

COMMENTARY

A Better Way to Assess Students and Evaluate Schools

By Monty Neill

Most Americans agree: We need a better way to assess students and evaluate schools. The latest Phi Delta Kappa/Gallup poll found that only one out of four respondents thought the No Child Left Behind law, the current version of the Elementary and Secondary Education Act, had helped schools in their community. Even U.S. Rep. George Miller, D-Calif., an original sponsor of that legislation and the chairman of the House Education and Labor Committee, agrees that NCLB may now be, as he put it, "the most negative brand" in the country.

As state testing intensified under the law and punitive sanctions were imposed, score gains on the National Assessment of Educational Progress slowed or halted for reading and math at all grade levels for almost all groups. Gap closing among demographic groups likewise slowed or stopped. Too much standardized testing damaged learning, particularly for the nation's neediest children. The test-and-punish approach distracted attention from more valuable reforms.

Yet, the underlying problems that propelled passage of NCLB remain. The nation still needs rational and effective approaches to school improvement, including strong curricula, skilled teaching, and equitable opportunities to learn. Society must address the consequences of poverty that undermine learning. Accountability systems and assessments should support high-quality teaching and learning.

Assessment functionally defines what we value in learning. As the old saying goes, "What you test is what you teach." With curriculum and instruction, it is a necessary component of the learning process. Assessment and evaluation inform the community about attainment of goals, including and those beyond academic outcomes. They signal problems that must be addressed and provide information on how to improve.

A healthy assessment and evaluation system would include three key components: limited large-scale standardized testing; extensive school-based evidence of learning; and a school-quality-review process.

- Large-scale tests. When it comes to assessment, the United States is an international outlier. As Stanford University's Linda Darling-Hammond has shown, many nations with better and more equitable educational outcomes test far less than we do. They typically test just one to three times before high school graduation, and use multiple-choice questions sparingly, if at all. Excessive testing wastes educational resources and fosters the use of cheap, low-level tests, while adding high stakes narrows and dumbs down the curriculum. The results provide little instructional value to students, teachers, schools, or districts.

Higher-quality tests would help. But based on the U.S. Department of Education's published criteria for awarding the \$350 million it will give to state consortia for test development, only modest improvements are likely to come from that program, far less than the qualitative leap schools need. Tests will continue to be administered too frequently.

Congress should return to the requirements of the 1994 version of the ESEA to test once each in elementary, middle, and high school. This would bring the United States in line with other nations, while freeing up resources for new assessment and evaluation approaches.

- Local and classroom evidence of learning. The primary public source of data about student achievement should be the work students do in the classroom. That kind of evidence reveals the range, depth, and quality of student learning. The United States has avoided taking this path, however, trekking instead through the wastelands of high-stakes standardized testing. This is largely because authorities have distrusted and not been willing to invest in teachers, unlike more successful nations, such as Finland. The pending ESEA reauthorization brings with it the chance to change direction and avoid another lost decade.

Classroom-based assessment by skilled teachers is of great value. Teachers assess frequently, but research shows that many have limited assessment skills. Thus, they need ongoing training to develop their assessment capabilities. In places as disparate as Nebraska, with its former STARS program of local, state-approved assessments, and New York state, where the New York Performance Standards Consortium replaces state tests with a mix of school- and consortium-based performance assessments, attention to assessment has been contributing to improved teaching, forging a stronger community of

educators, and producing improved results by a variety of measures from independent exams to college enrollment and success.

Classroom-based assessments can be adapted to students' varying needs while maintaining high standards. Assessing extended work, such as research projects, far more readily ensures evaluation of higher-order thinking skills than can large-scale standardized exams.

Of course, teachers cannot create every high-quality assessment they need. States should gather tasks that have been approved by skilled educators into "libraries" which teachers can access as they need. Using already-vetted instruments will contribute to ensuring the quality of classroom-based evidence of student learning.

In this country and around the world, a wide range of classroom- and school-based evidence—from exams, projects, "learning records," and portfolios—is audited and moderated. Essentially, a random sample from each classroom is rescored by trained readers to verify a teacher's initial scoring. This produces useful feedback to the originating teacher, score adjustments where needed, and professional development for the readers. Research in other nations and in this country shows that this process can be done with a degree of consistency more than sufficient for statewide comparability. What is standardized is not individual student work but the criteria for gathering and evaluating work products.

Schools would produce an annual report, including evidence of educational successes and ongoing problems, along with improvement plans. Documentation of student learning across the curriculum would then become publicly available. Such reports could be discussed by the school's community and reviewed by higher governmental authorities.

- School-quality reviews. Often called "inspectorates," these are the central tool for school evaluation in places such as England (which tests at a few grades), Wales (which tests only at grade 5, with no stakes), and New Zealand (which has only a NAEP-like national exam). Clearly, this is a very different mind-set: Instead of test results, the core of evaluation is a comprehensive review every four to five years covering the range of attributes parents and communities want for their schools. School-quality reviews have been proposed by the politically diverse signers of the Broader, Bolder Agenda. In the United States, these quality reviews would be complemented by limited large-scale testing and annual school reports, providing comprehensive evidence in which each component serves as a check on the others.

During inspections, skilled professionals, perhaps accompanied by parents and community members, conduct three- to five-day visits. The teams come prepared with other data (assessment results, graduation rates, school-climate surveys, opportunity-to-learn information, and so forth). They sit in on classes, review student work, and interview students, teachers, and other staff members. They prepare a draft report and discuss it with school personnel. The final report is a public document that includes an evaluation and recommendations for improvement. This approach is similar to college and school accreditation processes.

Schools with severe problems would be reviewed more frequently. States could specify how and when recommendations become mandates, some of which could require new resources, outside assistance, or strong interventions.

Since nations using a more balanced, comprehensive, improvement-focused assessment and evaluation system have produced better educational results with fewer harmful side effects, it makes good sense to restructure the current test-based U.S. system. The model outlined here can provide better assessment, comparability, and accountability. These improvements are needed by all schools, especially those which primarily serve low-income children.

Comprehensive data analysis can identify educational problems and solutions. Equitably distributed resources, strong collaborative leaders, professional learning communities of teachers, rich and challenging curriculum, strong parental involvement, and safe learning environments are also necessary.

But without healthy assessment and evaluation, the reform enterprise will fail again.

*Monty Neill is the interim executive director of **FairTest**, the National Center for Fair & Open Testing, in Boston. FairTest developed this proposal with help from allies, particularly the Massachusetts Coalition for Authentic Reform in Education.*

Vol. 29, Issue 36

<http://www.edweek.org/ew/articles/2010/06/18/36neill.h29.html>