4

Assessing Learning

Herbert J. Walberg

Along with high standards, accountability, and choice, improved assessment including testing is one the most effective and cost-effective investments to improve achievement, particularly for minority students. Contrary to widespread beliefs, rigorous assessment does not reduce promotion and graduation rates, and clear goals and assessment promote professional morale among educators. Data from twenty-five states show assessment costs averaging only about $20 per student, or a miniscule 0.3 percent of annual per student spending. For these reasons, the assessment reforms deserve attention.

Why is assessment so effective and cost-effective? Under the federal No Child Left Behind Act as well as the state's own accountability system, it provides valuable "high stakes" information for determining which schools should be on probation, which schools need interventions, and whose students can transfer to more successful schools. In the case of end-of-course tests for course credits, it provides incentives for both students and teachers. It is the central indicator for teacher pay incentives. Given

frequently, assessment tests provide ongoing diagnostic information to educators on students' strengths and weaknesses, which is the basis for efficient teaching and re-teaching.

Contrary to common belief, testing takes very little time and is also part of learning. Assuming six hours of school per day in a 180-day school year, six state-required two-hour tests would consume just 1.1 percent of school time. Students, moreover, learn not just from being taught and studying but also by preparing for and taking good examinations, which cause them to reinforce and consolidate what they have learned. Indeed, good students learn to assess themselves as they learn. Finally, assessment provides information on student progress to parents, school boards, legislators, and the public, and an objective basis for comparing performance among teachers, schools, districts, and states.

Arkansas Standards and Assessment

Arkansas' standards are improving. The state's extensive use of Advanced Placement, International Baccalaureate, and end-of-course tests is commendable, as are the detailed numerical specifications for measuring student achievement, including reporting of the percentages of students at the "Below Basic," "Basic," "Proficient," and "Advanced" levels on statewide criterion-referenced tests. Other chapters in this volume, nonetheless, reveal the need for additional progress in the state's academic standards, which in turn would dictate changes in the testing regimen and the criteria for achievement proficiency.

Even if the standards remain unchanged, however, Arkansas' current assessment system needs improvement. Many of the state's education leaders recognize this. In a February 9, 2005 memo, for example, a study group of Arkansas superintendents and other leaders concluded that the present system is too limited and provides too little information on student achievement and progress.

The tests are administered too early in the school year, and the results are too slow in being returned to be most useful to educators and informative to parents. Study group members were also concerned about the possible lack of alignment or linkage between the tests and the standards. They concluded that the assessment system does not meet the criteria called for in Act 35 and the federal No Child Left Behind Act.

Components of the Assessment System

Arkansas law mandates two types of tests: The norm-referenced test (currently a subset of the Iowa Test of Basic Skills [ITBS]) allows comparison of Arkansas students with others in the nation. The criterion-referenced tests are intended to assess the degree to which students have mastered Arkansas state standards. Even if these tests are faultless, they cannot compensate for weak standards. Tests cannot be aligned with vague standards and should not be aligned with ones that are flawed. Since Arkansas is plagued by both problems, substantial changes in the assessment system are needed.

Norm-Referenced Tests

The norm-referenced Iowa tests presently used in Arkansas are reliable, and permit comparisons of Arkansas students and schools with their national counterparts, but they are neither geared toward Arkansas curriculum frameworks and standards nor are they state-of-the-art examinations. Only two parts of the limited Basic Battery of the Iowa tests are used, which further limits possible state and national comparisons. Narrow testing tends to concentrate educators' efforts on teaching a limited set of skills at the possible sacrifice of other things that students should be learning.

It is now possible to use computer-administered, computer-adaptive (CA) tests. Reliable and nationally normed CA tests solve many problems of conventional paper-and-pencil tests, and they are recommended here. For years, they have been used by the military services and colleges and universities, and some 6000 K–12 schools now employ them in the United States. Idaho adopted such tests rather than engaging in difficult and costly development of its own examinations.[1] An early Arkansas adopter of CA testing, the Hot Springs school district, is highly pleased with this form of assessment.

CA tests can serve simultaneously as both norm- and criterion-referenced tests, and have several other impressive advantages.[2] Unlike many state criterion-referenced tests, they are technically reliable and can be calibrated with national and state content and performance standards. CA tests are better protected against cheating,[3] and individual students' scores are available immediately upon completing the test. Detailed school reports can be returned within twenty-four hours. By comparison, the procedures for conventional tests take more student time, and

1. The Idaho state webpage describes how the computer-adaptive tests work and confirms their advantages in a complete statewide application: "The Idaho Standards Achievement Tests (ISAT) provide teachers, students, and parents with an accurate assessment of student progress in mastering the skills of Mathematics, Reading, and Language Usage. When administered at regular intervals over time, it is possible to find out whether an individual student, or an entire grade level, is making satisfactory progress in these basic skills areas. Teachers can use this assessment information for instructional planning for individual students or an entire class." See http://www.nwea.org/support/idaho/ and its links.

2. Disclosure: The author has no financial interest or consulting relation with a K–12 computer-adaptive test publisher and serves on the board of the Chicago International Charter School Foundation, which evaluates the progress of its some 6,000 students on its eight campuses with computer-adaptive tests.

3. How does a CA test minimize cheating? The "item pool" for any given grade or subject is composed of thousands of continuously updated sets of items from which the computer first randomly and then *adaptively* selects items to administer individually to each student. For this reason, students sitting next to one another are highly unlikely to be exposed simultaneously to the same item.

their results take a long time to be returned,[4] as Arkansas state leaders noted.

CA tests can readily be "vertically scaled"—they yield scores comparable across grades, as is now required by Arkansas law.[5] CA tests, moreover, are particularly desirable for measuring value-added student progress. Like a yardstick capable of measuring both long and short objects, vertically scaled tests can compare the progress of schools and students regardless of low or high starting points. They are much more efficient than traditional tests because they avoid giving students tasks that are too hard or too easy; rather, CA tests continuously "adapt" to an individual student's proficiency as the test is taken and thus require less time to complete. They can be conveniently administered to transfer or migrant students when they enter school during the academic year.

Perhaps the greatest advantage of CA tests is that, while taking less time than traditional tests, they can be given as many as four times a year to individual students or groups. The cost for such multiple administrations is less than the typical cost of one administration of a traditional test, which is usually given only annually.

4. Conventional paper-and-pencil tests need to be shipped to schools, given close protection against breaches of test security, and shipped back to service centers for scoring. Additional and uncertain amounts of time are required for scoring, printing, and shipping score reports to districts, which in turn may take additional time to ship or transmit the scores and analyses to schools and state departments of education.

5. This wide-range feature is highly advantageous for several reasons: Students within a single class may vary across six grade levels in proficiency, which makes many conventional paper-and-pencil tests inappropriate and unreliable for more and less advanced learners since such tests are targeted on the average student. Because the computer-adaptive tests are reliable over a wide range, they provide a better basis for calculating "value-added" or progress scores than do conventional tests (see the chapter by Erik A. Hanushek and Caroline M. Hoxby).

Just as firms and organizations benefit from more frequent feedback on their results, educators and parents can make use of more frequent and detailed reports on individual students, classes, teachers, and the school as a whole. Testing thus serves the purpose of diagnosis and mid-course corrections for educators in addition to summative reporting at year's end. Both overall status and value-added reports on the school can be prepared for parents, school boards, legislators, and the public.

Based on his direct experience, Hot Springs schools' research director Joel Rush has found CA tests convenient to use. He finds them far superior to past and present Arkansas state assessments, which are often ill matched to state standards, inadequately cover course content, provide weak diagnostic assessment of student problems for instructional guidance, and take too long to report testing results. His experience confirms the study group's criticisms of the present Arkansas assessment system and provides a promising, experience-based solution.

Criterion-Referenced Tests

Like many other states, Arkansas has found it difficult to develop rigorous standards, particularly for students in the early grades. It also needs reliable tests aligned to the standards and useful to educators. It seems possible that computer-adaptive tests, as in Idaho, can be a cheap, reliable, and time-efficient substitute not only for the norm-referenced tests, but also for many of the criterion-referenced tests. How might this type of testing system be phased in?

Obviously, time is needed to improve the state standards before new Arkansas-unique criterion-referenced tests can be chosen, adapted or developed. While that process is underway, the state may need to maintain much of its present assessment system. In the meantime, however, as is already taking place in

Hot Springs, schools can make good use of CA tests for more frequent assessment as a basis for improving achievement. This costs little and requires minimal classroom time but is likely to confer significant benefits.

It seems likely that well-chosen CA tests can cover many of the Arkansas standards, but this requires evidence rather than presumption, and vendors should be asked to demonstrate this alignment. The state should independently study how well the (subscore) components of CA tests match (or correlate with) the emerging state standards[6] and what standards they may not adequately address.

The study may show that the CA tests require supplementation with unique Arkansas content. It is also likely that some skills, such as writing, are best assessed through separate tests of actual performance. Thus, a combination of the CA with other tests can serve several state and NCLB purposes while providing greater amounts of reliable information more quickly to educators, parents, students, legislators, and the public.

Reading Comprehension and General Knowledge Assessment

Reading proficiency may be the most important skill to learn in school and so it merits special attention when developing the standards and assessments by which it is taught and measured. This author agrees with Professor E. D. Hirsch Jr.'s views on the subject, expressed in his chapter. Reading "proficiency" has at least two crucial aspects: "process," referring to the early learning of such things as letter recognition, how letters combine, and how

6. The study could also be designed to "calibrate" the CA tests with criterion-referenced, previous norm-referenced tests, and NAEP tests to allow for various comparisons.

to sound out new words, and "comprehension," which concerns textual understanding.

After mastering the basic processes of reading, the main challenge for readers is to comprehend the information contained in a book, poem, reading passage or other body of text. To comprehend new material, learners must possess general background knowledge. Readers who have no knowledge of the Civil War, for example, are unlikely to understand reading passages dealing with its causes or consequences. Thus, general knowledge and the ability to comprehend what one reads are nearly inseparable.[7] General knowledge is vast, but Hirsch and his colleagues have devoted years to developing and perfecting Core Knowledge, a publicly accessible specification of what knowledge is most important for K–12 students to know in each subject and grade level, which goes beyond the narrow NCLB requirements.[8]

Arkansas should re-examine and revise its standards in all curriculum areas not only in light of the three chapters in this book dealing with standards but also Hirsch's authoritative specifications. In addition, assessments should be developed that are aligned with the revised standards. Holding schools accountable

7. Significant correlations of general knowledge, reading comprehension, and other language skills provide empirical support for this view. Students and adults who score high on tests of general knowledge also tend to have high scores on reading comprehension tests; students who gain on tests of general knowledge tend to gain on other requiring language proficiency. See, for example, the U.S. Department of Health and Human Services Administration for Children and Families, "Language, Literacy, and Numeracy Instruments," http://www.acf.hhs.gov/programs/opre/hs/faces/instruments/child_instru02/cognitive_woodcook.html; Scott K. Baker, "Vocabulary Acquisition: Synthesis of the Research," http://idea.uoregon.edu/~ncite/documents/techrep/tech13.html; Grover Whitehurst, "Much Too Late," http://www.educationnext.org/20012/8whitehurst.html; and E. D. Hirsch Jr., "The Tests We Need and Why We Don't Have Them," http://www.coreknowledge.org/CKproto2/about/articles/stateTests.htm.

8. See, for example, a listing of many books by Prof. Hirsch at http://www.campusi.com/author_E._D._Hirsch.htm such as *What Your Third Grader Needs to Know* and *Cultural Literacy: What Every American Needs to Know*.

on such tests would permit Arkansas to advance beyond the federal No Child Left Behind proficiency requirements to include such subjects as history, civics, geography, and other important school fields. Such an endeavor would help bring Arkansas to the first ranks of state assessment systems and effectively, cost-effectively, and more broadly promote student learning.

Recommendations

1. Continue reporting the percentages of students that attain "Below Basic," "Basic," "Proficient," and" Advanced" levels so that educators are encouraged to help students at all levels of proficiency rather than those that are just below the proficiency criterion.
2. Maintain the use of the Advanced Placement and International Baccalaureate exams, and extend end-of-course tests to more subjects and courses.
3. Launch and evaluate a pilot computer-adaptive examination system for its benefits and possible weaknesses.
4. Commission an analysis of the possibility of cost-effectively substituting computer-adaptive tests for the state's norm-referenced tests and substantial parts of its criterion-referenced tests.
5. Revise state standards in all curriculum areas in light not only of the recommendations elsewhere in this volume but also state-of-the-art Core Knowledge specifications.
6. By 2006–7, assess students' general knowledge beyond what is required by the federal No Child Left Behind Act, including history, civics, and geography.