The author outlines the limitations of high-stakes standardized tests and discusses the impact of such tests on grade retention and the narrowing of curriculum content and instruction. He argues that high-stakes testing is a failed school-improvement strategy and suggests an alternative accountability approach that would use multiple forms of formative as well as summative assessments.

Monty Neill

Many principals and teachers have concluded that high-stakes testing, particularly that mandated by the No Child Left Behind (NCLB) Act, is doing grave damage to education and the lives of children. Parents and other community members likewise worry about the consequences of schools focusing on test preparation. These educators and parents understand that assessment is a necessary and valuable part of teaching, and they do not object to accountability. But they oppose accountability that is reduced to test scores and punishments. Their opposition is based on two key points: the limited nature of standardized tests and the consequences of attaching high stakes to them.

IN BRIEF

The author outlines the limitations of high-stakes standardized tests and discusses the impact of such tests on grade retention and the narrowing of curriculum content and instruction. He argues that high-stakes testing is a failed school-improvement strategy and suggests an alternative accountability approach that would use multiple forms of formative as well as summative assessments.
In the United States today, the dominant mode of assessment is government-mandated standardized tests that are primarily composed of multiple-choice questions. Although such high-stakes tests can provide some useful accountability data, they are far too narrow to provide sufficient information for any fair or helpful educational decisions. In fact, their use can have harmful consequences for curriculum, instruction, classroom assessment, and student learning. What is needed are high-quality assessments that can make a powerful contribution to student achievement and provide useful information for school improvement when combined with other information, such as grade-progression rates, teacher quality, and school climate.

**The Limits of Testing**

State standards are too long and detailed to ever be completely taught. Many fail to distinguish what is important from what is unimportant, and much of what is important cannot be tested with a paper-and-pencil test of a few hours duration. In the best elementary and middle school education, students read, talk, and write about real stories. They conduct science experiments, write papers, make oral presentations, prepare computer-based presentations, evaluate and synthesize information from a variety of fields, and apply their learning to new situations. Standardized tests are poor tools for evaluating these important kinds of learning. If instruction focuses on the test, students have few opportunities to display the attributes of higher-order thinking, such as analysis, synthesis, evaluation, and creativity, which are needed for success in school, college, and life.

Assessments of educational strengths and weaknesses can be useful at the individual, classroom, school, or district levels to shape teaching and improve schooling. However, that information must be sufficiently timely, accurate, meaningful, detailed, and comprehensive. The lengthy turnaround time for scoring most standardized tests makes them nearly useless for helping a particular individual. In addition, standardized tests usually include only a few questions on any particular topic, not enough to produce accurate, comprehensive, or detailed results. Many topics in state standards are not addressed at all in testing.

Test validity 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 result in incorrect inferences, which in turn can lead 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.

Reliability, or consistency of information, is sometimes regarded as the most important aspect of testing. However, consistent information about too narrow a range of topics, skills, or knowledge cannot provide an adequate basis for credible decisions. A doctor, for example, needs more than just reliable blood pressure results to diagnose a patient. Well-designed classroom-based assessments can provide richer, more consistent information that enhances validity, diagnostic capacity, and the ability to assess progress toward meaningful standards. Researchers have concluded that such formative assessment can have a powerful impact on student achievement, especially for low-achieving students (Black and Wiliam 1998). Unfortunately, too few teachers know how to use such assessment effectively.

In response to the pressure to raise test scores, districts are increasingly relying on short, standardized multiple-choice tests, often labeled “benchmark tests,” that are intended to predict how well a student will do on high-stakes exams. These tests also fail to adequately assess the range of important student learning.

**High-Stakes Consequences**

High-stakes testing leads to increased grade retention, which has repeatedly been proved to be counterproductive. Students who are retained do not improve academically, are emotionally damaged by retention, lose self-esteem, and are more likely to drop out of school.

High-stakes testing also produces “teaching to the test.” The higher the stakes, the more schools focus instruction on the tests. Whole subjects, such as science, social studies, art, or physical education may be reduced or eliminated if only language arts and math are tested. Even in tested subjects, many important topics or skills that are not covered by the exam are not taught. Instruction starts to resemble the tests. For example, reading is reduced to short passages followed by multiple-choice questions, while writing becomes repeated practice in the “five-paragraph essay.” Benchmark testing further exacerbates these problems.

Such narrowing of curriculum and instruction happens mostly in classrooms serving low-income students. In schools serving more affluent communities, teachers and parents make sure most students gain the skills and knowledge they need to succeed in high school, college, and life. Too often, poor kids get little more than test coaching and scripted instruction that does not adequately prepare them for further learning.

Proponents of high-stakes testing have been reduced to touting marginal gains, such as the modest growth in math scores recorded by the independent National Assessment of Educational Progress, or claiming that it is...
“THE MORE THINGS CHANGE, THE MORE THEY STAY THE SAME.”

—ALPHONSE KARR

National Educational Service is proud to announce that we are now Solution Tree.
Yes, we’ve changed our name, but not what’s really important.

We are still the home of Professional Learning Communities at Work™, leading educational institutes and summits, and a world-class variety of on-site professional development workshops for educators and administrators. We continue to publish and represent some of the best and the brightest authors and practitioners in education: Rick DuFour, Bob Eaker, Becky DuFour, Larry Brendtro, Lee Canter, Anne Conzemius, Rick Curwin, Michael Fullan, Larry Lezotte, Cheri Lovre, Allen Mendler, Jan O’Neill, Mike Schmoker, Dennis Sparks, Rick Stiggins, and more. Visit us online at www.solution-tree.com to check out our array of offerings relating to school improvement and youth at risk.

Visit us online at www.solution-tree.com or call 800.733.6786 today to find out more.
too early to tell if high-stakes testing will work. It is not too early to tell: high-stakes testing is a failed school-improvement strategy. The Texas “miracle,” for example, was promoted as evidence the strategy worked. But claims for the Texas experience have been buried under charges of cheating, falsified dropout rates, and teaching to the test. Middle school teachers in that state have described students coming into their classrooms with high test scores but not able to read.

What Can Be Done?
If high-stakes testing does not lead to school improvement, can some other form of assessment work contribute to positive changes? A better strategy would be to build on the evidence for the success of formative assessments. When done well, formative assessments produce significant gains. However, such testing requires teachers to be skilled in using multiple forms of assessment. Unfortunately, colleges do not prepare prospective teachers well in this area, and professional development opportunities are limited. As a result, too few teachers are sufficiently literate in assessment practices. Meanwhile, the pressure to raise test scores pushes teachers to imitate the state exams or benchmark tests in their own classroom assessments.

Real education reform cannot be accomplished by handing teachers scripted curricula and commands to teach to the test. It requires a commitment to spend the time and money needed to ensure that all teachers are knowledgeable in assessment as well as curriculum content. This involves not only ongoing professional development, but also time for teachers to work together to improve curriculum, instruction, and assessment.

The use of summative assessments—standardized or classroom exams used to measure what students have learned to date—must also change. Rather than assuming that one test at the end of the year is an adequate basis for making decisions about students, summative assessments should be mixed with ongoing classroom-based evidence of student learning. Together, formative and summative assessments provide extensive information for teachers, principals, and district administrators to use in improving school performance.

Accountability should no longer be based on test scores alone. Rather, it should include classroom and school-based evidence of achievement, standardized test scores, and regular detailed investigations by trained observers. The Massachusetts Coalition for Authentic Reform in Education has detailed how such an approach can work.

To begin moving toward this vision of assessment and accountability, an alliance of more than 60 national education, civil rights, religious, and other groups, initiated and chaired by FairTest, has released the Joint Organizational Statement on No Child Left Behind that says, “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.” The alliance recognizes that these changes will not come simply, quickly, or cheaply. But they are necessary if our nation is to improve, not diminish, the quality of learning and ensure that all children have access to a high-quality education.

Web Resources
FairTest’s report Failing Our Children details the flaws in NCLB and high-stakes testing and provides more information on assessment alternatives. http://fairtest.org/Failing_Our_Children_Report.html

The Joint Organizational Statement on No Child Left Behind is available online. www.fairtest.org

The British Assessment Reform Group seeks to ensure that assessment policy and practice at all levels take account of relevant research evidence. Its Web site offers a list of publications as well as links for further information. http://arg.educ.cam.ac.uk

The Nebraska State Department of Education Web site includes informative material on accountability and a report on the state’s Classroom Assessment Conference in September 2005. www.nde.state.ne.us
ASSESSMENT ISN’T A FOUR LETTER WORD.

Help your teachers manage the increasing demands of assessment standards with Mastering Assessment, a practical, easy-to-use set of fifteen booklets on teaching and student evaluation. Written by the foremost authority on educational assessment, W. James Popham in his characteristic witty style, these short books put the most relevant information on assessment at teachers' fingertips. Reading through a booklet on reducing assessment bias, how to build constructed response tests, or scoring rubrics will provide teachers a quick, engaging means of becoming "assessment literate."

“I must say that I found Popham’s writing style refreshing. His use of humor and ‘colleague to colleague’ format made this booklet an easy read. Too often, in educational publications, the use of ‘jargon’ and technical data is overwhelming. The simple explanations of the testing jargon made me feel comfortable right away.”

-- Jerome Cammarata
Principal, North Boulevard School
Pompton Plains, NJ

Mastering Assessment
W. James Popham

1 Pack 15 Booklets 1 Solution
Appropriate and Inappropriate Tests for Evaluating Schools
Assessing Students’ Affect
Assessing Students with Disabilities
Assessment Bias: How To Banish It
Classroom Evidence of Successful Teaching
College Entrance Examinations: The SAT and the ACT
Constructed-Response Tests: Building and Bettering
How Testing Can Help Teaching

Interpreting the Results of Large-Scale Assessments
Portfolio Assessment and Performance Testing
Reliability: What Is It and Is It Necessary?
Selected-Response Tests: Building and Bettering
Test Preparation: Sensible or Sordid?
The Role of Rubrics in Testing and Teaching
Validity: Assessment's Cornerstone

For more information on Mastering Assessment please contact:
Routledge Education • 270 Madison Avenue, New York, NY 10016
colin.reublinger@taylorandfrancis.com • routledge-ny.com