once again, please notice who is using test results to decide whether to strive for excellence or give up in hopelessness. The "data-based decision makers" in this process are not teachers, not school leaders, and not policy makers. Rather, they are students themselves. Students are deciding whether success is within or beyond reach, whether the learning is worth the required effort, and so whether to try or not. The critical emotions underpinning the decision-making process include anxiety, fear of failure, uncertainty, and unwillingness to take risks — all triggered by students' perceptions of their own capabilities as reflected in assessment results.

Some students responded to the demands of such environments by working hard and learning a great deal. Others controlled their anxiety by giving up and not caring. The result for them? Exactly the opposite of the one society wants. Instead of leaving no child behind, these practices, in effect, drove down the achievement of at least as many students as they successfully elevated. And the evidence suggests that the downside victims are more frequently members of particular socioeconomic and ethnic minorities.

A NEW MISSION AND ITS EMOTIONAL PROMISE

In recent years, however, society has come to understand the limitations of schools that merely sort and rank students. We have discovered that students in the bottom one-third to one-half of the rank order — plus all who drop out before being ranked — fail to develop the foundational

Illustration: Artville

reading, writing, and mathematical proficiencies needed to survive in, let alone contribute to, an increasingly technically complex and ethnically diverse culture. So today, in asking schools to leave no child behind, society is asking that educators raise up the bottom of the rank-order distribution to a specified level of competence. We call those expectations our "academic achievement standards." Every state has them, and, as a matter of public policy, schools are to be held accountable for making sure that *all students* meet them.

To be clear, the mission of sorting has not been eliminated from the schooling process. For the foreseeable future, students will still be ranked at the end of high school. However, society now dictates that such a celebration of differences in amount learned must start at a certain minimum level of achievement for all.

The implications of this change in mission for the role of assessment are profound. Assessment and grading procedures designed to permit only a few students to succeed (those at the top of the rank-order distribution) must now be revised to permit the possibility that all students could succeed at some appropriate level. Furthermore, procedures that permitted (perhaps even encouraged) some students to give up in hopelessness and to stop trying must now be replaced by others that promote hope and continuous effort. In short, the entire emotional environment surrounding the prospect of being evaluated must change, especially for perennial low achievers.

The students' mission is no longer merely to beat other students in the achievement race. At least part of their goal must be to become competent. Teachers must believe that all students can achieve a certain level of academic success, must bring all of their students to believe this of themselves, must accommodate the fact that students learn at different rates by making use of differentiated instruction, and must guide all students toward the attainment of standards.

The driving dynamic force for students cannot merely be competition for an artificial scarcity of success. Because all students can and must succeed in meeting standards, cooperation and collaboration must come into play. The driving forces must be confidence, optimism, and persistence — for all, not just for some. All students must come to believe that they can succeed at learning if they try. They must have continuous access to evidence of what they believe to be credible academic success, however small. This new understanding has spawned increased interest in recent years in *formative assessment.*

FORMATIVE ASSESSMENT REVISITED: A KEY TO SUCCESS

Since 1967, when Michael Scriven articulated the distinction between summative and formative program evaluation, and since 1971, when Benjamin Bloom, Thomas Hastings, and George Madaus extended the differentiation to various forms of assessment, *summative* assessment has referred to tests administered after learning is supposed to have occurred to determine whether it did.[1] Meanwhile, *formative* assessment has been the label used for assessments conducted during learning to promote, not merely judge or grade, student success.

Clearly, over the decades, the interest (and investment) in summative assessment has far outstripped that accorded to formative assessment, as layer upon layer of tests have been used for classroom grading, as well as for local, state, national, and international accountability testing. Within the past few years, however, formative assessment has emerged as an increasingly prominent tool for school improvement.

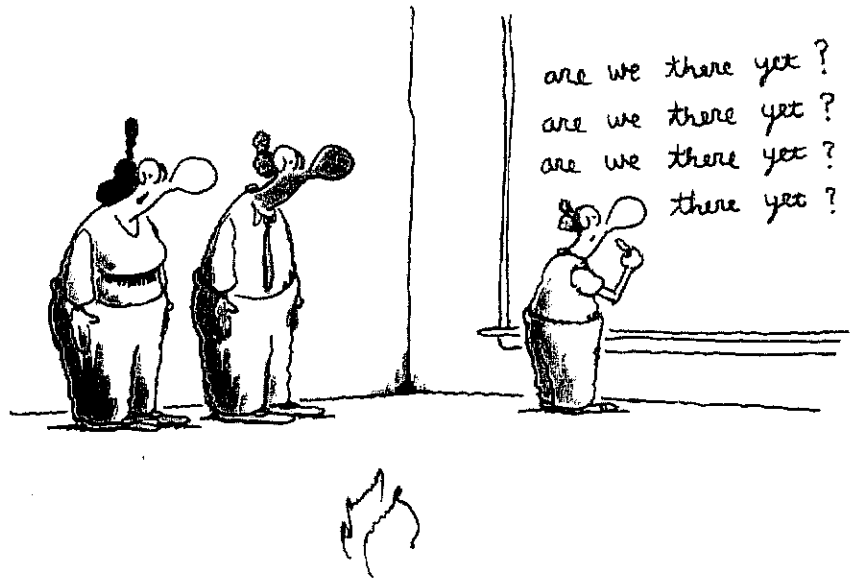
In its traditional form, formative assessment has been thought of as providing teachers with more frequent evidence of students' mastery of standards to help teachers make useful instructional decisions. In this way, formative assessment is intended to enhance student learning.

One reason for the recent resurgence of interest in formative assessment has been educators' realization that once-a-year summative standardized testing doesn't happen frequently enough to affect specific day-to-day, week-to-week, or even month-to-month instructional decisions. Besides, such testing fails to provide a sufficiently detailed picture of student learning to enable teachers to identify ways to help individual students. Typically, state accountability assessments include items covering many standards, and these are summed to yield a single overall proficiency score that is used to judge the sufficiency of student learning. These assessments tend not to provide evidence of each student's mastery of individual standards. Thus state summative test scores can inform accountability decisions but are not very helpful in guiding learning at the classroom level. So annual standardized tests have lacked sensitivity to instruction. Recently, attempts have emerged that are intended to overcome these inadequacies and make these large-scale assessments more practically useful.

Test more frequently. One approach that is beginning to emerge is to increase the frequency of summative assessments of standards from once to several times a year. Such evidence, it is argued, can give notice of instruction that is not working and so can inform programmatic changes that can increase the proportion of students

who meet standards. A few examples are short-cycle assessments, common assessments, benchmark tests, end-of-course examinations, and the quarterly or monthly formative standardized tests offered by some test publishers.

From a slightly different perspective, state departments of education embrace this approach when they release old state tests for local school practice exams. The largest publishers of standardized tests currently are using their immense banks of multiple-choice items to develop — and sell — new formative tests or computerized collections of test items that are aligned to state standards. The idea is that these will be purchased for more frequent local formative testing.



"He's being punished for his annoying behavior on the field trip."

Those who adopt this practice see the benefit of using summative assessments in formative ways. They can identify state standards not yet being mastered by examinees early enough to permit teachers to make adjustments to promote greater success for their students. Similarly, they can identify students not progressing appropriately and can bring support services to bear. These are potent arguments in favor of this approach.

Manage data more effectively. A second trend in the emergence of formative assessment arises from the belief among some that the key to success resides not in the evidence gathered but in how that evidence is managed. Success in this camp is achieved by accumulating, summarizing, analyzing, and reporting assessment results with maximum efficiency. The more data-based the instructional decisions, advocates contend and research shows, the more effective will be instruction.

So local school districts and commercial software developers create and offer computer-based and online warehouse and management systems for handling student test scores. These systems are designed to deliver timely evidence of student progress into the hands of the right instructional decision makers. Typically, the assessment method used is multiple-choice tests generated from items in computerized banks that are aligned to standards. The result is a score that reflects student mastery of those standards. In this case, the intended users are teachers and school leaders working in teams to examine test-score trends, identify gaps in student learning, and translate test results into conclusions about program improvements.

Assessment FOR learning. A third approach to formative assessment contends that access to more frequent evidence of student mastery of state standards gathered using multiple-choice tests and placed in the hands of teachers, while potentially helpful, falls short of tapping the immense potential of formative thinking. The alternative is to use many different assessment methods to provide students, teachers, and parents with a continuing stream of evidence of student progress in mastering the knowledge and skills that underpin or lead up to state standards. This option has been labeled *assessment FOR learning.*

In this approach, students learn about achievement expectations from the beginning of the learning by studying models of strong and weak work. And they don't merely learn about the standards. Rather, they come to see and understand the scaffolding they will be climbing as they approach those standards. Students partner with their teacher to continuously monitor their current level of attainment in relation to agreed-upon expectations so they can set goals for what to learn next and thus play a role in managing their own progress. Students play a special role in communicating evidence of learning to one another, to their teacher, and to their families, and they do so not just after the learning has been completed but all along the journey to success. In short, during the learning, students are inside the assess-

ment process, watching themselves grow, feeling in control of their success, and believing that continued success is within reach if they keep trying.

When consistently carried out as a matter of routine within and across classrooms, this set of practices has been linked to achievement gains of one-half to two standard deviations on high-stakes tests, and the largest gains made are by low achievers.[2]

The most important difference between the first two formative assessment approaches described above and assessment FOR learning is that the former intend to inform the teachers about student achievement, while the latter also wants to inform students about their own learning. Assessment FOR learning rests on the understanding that students are data-based instructional decision makers too, a perspective all but ignored in our assessment legacy and in previous approaches to school improvement.

Another difference is that traditional formative thinking tends to want more frequent assessment of student mastery of the standards themselves, while assessment FOR learning focuses on day-to-day progress in learning as students climb the curricular scaffolding leading up to state standards. It tells users if and when students are attaining the foundations of knowledge, the reasoning, the performance skills, and the product development capabilities that underpin the mastery of essential standards.

In short, student success does not hinge merely on testing more frequently, on what teachers and principals do with the results, or on how efficiently the data are managed, although these things can contribute to student success. Rather, success also rests, at least in part, on what students do with and about those results. The actions students take — and therefore their ultimate success at learning — are determined by their emotional reaction to the assessment results. That response can be optimistic or pessimistic. An optimistic response leaves learners ready to keep trying and knowing what to do next: students maintain their desire to achieve and press on. A pessimistic response leaves learners feeling that the target remains beyond reach: students stop trying.

When used effectively, assessment FOR learning always triggers an optimistic response to assessment results from within the learner. It starts by providing students with a clear, student-friendly vision of the achievement target to be mastered, including models of strong and weak work. These examples reveal to learners where we want them to end up. Then the teacher provides learners with continuous access to descriptive feedback, which consists not merely of grades or scores but also of focused guidance specific to the learning target. Thus a foundation is laid for students to learn to self-assess and set goals. In this way, assessment FOR learning keeps students posted on where they are in relation to where they want be. By teaching students how to improve the quality of their work one dimension at a time and teaching them to monitor their own improvement over time, assessment FOR learning helps them close the gap between where they are now and where we want them to be.

But to use assessment productively to help achieve maximum student success, certain conditions need to be satisfied. Our achievement targets need to be clear. State standards need to be deconstructed into curriculum maps that are articulated within and across grade levels, and the resulting classroom-level achievement targets must be translated into student- and family-friendly versions. Furthermore, assessment and information management systems must be created to meet the needs of all instructional decision makers, including students. All assessments — especially those created by classroom teachers — must be accurate, producing dependable evidence of learning in all contexts at all times. The timing and nature of student involvement in assessment, record-keeping, and communication must be effectively managed by teachers.

Obviously, this list includes assessment responsibilities that are different from what has been expected of teachers in the past. Very few teachers have been given the opportunity to learn to apply the principles of assessment FOR learning. But with proper professional development and support from school leaders, teachers can be provided with the opportunity to use the classroom assessment process and its results in ways that honor their professionalism and promote maximum student success.[3]

---

1. Michael Scriven, "The Methodology of Evaluation," in Ralph W. Tyler et al., eds., *Perspectives in Evaluation*, American Educational Research Association Monograph Series on Curriculum Evaluation, No. 1 (Chicago, Ill.: Rand McNally, 1967), pp. 39-83; and Benjamin S. Bloom, J. Thomas Hastings, and George F. Madaus, *Handbook of Formative and Summative Evaluation of Student Learning* (New York: McGraw-Hill, 1971).

2. Benjamin S. Bloom, "The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," *Educational Leadership*, May 1984, pp. 4-17; Paul Black and Dylan Wiliam, "Assessment and Classroom Learning," *Educational Assessment: Principles, Policy, and Practice*, vol. 5, 1998, pp. 7-74; idem, "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan*, October 1998, pp. 139-48; Samuel Meisels et al., "Creating a System of Accountability: The Impact of Instructional Assessment on Elementary Children's Achievement Scores," *Educational Policy Analysis Archives*, vol. 11, 2003, http://epaa.asu.edu/epaa/v11n9; and Michael C. Rodriguez, "The Role of Classroom Assessment in Student Performance on TIMSS," *Applied Measurement in Education*, vol. 17, 2004, pp. 1-24.

3. Richard J. Stiggins et al., *Classroom Assessment for Student Learning: Doing It Right — Using It Well* (Portland, Ore.: Assessment Training Institute, 2004). K