

FairTest

National Center for Fair & Open Testing

The Value of Formative Assessment

The current wave of test-based "accountability" makes it seem as though all assessment could be reduced to "tough tests" attached to high stakes. The assumption, fundamentally unproven, is that such tests produce real improvements in student learning better than do other educational methods.

In this environment, Paul Black and Dylan Wiliam's "Inside the Black Box: Raising Standards Through Classroom Assessment" (Phi Delta Kappan, October 1998) provides strong evidence from an extensive literature review to show that classroom "formative" assessment, properly implemented, is a powerful means to improve student learning -- but summative assessments such as standardized exams can have a harmful effect.

Summative assessment is the attempt to summarize student learning at some point in time, say the end of a course. Most standardized tests are summative. They are not designed to provide the immediate, contextualized feedback useful for helping teacher and student during the learning process. High quality summative information can, of course, shape how teachers organize their courses or what schools offer their students.

By contrast, formative assessment occurs when teachers feed information back to students in ways that enable the student to learn better, or when students can engage in a similar, self-reflective process. If the primary purpose of assessment is to support high-quality learning (principle one in Principles and Indicators for Student Assessment Systems), then formative assessment ought to be understood as the most important assessment practice.

The evidence shows that high quality formative assessment does have a powerful impact on student learning. Black and Wiliam report that studies of formative assessment show an effect size on standardized tests of between 0.4 and 0.7, which is larger than most known educational interventions. (The effect size is the ratio of the average improvement in test scores in the innovation to the range of scores of typical groups of pupils on the same tests; Black and William recognize that standardized tests are very limited measures of learning.) Formative assessment is particularly effective for students who have not done well in school, thus narrowing the gap between low and high achievers while raising overall achievement.

Unfortunately, the research shows that high-quality formative assessment is relatively rare in classrooms, and that most teachers do not know well how to engage in such assessment. Black and Wiliam also found that most classroom testing encourages rote and superficial learning. Teachers do not help each other become good assessors, and they often emphasize quantity of work over high quality. Actual assessment practices are often harmful: marking and grading are overemphasized while giving useful advice is underemphasized, and comparing students competitively causes low-achieving students to believe they cannot learn. Teachers, it turns out, generally replicate standardized tests in their own assessment practices and therefore lack sufficient information about their students.

In sum, the state of formative assessment is quite weak. The massive emphasis on external, standardized exams, often with high stakes attached as with the federal No Child Left Behind law, has intensified the domination of summative tests over curriculum and instruction -- even though the

research examined by Black and William supports the conclusion that summative assessments tend to have a negative effect on student learning. Meanwhile, there has been some lip service but little real attention to strengthening teachers' formative, classroom assessment capabilities. Worse, there has recently been a proliferation of so-called "benchmark" tests that are claimed to be formative but are no more than mini-standardized tests intended to predict how well the test-taker will do on the end-of-year big test.

Changing Assessment

Black and Wiliam offer recommendations for changing this state of affairs. They begin with the student, "the ultimate user" of that assessment information which is geared toward improving learning. To be truly helpful to students, formative information should be focused on the task, not the student; and the student must understand the feedback so as to make use of it. To really succeed, however, students must learn to self-assess " so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve."

Black and Wiliam discuss what makes for effective feedback from teachers, such as opportunity for students to express their understanding, classroom dialogue that focuses on exploring understanding, and feedback which includes opportunities to improve and guidance on how to improve. They conclude with policy prescriptions, starting with the need to change from a focus on standardized tests to what is still the "black box" of actual student learning and to changing classroom practices, particularly formative assessment.

"Inside the Black Box" is a must read for educators and policymakers. It is a strong and clear summary of the vast array of evidence that Black and Wiliam review in the journal *Assessment in Education* (March 1998). The latter publication also includes commentaries on the Black and Wiliam review. In one, Phillippe Perrenoud introduces concepts and research evidence from the French. He argues that schools fail students mostly by failing to respond to their differences. He calls for "diversity in approach at the service of equality" and summarizes with a simple equation: "diversity in people + appropriate treatment for each = diversity in approach." As with Black and Wiliam, the purpose of diversity is not negatively discriminatory tracking and the denial of opportunity, but using classroom assessment integrated with high quality curriculum and instruction to ensure that all students reach high levels of important learning.

The articles in *Assessment in Education* are often quite complex, but should be of great value to researchers and to educators working to improve classroom assessment practices. Taken together, the Kappan article and the pieces in *Assessment* make a powerful contribution to the knowledge base for assessment reform and a powerful rebuttal to the now-dominant political forces which reduce assessment to "tough" tests.

- "Inside the Black Box" can be found on the Kappan web site at <http://www.pdkintl.org/kappan/kbla9810.htm>.
- *Assessment in Education*, Vol. 5, No. 1, can be obtained from Carfax Publishing, 875-81 Mass. Ave., Cambridge, MA 02139; 800-354-1420; their web site is <http://www.carfax.co.uk>.
- Other important work on formative assessment includes material from the British Assessment Reform Group at <http://arg.educ.cam.ac.uk/> (especially the publication, *Assessment for Learning*) and publications by Richard Stiggins.

Joint Organizational Statement on *No Child Left Behind (NCLB) Act*

The undersigned education, civil rights, religious, children's, disability, and civic organizations are committed to the No Child Left Behind Act's objectives of strong academic achievement for all children and closing the achievement gap. We believe that the federal government has a critical role to play in attaining these goals. We endorse the use of an accountability system that helps ensure all children, including children of color, from low-income families, with disabilities, and of limited English proficiency, are prepared to be successful, participating members of our democracy.

While we all have different positions on various aspects of the law, based on concerns raised during the implementation of NCLB, we believe the following significant, constructive corrections are among those necessary to make the Act fair and effective. Among these concerns are: over-emphasizing standardized testing, narrowing curriculum and instruction to focus on test preparation rather than richer academic learning; over-identifying schools in need of improvement; using sanctions that do not help improve schools; inappropriately excluding low-scoring children in order to boost test results; and inadequate funding. *Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement.*

Recommended Changes in NCLB

Progress Measurement

1. Replace the law's arbitrary proficiency targets with ambitious achievement targets based on rates of success actually achieved by the most effective public schools.
2. Allow states to measure progress by using students' growth in achievement as well as their performance in relation to pre-determined levels of academic proficiency.
3. Ensure that states and school districts regularly report to the government and the public their progress in implementing systemic changes to enhance educator, family, and community capacity to improve student learning.
4. Provide a comprehensive picture of students' and schools' performance by moving from an overwhelming reliance on standardized tests to using multiple indicators of student achievement in addition to these tests.
5. Fund research and development of more effective accountability systems that better meet the goal of high academic achievement for all children

Assessments

6. Help states develop assessment systems that include district and school-based measures in order to provide better, more timely information about student learning.

7. Strengthen enforcement of NCLB provisions requiring that assessments must:
 - Be aligned with state content and achievement standards;
 - Be used for purposes for which they are valid and reliable;
 - Be consistent with nationally recognized professional and technical standards;
 - Be of adequate technical quality for each purpose required under the Act;
 - Provide multiple, up-to-date measures of student performance including measures that assess higher order thinking skills and understanding; and
 - Provide useful diagnostic information to improve teaching and learning.
8. Decrease the testing burden on states, schools and districts by allowing states to assess students annually in selected grades in elementary, middle schools, and high schools.

Building Capacity

9. Ensure changes in teacher and administrator preparation and continuing professional development that research evidence and experience indicate improve educational quality and student achievement.
10. Enhance state and local capacity to effectively implement the comprehensive changes required to increase the knowledge and skills of administrators, teachers, families, and communities to support high student achievement.

Sanctions

11. Ensure that improvement plans are allowed sufficient time to take hold before applying sanctions; sanctions should not be applied if they undermine existing effective reform efforts.
12. Replace sanctions that do not have a consistent record of success with interventions that enable schools to make changes that result in improved student achievement.

Funding

13. Raise authorized levels of NCLB funding to cover a substantial percentage of the costs that states and districts will incur to carry out these recommendations, and fully fund the law at those levels without reducing expenditures for other education programs.
14. Fully fund Title I to ensure that 100 percent of eligible children are served.

We, the undersigned, will work for the adoption of these recommendations as central structural changes needed to NCLB at the same time that we advance our individual organization's proposals.

Signed by 149 organizations as of November 12, 2008, including Council for Exceptional Children, Learning Disabilities Assn., NAACP, Urban League, League of United Latin American Citizens, Childrens Defense Fund, Parent Teacher Assn., National Council of Churches, National Education Assn., American Federation of Teachers, National School Boards Assn., American Assn. of School Administrators, Public Education Network, and FairTest. A full list is at www.fairtest.org, or Forum on Educational Accountability, www.edaccountability.org.

FairTest

National Center for Fair & Open Testing

The Limits of Standardized Tests for Diagnosing and Assisting Student Learning

Standardized tests have historically been used as measures of how students compare with each other (norm-referenced) or how much of a particular curriculum they have learned (criterion-referenced). Increasingly, standardized tests are being used to make major decisions about students, such as grade promotion or high school graduation, and schools. More and more often, they also are intended to shape curriculum and instruction.

Proponents of the expanded uses and consequences of tests claim that newer exams are superior to the flawed exams of the past, measure what is important, and are worth teaching to. These arguments ignore the real-world limits to what standardized tests can usefully do. Repeating such false claims perpetuates test misuse and the dangerous belief that what is worth teaching is that which can be assessed by a standardized test.

Under a new federal law, state assessments of reading and math must be administered for accountability annually in grades 3-8 and once in high schools. The assessments must be based on state content and performance standards; measure higher order thinking; provide useful diagnostic information; and be valid and reliable. While the law does not mandate the use of standardized tests, many states will be inclined administer them to meet the federal law. An examination of each requirement, however, reveals the limits of standardized tests.

Tests are to be based on state standards

State standards are often too long and detailed to ever be taught. Many fail to distinguish what is important from what is unimportant or to separate what all students ought to learn in a subject from what only the most interested might learn. In part because of the level of detail, much of the content in state standards is not assessed by state tests.

Moreover, much of value in state standards cannot be tested with any paper-and-pencil test of a few hours duration. In a high quality education, students conduct science experiments, solve real-world math problems, write research papers, read novels and stories and analyze them, make oral presentations, evaluate and synthesize information from a variety of fields, and apply their learning to new and ill-defined situations. Standardized tests are poor tools for evaluating these important kinds of learning. If instruction focuses on the test, students will not learn these skills, which are needed for success in college and often in life.

Measure higher-order thinking

Standardized exams offer few opportunities to display the attributes of higher-order thinking, such as analysis, synthesis, evaluation, and creativity. Higher order thinking is encouraged and revealed by in-depth and extended work, not by one-shot tests.

Provide useful diagnostic information

Assessments of educational strengths and weaknesses can be useful at the individual, classroom, school or district levels. However, information needs to be sufficiently timely, accurate, meaningful, detailed and comprehensive for the kind of diagnosis being made. The

lengthy turn-around time for scoring most standardized tests makes them nearly useless for helping a particular individual, though the information might be of some value to teachers and schools for longer-range planning.

In addition, standardized tests usually include only a few questions on any particular topic. This is too little information to produce accurate, comprehensive or detailed results. Many topics in state standards are not addressed at all in state exams, so the tests provide no diagnostic information about them.

Diagnosis suggests the use of “formative” assessment – assessments that can help a teacher and student know what to do next. Standardized tests administered at the end of the year – “summative assessment” – cannot possibly meet this need. Sound diagnostic practices also include understanding why a student is having difficulty or success and determining appropriate action. As snapshots with limited information, standardized tests provide neither an answer to “why” nor little guidance for successful instruction.

Be valid and reliable

Test validity, experts explain, resides in the inferences drawn from assessment results and the consequences of their uses. Relying solely on scores from one test to determine success or progress in broad areas such as reading or math is likely to lead to incorrect inferences and then to actions that are ineffective or even harmful. For these and other reasons, the standards of the testing profession call for using multiple measures for informing major decisions – as does the ESEA legislation.

Reliability, or consistency of information, is sometimes treated as the most important aspect of testing. However, consistent information about too narrow a range of topics, skills or knowledge cannot provide adequate information for credible decisions: a doctor needs more than just reliable blood pressure results to treat a patient. Well-designed classroom-based assessments can provide richer, consistent information that enhances validity, diagnostic capacity, and the ability to assess progress toward meaningful standards.

Conclusion

When standardized tests are the primary factor in accountability, the temptation is to use the tests to define curriculum and focus instruction. What is not tested is not taught, and what is taught does not include higher-order learning. How the subject is tested becomes a model for how to teach the subject. At the extreme, school becomes a test prep program – and this extreme already exists.

It is of course possible to use a standardized test and not let its limits control curriculum and instruction. However, this can result in a school putting itself at risk for producing lower test scores. It also means parents and the community are not informed systematically about the non-tested areas, unless the school or district makes a great effort.

To improve learning and provide meaningful accountability, schools and districts cannot rely solely on standardized tests. The inherent limits of the instruments allow them only to generate information that is inadequate in both breadth and depth. Thus, states, districts and schools must find ways to strengthen classroom assessments and to use the information that comes from these richer measures to inform the public.