



*From standards to assessment to results—what school leaders need to know in the NCLB era*

BY JOAN L. HERMAN AND RONALD DIETEL

# A Primer on Accountability

**E**stablish goals, make plans to reach them, measure progress, revise programs based on results. The principles of No Child Left Behind? Benjamin Bloom's mastery learning? Management by objectives?

Actually, this process—commonly known today as standards-based accountability—was outlined by Ralph Tyler in his classic 1949 book, *Basic Principles of Curriculum and Instruction*. No Child Left Behind adds some new twists, including major consequences for failing to increase test scores, but the components of standards-based accountability are well-known, having been integrated into virtually all state testing programs, as well forming the backbone of NCLB.

Research provides some fundamental knowledge that may help you as you grapple with the law's requirements. Here are some of the things we know:

**State assessments focus instruction.** Research and practical experience show that teachers and principals indeed pay attention to what is tested and adapt their curriculum and teaching accordingly.

We also know that prior to a state test, nearly all schools engage in direct test preparation and that they are increasingly analyzing test scores, then adapting instruction to help students. "Analyzing data [is] of little value if it doesn't lead to instructional change," said Rochester (N.H.) Superintendent Raymond

Yeagley in response to a research study being conducted by the National Center on Evaluation, Standards, and Student Testing. Like a growing number of districts, Rochester uses data to help identify districtwide weaknesses in addition to helping individual students.

**Teachers adjust their approach according to what is assessed.** When states developed performance assessments in the early 1990s, many classroom teachers revised their instruction and assessment accordingly. In fact, teachers scrambled to replace their own multiple-choice tests with the same types of open-ended items and extended writing questions that state tests had begun to use. As one teacher who was piloting a new science performance assessment said in a study, "If this is the way you are testing, then this is the way we are going to teach."

**Test scores show initial increases.** In state after state, when new assessments and accountability provisions are put into place, student scores show an increase, at least for the first few years. For example, California elementary school students showed a 12-point gain in the percentage of students scoring at or above the 50th percentile from 1998 to 2001, as second-graders progressed through the fifth grade.

**Schools focus on the test rather than the standards.** Although Washington state took great effort and many years to develop its state standards, it wasn't until state tests were introduced in the form of the Washington Assessment of Student

Learning and the Iowa Test of Basic Skills that teachers changed their instruction.

The pressure to perform well on state tests had some potentially undesirable curriculum effects in Washington and Kentucky. Researcher Brian Stecher found that when testing different topics in different years, teachers shifted their instruction to the tested subjects and away from others. In Kentucky, because math was tested in the fifth and eighth grades, it received more emphasis than subjects not tested in those grades, including reading, writing, and science. These changes were not motivated by any coherent sense of curriculum nor driven by the need to continuously develop students' learning.

**What is not tested becomes invisible.** Performance in subjects the state does not test, such as social studies and art, may languish unseen. Can you say as a board member that you know how students are doing in all subjects? Or do you mostly know about language arts and math? Even within a topic, such as math, important skills may be missed. If the state mathematics test focuses on basic computation, and not practical math applications, guess which will receive more classroom emphasis.

**State assessments show uneven alignment with standards.** True alignment has multiple requirements, including *relevance*, *breadth*, and *balance*. Relevance is the degree to which the assessment items seem to match some specific state standards. Most state tests do quite well by this criterion, but relevance by itself is insufficient. If relevance were all that mattered, a test would be considered aligned even if all the items on the test addressed a single standard.

Breadth is the extent to which all standards or benchmarks are measured by a test or set of assessments, and balance reflects the extent to which some standards or benchmarks may be prioritized over others in terms of relative emphasis or number of items.

Based on a review of standards and tests in a dozen states, the education organization Achieve said that it "found few examples of tests that truly measure the depth and breadth of what states expect all children to learn."

Moreover, it appears that existing tests tend to emphasize lower levels of knowledge and skills. What is tested seems at least as much a function of the particular items that writers are best at producing and those that survive field testing.

As Stephen Dunbar, a principal coauthor of the Iowa Test of Basic Skills, told a 2002 conference, "The unprecedented scope of testing under the No Child Left Behind Act will place heavy demands on test development and result in less field testing and perhaps a dwindling quality of the items."

**Test score increases may not signal real increases in learning.** In the first three years of the current California accountability system, elementary schools showed the greatest increase in the first year, less in the second, and less still in the third. Scores at middle and high school levels tended to level out even sooner. Nearly all state tests show the same pattern, strong evidence that most score increases are due more to

good test preparation than to true improvements in learning. Researcher Robert Linn has found that when students are given a new version of the same test, scores plummet.

Similarly, external assessments seldom confirm state test score increases. Harvard education professor Dan Koretz and his colleagues found that Kentucky fourth-graders showed nearly four times the growth on Kentucky's state math assessment over the period 1992 to 1996 than was measured by an external test, the National Assessment of Educational Progress. Board members are right to question large gains on any single test.

**The reliability of school score changes from year to year is uncertain.** Like surveys, tests have a margin of error. Small sample sizes—say, fourth-grade special education students at a single school—will have more error than the same group at the state level. If your total population size is just 10 students, a change in a single student's score may make a substantial difference. Teacher mobility may also contribute to test score changes beyond a school or district's control.

**NCLB's goals may be unreachable.** Having *all* students proficient by 2014 represents a gargantuan challenge. Two of the groups facing large hurdles are English Language Learners, who must learn English in addition to their regular course content, and students with disabilities who have substantial learning problems. In Massachusetts, for example, no measurable improvement was made from 2002 to 2004 in the language arts scores of 10th-grade ELL students; over seven years, the increase was just 6 percentage points.

## WHAT YOUR BOARD CAN DO

We may all agree on the underlying goals of leaving no child behind, but the research shows that current accountability systems, including NCLB, are imperfect mechanisms for achieving the intended goals. While some aspects of the system are working as intended to stimulate achievement and accurately monitor progress, others need improvement.

Here are some ways you and your board can help make accountability work better:

**Support alignment of standards, assessments, and instruction.** Focusing on tests rather than standards may mean that students learn only a relatively narrow, test-based curriculum. Doing well on the state test doesn't mean students will do well on other tests, which may measure different skills. Mismatches between tests and standards can lead educators and policymakers to misinterpret rising test scores and fail to address genuine needs.

Work collaboratively with your district to support goals and instruction that give students adequate content coverage and the deep learning skills that will help them perform well on any test. Be cautious of the decisions you make based on single test scores.

**Support a complementary district assessment program.** The multiple-choice and short-answer questions that predominate on state tests can only go so far in tapping the complex thinking, communication, and problem-solving skills students

need for future success. We encourage district assessments that include other types of performance measures—essays, applied projects, portfolios, or demonstrations—as a complement to state tests. Multiple types of measures can better respond to individual differences than can a single test. Just as not all students learn in the same way, not all students can demonstrate their proficiency in the same way. Some may do better in some formats and contexts than in others.

**Develop good classroom assessment and data management.** District test results that come just once or twice a year are often too late to make improvements in student learning. Well-designed classroom assessment, aligned with standards and instruction, can give teachers ongoing information to monitor specific student needs, focus scarce resources, and improve student learning. A wide variety of data management software is now available to track student progress, integrating results from state, district, and classroom assessments.

**Empower educators to improve teaching and learning.** While accountability is a top-down policy strategy to promote improvements in student learning, it will not work in the absence of talent and resources at the local level. Both new and veteran teachers need the capacity to improve teaching and assessment, including high-quality professional development, time, and leadership support. Good teaching is an intense problem-solving endeavor. It requires sophisticated content knowledge and pedagogical finesse, including sensitivity to students' needs and motivations, to take students from where they are to where they need to be.

**Pay attention to social as well as academic capital.** Social capital is the glue that holds a community together. It creates a sense of mutual obligation and a network of support for reaching goals. As we plan to help students achieve high academic standards, we must do so in ways that develop their social capital as well. Research at the middle school and high school levels shows the importance of fostering a sense of connection and commitment to schooling.

**Learn from success.** While the NCLB goals are ambitious, some schools are well on their way to achieving them. Leaders at these schools have an impressive “can-do” attitude and an ability to inspire their staffs. They constantly ask, “What should we be doing differently in teaching and learning?” And then they try to do it, while carefully monitoring the success of their strategies.

“Whenever you tell me that 25 percent of your kids are going to fail, that is the most awful thing I’ve ever heard,” a Kentucky principal told researchers Hilda Borko and Monette McIver. “You cannot call yourself a good educator and lose 25 percent. If you go to your surgeon and he says 25 percent of his patients died, you don’t want to go back ... I expect you to use the same intensity as that show on TV, *ER*. When they get [patients] in that emergency room, they don’t care how much money they have. They don’t care what ethnic background they have. They blitz them. And they use every strategy. They consult with each other. They do every possible thing to save that patient, and

that is your job here, as teachers. If you lose one of them ... then part of you goes with it.”

### A THREE-PART AGENDA

So how does standards-based accountability measure up? Is it meeting its dual role of motivating improvement while providing accurate accountability information?

Standards-based accountability, like Ralph Tyler’s instructional principles of 1949, has helped states focus their accountability programs and stimulate achievement and teaching, albeit with serious alignment and instructional imperfections. The accuracy of the information about school quality and student improvement remains in some doubt, hampered by the use of 50 different accountability systems and questions over whether the gains are true or not.

This is certainly not unique to standards-based accountability, but rather to our decentralized educational system. Regardless, the key question is: Can standards-based accountability, as embodied in NCLB, fulfill its dual roles? We believe three actions are crucial from school boards and districts.

First, develop the best possible assessment programs in your school district, aligning standards and state tests with complementary district assessments that measure deep knowledge and the ability of students to apply skills across content areas. Provide resources for the development of high-quality classroom assessments, aligned to state standards and textbooks. Support the system all along the way with professional development, time, and—most important—leadership. Realize that a high-quality system needs monitoring, feedback, and continual improvement.

Second, continue to push policymakers at state and federal levels for flexibility in current NCLB requirements. Unless the requirements are changed, virtually all schools, districts, and states will fail to make adequate yearly progress in coming years. And if even high-performing schools fail, the public will lose confidence in NCLB as a valid indicator of school quality.

Finally, develop a school board and district commitment to the same can-do spirit of another Kentucky principal whose children are beating the odds:

“...This is our home. These are all our kids. We love this school. We have a school spirit here that cannot be equal to any other, anywhere else. We’re not perfect. ... But we try all the time, each and every day, to keep growing and keep learning. And in the face of any type of obstacle, we’ve always stuck together. We’ve laughed together. We’ve cried together. But the bottom line is, we love these kids, we love this school, and there is nothing we won’t do to make it a success.”

---

Joan L. Herman (joan@cse.ucla.edu) is codirector of the National Center on Evaluation, Standards, and Student Testing (CRESST) at the UCLA Graduate School of Education & Information Studies.

Ronald Dietel (ron@ucla.edu) is CRESST’s assistant director for research use and communications and former president of the La Cañada (Calif.) Unified School District Governing Board.