North Central Regional Educational Laboratory | LEARNING POINT ASSOCIATES™

Critical Issue: Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics

This Critical Issue was written by Asta Svedkauskaite, program associate at Learning Point Associates, and researched by Mary McNabb, Ph.D., director of Learning Gauge Inc.

Editorial guidance was provided by Gilbert Valdez, Ph.D., director of the North Central Regional Technology in Education Consortium and codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC); and Barbara Youngren, NCEMSC director.

The Critical Issue team would like to acknowledge the following experts for reviewing this article: Arlene Hambrick , Ph.D., National-Louis University faculty member and a private consultant in the field of equity in mathematics and science education; Terese A. Herrera, Ph.D., mathematics resource specialist at Eisenhower National Clearinghouse; and Nijole R. Mackevicius, External Resources coordinator at Chicago Public Schools' Office of Mathematics and Science .

Pathways Home

ISSUE:

Reform documents such as the *National Science Education Standards* (National Research Council, 1996), *Benchmarks for Science Literacy* (American Association for the Advancement of Science, 1993), and *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2000) suggest that the focus of planning and teaching should be on providing all students with optimal opportunities to learn to their maximum potentials. High-stakes testing can easily narrow the scope and depth of the curriculum and thus in-depth learning of students. Tests used for accountability purposes to meet federal and state requirements need to be aligned with curriculum standards that explore science and mathematics in a comprehensive fashion. Furthermore, to continually improve curriculum and instruction, some tests need to provide ongoing (formative) feedback that a high-stakes test once or twice a year cannot provide. Taking into account the continuum between formative assessment to improve learning and summative assessment for accountability, this Critical Issue explores multiple dimensions of assessment from an overall perspective and also with specific references to science and mathematics.

Overview | Goals | Action Options | Pitfalls | Different Viewpoints | Cases | Contacts | References

OVERVIEW:

Introduction

Although standards-based reforms have been the focus of school improvement agendas in districts across the country for more than a decade, the No Child Left Behind (NCLB) Act (NCLB, 2002) makes content standards a required part of federal and state accountability systems. NCLB mandates that states have content standards in reading, writing, mathematics, and science, and that states develop or adopt tests to measure students' achievement of those standards. Specifically, it requires states to implement assessments in mathematics and reading in Grades 3–8 and at least once in Grades 10–12 by the 2005–06 school year (U.S. Department of Education, 2002). Beginning in the 2007–08 school year, schools must administer annual tests in science achievement, at least once in Grades 3–5, 6–9, and 10–12 (U.S. Department of Education, 2002). NCLB requires states to make a number of changes to their accountability systems in order to comply. Most notable changes include the following:

- Identifying adequate yearly progress (AYP) goals.
- Participating every other year in the National Assessment of Educational Progress (NAEP) for reading and mathematics in Grades 4 and 8.
- Disaggregating test scores for subgroups of students according to disability status, English language proficiency, socioeconomic background, and racial and ethnic backgrounds of students.

Besides state-administered testing, NCLB legislation calls for evaluation of classroom-level opportunities to learn for all students. Such evaluation requires the use of multiple forms of assessment for screening and diagnostic purposes, and it puts new emphasis on uses of formative assessment that can inform instructional decisions. Checking for student understanding almost always takes place in the classroom where teachers can probe for various explanations of students.

As veteran elementary teacher Barbara Campbell explains, all kinds of understanding check-ins are important as they can show whether the student is learning. [Video: 1:03]

New classroom-based as well as large-scale assessments are continually under development to measure students' achievement in alignment with content standards, such as in science and mathematics. Under NCLB, states are required to further define their content standards into measurable grade-level standards for tracking AYP. Classroom-based assessments aligned to a continuum of within-grade progress benchmarks or goals can provide teachers with effective tools for helping students progress toward their expected year-end outcomes, provided those outcomes account for the more complex and enduring concepts, principles, and skills within the science and mathematics curriculum.

It is important to reemphasize that under NCLB legislation, both classroom-based assessments and state-mandated standardized tests are increasing in frequency and importance. Classroom-based assessments provide teachers with feedback about the quality of responses students generate for open-ended, real-world tasks and may be flexible enough to go beyond basic skills testing.

According to Gilbert Valdez, Ph.D., learning and assessments need to be related to "big ideas" that students encounter in their daily lives beyond the classroom. [Video: :35]

The fear many educators experience is that the increase in standardized testing may narrow the curriculum as a response to policy pressures surrounding students' standardized basic test scores. Herman (1997) reports cumulative findings of researchers who investigated the impact of mandated, public testing and found that standardized testing encourages teachers and administrators to focus instruction on narrow test content as they tend to incorporate the following strategies:

- Encouraging drill and practice on rote memory tasks related to content.
- Avoiding inquiry-based learning activities in which complex concepts and skills are developed.
- Skipping content that may be in the standards but not on the test.

Yet public testing cannot be treated dismissively. Standardized tests *are* important AYP measures and can become a positive force in education and provide an overall "positive enhancement of learning" (Wright, 2001, p. 60). They can provide an overall health report of K–12 education nationally and globally as well as compare the United States to other countries. As an example, in their *Science* editorial discussing the Trends in International Mathematics and Science Study (TIMSS) test for 2003, Bybee and Kennedy (2005) summarize: "At Grade 4, between 1995 and 2003, U.S. student scores held constant, although their international ranking declined slightly. But the average scores of U.S. eighth-graders made statistically significant improvements between 1995 and 2003 in both mathematics and science" (p. 481).

Results on such a scale show that standardized tests may be helpful to build a big picture of student achievement. Large-scale testing is helpful in assessing skills and concepts that can be measured by multiple-choice and other standardized format tests because there are skills and concepts that *can* be tested that way. Also, tests in these formats may be a wise use of time and dollars as well as bring value for instruction. According to the Commission on Instructionally Supportive Assessment (2001), "state-mandated accountability tests must be useful to educators concerned about improving the instruction of children." The Commission presented nine requirements (or steps) for preparing statewide achievement tests that would provide educators with information they need to improve their instruction and benefit students in their own classroom. Applicable to this Critical Issue, there are two requirements that need to be highlighted:

1. The state must provide educators with optional classroom assessment procedures that can measure students' progress in attaining content standards not assessed by state tests.
2. A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments.

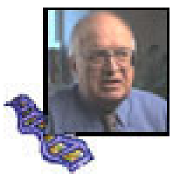
The national push for content standards and their implementation for the last 15 years have certainly had a cumulative effect on the overall educational system, influencing state standards, curriculum materials selection, and assessments –as well as teacher education (Bybee & Kennedy, 2005). The following sections of the Critical Issue provide information on assessment as it further evolves in the era of NCLB and national accountability mandates.

Assessment Reform Trends

Decade after decade, the tendency in this country was to implement standardized testing as a means to improve schools —starting with a district level and going up to an international level (Stiggins, 2004). The mistake, it appears, was to "believe that once-a-year standardized assessments alone can provide sufficient information and motivation to increase student learning" (Stiggins, 2004, p. 22). In fact, literature reports few studies showing student test scores and improvement in learning that is attributable to the presence of high-stakes tests alone (Stiggins, 2004).

In 1998, the National Research Council with the support of the National Science Foundation convened a Committee on the Foundations of Assessment to explore implications from advances in the learning sciences for improving educational assessment practices in schools. The committee focused on assessment of students' achievement rather than on tests that predict students' capacity for future achievement such as college entrance examinations. Starting with the realization that existing assessments are grounded in outdated theories of how people learn and how to measure what students know, the committee established a new foundation for assessment designs and related practices that better support 21st century learning goals.

The committee explains that assessment reform is necessary because of new expectations for 21st century knowledge and skills that arose, in part, from our increasingly competitive economy. The challenge of the 21st century is to ensure every student develops an appreciation for and mastery of science and mathematics subjects. It also is critical they "become the creative, problem-solving, critical-thinking workforce of the 21st century" (Burmaster, 2003). Teachers therefore are faced with an enormous task of testing their students in light of the 21st century requirements for learning, work, citizenship, and life.



It is critical that assessment is diverse and divergent and develops through many pathways rather than from a single source, says Dr. Valdez. [Video: 1:07]

Assessment reform is a key component of a larger educational reform trend involving advancements in science and mathematics, technology innovations, and new accountability policies. The authentic assessment (or alternative assessment)] movement responded to the lack of tests addressing higher-order thinking and other vital 21st century skills, as outlined through the work of the Partnership for 21st Century Skills. Herman (1997) stated:

"… There has been great enthusiasm for alternative assessments, which ask students to create their own responses rather than simply selecting them, assessments that many believe best represent the kinds of skills students will need for future success" (p. 2).

The Committee of the Foundations of Assessment recognizes and supports the increasing demand for better student outcomes in effective thinking and reasoning, communication and team-building, complex problem solving, higher order literacy and computation, and self-directed life-long learning skills (CEO Forum on Education and Technology, 2001; Haertel & Means, 2000; North Central Regional Educational Laboratory & Metiri Group, 2003). In support of the 21st century skills, the committee advocates that assessment be aligned with a range of skills: "Assessments must tap a broader range of competencies than in the past. They must capture the more complex skills and deeper content knowledge reflected in new expectations for learning" (National Research Council, 2001, pp. 22–23). It is therefore vital state-level tests align to content standards and that school and classroom curriculum and assessments align to the same standards as well.

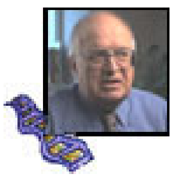
One of the most recent and influential document that determines the knowledge and skills each student should learn for school and the workforce is a report published by The Partnership for 21st Century Skills, a business-education-policymaker organization, titled *Learning for the 21st Century*. The report articulates a collective vision of 21st century learning and provides an assessment tool that can help determine those skills. To help stakeholders create a 21st century learning environment, the Partnership has released several support documents, *The Road to the 21st Century: A Policymaker's Guide to the 21st Century Skills*, *Route 21: An Interactive Guide to 21st Century Learning*, and *Milestones for Improving Learning and Education Guide for 21st Century Skills*. This set of comprehensive tools will help stakeholders in education integrate 21st century learning skills into schools and beyond.

Trends in assessment reform have pushed some states to explore alternative assessments options and implement standards-based curriculum at the state level. Whether states do or do not have content standards articulated, under the NCLB legislation all states and school districts are required to gather data about student achievement over time and to hold schools, teachers, and students accountable for meeting curricular standards and for making AYP.

Summative Versus Formative Assessment. It is critical to distinguish between the functions and purpose of both summative and formative assessments.

Dr. Valdez explains, they are both important yet for different purposes and intents. [Video: :27]

Assessment becomes formative "when the evidence is actually used to adapt the teaching to meet student needs" (Black & Wiliam, 1998b). It stands in contrast to summative assessment, which generally takes place after a period of instruction and requires making a judgment about the learning that has occurred (e.g., by grading or scoring a test or paper). In other words, while summative assessment is used for accountability to showcase and prove that students gained knowledge over a certain time, formative assessment is valuable in its intent to improve learning and to change instruction based on results. In fact, research shows that formative assessment has one of the biggest effects on learning, even equal to the effect of parental influence.



Dr. Valdez further discusses the two functions of assessment. [Video: 1:20]

Certain uses of achievement test results are termed "high stakes" if they carry serious consequences for students or for educators. High schoolwide scores may bring public praise or financial rewards, and low scores may bring public embarrassment and heavy sanctions. For individual students, high scores may bring a diploma attesting to exceptional academic accomplishment, while low scores may result in students being held back in grade or denied a high school diploma and, consequently, higher education opportunities.

The mandate to track students' yearly progress in academic subjects suggests a place for growth models of assessment. For growth to be seen, assessment cannot be sporadic. As Seltzer, Choi, and Thum (2002) explain it:

"… Much can be learned by moving beyond snapshots of student achievement at single points in time to analyses and summaries of student growth. To be sure, the notion of growth in knowledge and skills lies at the heart of definitions of learning and education" (p. 2).

In their Longitudinal Study of American Youth, Seltzer, Choi, and Thum (2002) demonstrate the use of within-student growth and between-student growth models of assessment, in which they captured time series data for students in several schools. Within-student growth models enable teachers to understand students' initial status in relationship to a set of content standards and formatively assess their rate of change toward increasing levels of competency. Using classroom-based assessments designed for the purpose of identifying students' initial status and growth over time can enhance teachers' ability to provide equitable opportunities to learn for students with varying instructional needs. On the other hand, between-student growth models enable schools, districts, and grade levels to estimate the "mean rate of change for a group of students, assess the extent to which students vary in their rates of change, and identify important correlates of change" (Seltzer, Choi, & Thum, 2002, p. 6).

National standards have been developed in science and mathematics by the National Research Council and the National Council of Teachers of Mathematics, respectively.

Formative Assessment

Accountability for schooling often focuses teachers' attention on basic-skills test scores, leaving little time for promoting deep understanding, inquiry, or problem solving in the classroom. Yet it is important to consider that during the course of a year, teachers can build in many opportunities to assess how students are achieving in relation to AYP requirements and then use this information to make beneficial changes in instruction. Ideally, most teachers would like to find ways to balance both, even though assessment requirements can be overwhelming.

Teachers and students alike need to know what students are learning. When teachers know how students are progressing and where students are having trouble, they can use this information to make necessary instructional adjustments such as reteaching, trying alternative instructional approaches, or offering more opportunities for practice. These activities can lead to improved student success. The systematic and regular measurement of students' progress that occurs at the classroom level and the process which allows use of test results to shape instruction is the basis of formative assessment (Barchfeld-Venet, 2005).

According to Braun (2001), classroom-based assessment has three main functions. From the perspective of individual students, these functions are to choose, learn, and qualify. To "choose" means to use assessment data to help inform the selection of an appropriate course of study. To "learn" refers to using assessment data as feedback about how well one is learning. To "qualify" refers to using assessment data to certify an accomplishment. From a program perspective, assessment data are typically used to "place" a student into an appropriate course of study, to "monitor" the student's learning progress, and to "report" the accomplishments of the students in the program or school (Braun, 2001).

Along with definitions of the three functions of assessment, Braun (2001) also explains the importance of goal defining when selecting or designing classroom-based assessments: "The success of an assessment or an evaluation depends in large measure on being clear about the goals and then engaging in a disciplined design and development process that is focused on these goals." Much can be accomplished with such assessments that both support students' learning and inform teachers' instruction. That is why teachers may value classroom-based assessments and make better use of them than of the high-stakes standardized tests.

Regardless of the type of assessment, it is important to ensure alignment across content standards, opportunities to learn in the classroom, and the achievement measures –both published and teacher-created—used to evaluate progress toward proficiency in the disciplines. Providing students opportunities to learn especially is critical because the content calls for differentiated instruction to meet individual learner needs.

In addition, providing students with differentiated instruction to meet multiple individual needs requires educators to measure student learning through effective assessment. Foertsch (1999) defines 10 major underlying principles of effective testing and assessment in the classroom, including any achievement measure:

1. Clearly define what you will assess. What do you expect your students to be able to do?
2. Define the purpose of your assessment. Are you intending to conduct placement, formative, diagnostic, or summative assessment?
3. Select or develop assessment procedures that closely match targeted learning goals or abilities. Are the tasks accurately reflecting the learning goals you wish to assess? Are the formats of tests and items affecting student responses? What irrelevant factors do interfere with assessment? Does the test contain appropriate sampling of test or item difficulty?
4. Know the limitations of assessment procedures used. Does the way the test is developed, administered, or interpreted present any limitations?
5. Use a variety of assessment procedures. Is your assessment comprehensive (e.g., do you make use of observations, class work, professional judgment, student and parent input?)
6. Evaluate the assessment or test you develop or use. Are they valid and reliable?

7. Communicate assessment results clearly to all users. Do the students, parents, teachers, and other stakeholders understand the results?
8. Consider and address personal implications. Are your biases influencing your professional judgment?
9. Strengthen the link between assessment and learning. Is assessment helping improve instruction and learning?
10. Assessment should serve a useful purpose and not be an end in itself. Are tests an integral part of your instruction and are they helping you guide instruction?

The 10 principles offer teachers guidance on how to create, use, and interpret classroom tests purposefully and judiciously. NCLB mandates for meeting opportunity-to-learn requirements reinforce the principle of fair testing. In the classroom, fair testing practices hold true for both formative and summative tests of all kinds. Assessment experts advocate that test content incorporate what is taught through the curriculum, materials, and instruction, and that students have opportunities to really learn before high-stakes consequences are imposed for failing any test (American Educational Research Association, 2000).

As a result of NCLB, academic standards are being translated into goals for AYP. Benchmarks aligned to state-level AYP requirements can provide the criteria for formative assessment objectives that identify students' level of progress throughout the school year. In science, teachers can use benchmarks that describe cognitive assessment objectives important to the science curriculum and then identify assessment tasks students can perform to provide evidence of their benchmark competence. Teachers and administrators can use progress mapping to track students' learning of science benchmarks, to provide students with feedback, and to set future learning goals. Dependence on one type of assessment does not provide a comprehensive view of student learning progress (Wright, 2001). "To be comprehensive, assessment must be authentic, meaning it resembles the classroom experience or 'real-life experience'" (Wright, 2001, p. 61). Use of classroom-based assessments that provide feedback about the quality of responses students generate for open-ended, real-world tasks and that go beyond basic-skills testing will help schools avoid the pitfalls of narrowing the curriculum as a response to policy pressures surrounding students' standardized basic test scores.

In recent years, there has been increasing interest in finding methods of assessment that go beyond the traditional ones. The need for teachers to adapt the teaching and to know what students are learning requires innovative methods of assessment. In science and mathematics, quizzes, tests, and teacher evaluations of homework, reports, and projects are often used to assess students' progress. Knowing that such traditional assessment devices do not accomplish everything, other assessment methods also should be used, such as concept mapping or creating science portfolios (Hickey, Kindfield, Horwitz, & Christie, 2003; Atkin, Black, & Coffey, 2001). Portfolios, or collections of student work, can be used formatively if students and teachers annotate the entries and observe growth over time and practice (Duschl & Gitomer, 1997).

In science education, concept mapping has been widely recommended and used in a variety of ways to observe change in students' understanding of concepts over time, to assess what the learner knows, and to reveal their unique thought processes. Anderson-Inman, Ditson, and Ditson (1999) cite considerable evidence that concept mapping promotes meaningful learning in science. Formative feedback through the means of concept mapping and science portfolios, for example, form a significant basis for success—both for students and for teachers.

In general, well-designed formative assessments positively affect a student's role, motivation, and self-perception, which allow students to view assessment as supportive rather than punitive (Sadler, 1989; Barchfeld-Venet, 2005). Studies indicate that evaluation and reflection involving analysis and feedback are important aspects of effective teaching (McAninch, 1993). Performance feedback benefits teachers and students (Bransford, Brown, & Cocking, 1999; Pellegrino, Chudowsky, & Glaser, 2001) by doing the following:

- Identifying the diversity in students' prior knowledge, which serves as the basis for tailoring instruction to draw upon it.

- Acknowledging and building upon the range of students' achievement levels.
- Recognizing students' misconceptions or gaps in their knowledge that instruction can address.
- Stimulating students' future learning by engaging and motivating learners to deeper levels of analysis on a personalized basis.

The potential benefits of formative assessment are so significant that teachers cannot ignore this assistance to students. Black and Wiliam (1998a) conducted an extensive research review of 250 journal articles and book chapters winnowed from a much-larger pool to determine whether formative assessment raises academic standards in the classroom. They concluded that efforts to strengthen formative assessment produce significant learning gains as measured by comparing the average improvements in the test scores of the students involved in the innovation with the range of scores found for typical groups of students on the same tests. Black and Wiliam (1998b) found effect size between .4 and .7, with the greatest impact for low-achieving students: "… Improved formative assessment helps low achievers more than other students and so reduces the range of achievement while raising achievement overall." They recommend setting up local groups to tackle formative assessment at the school level while collaborating with other local schools. In science or mathematics, teacher groups can collaboratively design classroom-based assessments that will help them understand their students' progress toward the core subject concepts and skills, and plan their instruction accordingly.

Collaborative work on assessment is one of the greatest forms of professional development, according to Dr. Valdez. [Video: :53]

While feedback generally originates from a teacher, students also can play an important role in formative assessment through self-evaluation. Two experimental research studies have shown that students who understand the learning objectives and assessment criteria and have opportunities to reflect on their work show greater improvement than those who do not (Fontana & Fernandes, 1994; Frederiksen & White, 1997). Students with learning disabilities who are taught to use self-monitoring strategies related to their understanding of reading and writing tasks also show performance gains (McCurdy & Shapiro, 1992; Sawyer, Graham, & Harris, 1992). Another student group that is commonly underrepresented is gifted learners, both identified and unidentified. Frequently, gifted students are misidentified as struggling learners because of an inappropriate assessment measure, when in fact they are not.

As a result, such students end up being bored and having high drop-out rates, Dr. Valdez comments. [Video: :54]

Since the goal of formative assessment is to gain an understanding of what students know and don't know, Black and Wiliam (1998b) encourage teachers to use questioning and classroom discussion as an opportunity to increase the knowledge of all students and improve their understanding. They caution, however, that teachers need to be sure to ask thoughtful, reflective questions rather than simple, factual ones and then give students adequate time to respond. In order to involve everyone, they suggest strategies such as the following:

- Invite students to discuss their thinking about a question or topic in pairs or small groups, then ask a representative to share the thinking with the larger group (sometimes called think-pair-share).
- Present several possible answers to a question, then ask students to vote on them.
- Ask all students to write down an answer, then read a selected few out loud.

Teachers might also assess students' understanding in the following ways:

- Have students write their understanding of vocabulary or concepts before and after instruction.
- Ask students to summarize the main ideas they have taken away from a lecture, discussion, or assigned reading.
- Have students complete a few problems or questions at the end of instruction and check answers.
- Interview students individually or in groups about their thinking as they solve problems.
- Assign brief, in-class writing assignments (e.g., "Why is this person or event representative of this time period in history?").

In addition to these classroom techniques, tests and homework can be used formatively if teachers analyze where students are in their learning and provide specific, focused feedback regarding performance and ways to improve it. Black and Wiliam (1998b) make the following recommendations:

- Frequent short tests are better than infrequent long ones.
- New learning should be tested within about a week of first exposure.
- Be mindful of the quality of test items and work with other teachers and outside sources to collect good ones.

Emerging Assessment Concerns and Effect Factors

One of the biggest concerns in assessment is the validity and reliability of measures as well as instruments of those measures. The terms "test validity" and "test reliability" are often used but seldom understood by teachers.

Dr. Valdez explains, the two terms have very specific meanings in statistics and are commonly misused in common language. [Video: 1:30]

Reliability is basically: Can that instrument be used reliably in other settings by different people? In other words, if you go to a different classroom and I go to a different classroom and the students are "knowledge," would the fact that I'm administrating the test versus you give you a different thing? If a test shows the same results with different people administering it in different settings, then it's reliable. So the concept of interrelated reliability that you often hear implies that regardless of who is giving the test, it's got reliability between the settings. So validity is basically about: Is a test doing what it's intended to do? And reliability is: Will it do, be consistent, across different settings and different people administering the test?

Russell and Haney (1997) studied the effects of test administration mode to see whether tests administered on computer versus paper-and-pencil have an effect on student performance on multiple-choice and written test questions. The study found the effect of responses written on computer are significantly higher than those written by hand: "The size of the effects was found to be 0.94 on the extended writing task and .99 and 1.25 for the NAEP language arts and science short answer items. Effect sizes of this magnitude are unusually large and of sufficient size to be of not just statistical, but also practical significance (Cohen, 1977; Wolf, 1986)" (Russell & Haney, 1997). Such findings suggest that while the medium of instruction is important, the method of assessing student knowledge is critical. Russell and Haney (1997) concluded, "As more and more students in schools and colleges do their work with spreadsheets and word processors, the traditional paper-and-pencil modes of assessment may fail to measure what they have learned." Studies like this one suggest that more of today's paper-based assessments are becoming a thing of the past for various reasons, including outmoded test designs, mismatches between standards-based curriculum and assessments, differences in population groups' performance, a lack of feedback to help students improve, or inefficiency in administration, analysis, and reporting (Bennett, 2000).

In a later study, Russell (2000) claims that technologies used during learning activities also should be used during testing. He contends that student assessment methods should match the medium in which students typically work and advocates for state and local assessment programs to allow students the same technology assistance in the assessment process as they get in the learning process.

To take this argument further, Rose and Meyer (2002) explain that selecting appropriate testing accommodations for students on an individualized basis is a very complex endeavor involving the following factors that can confound results of traditional academic assessments:

- Individual student learning differences.
- Media characteristics of the testing technology.
- Type and amount of student supports.
- Poor alignment between curriculum and the test used to measure achievement.

Media characteristics of the technology used to administer the test also can present factors confounding test results when students are more apt in one medium (i.e., paper-and-pencil tests) than another medium (i.e., computerized tests that require keyboarding skills). This is a concern that will continue to grow as more assessments become technology based. Technology is a powerful tool not only in instruction but in assessment, hence it should be equally used in both, says Dr. Valdez:

Technology is a powerful tool not only in instruction but in assessment, hence it should be equally used in both, says Dr. Valdez. [Video: 1:20]

Bennett (2003) explains that as technology has become ever central to schooling, assessing students via technology-based methods will be increasingly required. Education leaders in several states and numerous school districts already are implementing technology-based tests for low- and high-stakes decisions in elementary and secondary schools in all key content areas. Most of these tests are computer-adaptive tests using multiple-choice items such as the one North Carolina piloted during the 2000–01 school year. The North Carolina Computerized Adaptive Testing System (NCCATS) pilot evaluated the feasibility, validity, and reliability of a computerized adaptive testing system to be used as an accommodation for students with disabilities to respond to multiple-choice test questions for the North Carolina End-of-Grade Tests of Reading and Mathematics in Grades 3–8 and the North Carolina High School Comprehensive Test in Grade 10. The NCCATS is one state initiative among eight projects in K–12 technology-based assessments under way at the state level of education, according to Bennett (2003).

Recent research into computer applications for assessing students' writing, such as automated essay scoring methods, has achieved a reliable level of agreement with a human rater's essay score, which has far-reaching implications for assessments that include open-ended responses (Burstein, Marcu, Andreyev, & Chodorow, 2001; Foltz, Gilliam, & Kendall, 2000; Kintsch, Steinhart, Stahl, & LSA Research Group, 2000). Another area of assessment research involves artificial neural networks to generate performance models of students' complex problem solving during performance of simulated science tasks that do not have predetermined solution paths (Vendlinski & Stevens, 2003). According to Bennett (2003), state efforts will need to go beyond the initial achievement of computerizing traditional multiple-choice tests to create assessments that facilitate learning and instruction in ways that paper measures cannot.

Besides the method of test administration, assessment also is influenced by varying approaches to curriculum and instruction. Different approaches to curriculum and instruction inherently carry assumptions about underlying learning theory and teaching philosophy that influence the types of assessments needed for alignment (Shepard, 2000). Bennett (2000) points out that tests also are affected by factors or cognitive constructs that include knowledge organization, problem representation, mental models, and automaticity, and which are not accounted for explicitly by many tests.

GOALS:

Teachers and administrators should do the following:

- Understand the role of high-stakes testing in relation to their classroom science and mathematics curriculum and the different uses of data from high-stakes testing and from classroom-based assessments.
- Identify the potential gaps in high-stakes testing content in relation to science and mathematics curricula and related content standards.
- Foster science inquiry and mathematics problem-solving skills despite the imperfections of high-stakes testing practices.
- Use valid and reliable classroom-based assessments that reflect the goals of science and mathematics curriculum and instruction (e.g., scientific knowledge, scientific methods, scientific habits of mind, societal issues, personal needs, career awareness).
- Reflect on evidence-based science and mathematics pedagogy and find science and mathematics lessons that are relevant, interesting, and that engage students in thinking and doing what real scientists and mathematicians do.
- Use an array of formative and summative assessments to determine students' strengths and weaknesses in meeting the science and mathematics standards and shape instruction conscientiously.
- Learn how to optimize science and mathematics curricula to meet several standards in one lesson, how to enhance the effectiveness of learning groups, and how to create a standards-based classroom that uses multiple measures aligned with AYP requirements.

ACTION OPTIONS:

- Administrators, teachers, parents, and community members need to become familiar with their state's NCLB plan (and its implications for their school district) as well as classroom practices and students' AYP requirements.
- Administrators and teachers need to understand the requirements for student subgroup data analysis and

accommodations that support all students reaching their achievement goals. They should document all students' opportunities to learn within science and mathematics curricula for program evaluation purposes.

- Administrators and teachers need to understand assessment literacy skills so that they can distinguish between classroom-based assessments that are aligned with state science and mathematics standards and assessments that are not aligned.
- Administrators and test designers need to work toward verifying the reliability and validity of the classroom-based assessments used in science and mathematics, and be certain that they are in compliance with fair use of tests.
- Administrators and teachers need to engage in data-driven decision making based on collecting and analyzing the data to be able to draw evidence-based implications for instruction and to provide students with adequate opportunities to learn in the science and mathematics classroom.
- Administrators and teachers need to be knowledgeable about how technology, both educational and assistive, can enhance assessment and data-driven decision-making practices in science and mathematics.
- Administrators and teachers need to align their science and mathematics curriculum with the state standards.
- Administrators and teachers need to provide useful feedback to students and parents about student performance in learning science and mathematics.

IMPLEMENTATION PITFALLS:

There are some common implementation pitfalls that schools encounter when trying to improve their use of assessments for decision making and providing timely and ongoing feedback to students about their progress, strengths, and areas for improvement. Some of these barriers that teachers face can include the lack of time and limited assessment literacy skills, delayed access to student test data, assessment that loses its meaning in the real world, or curriculum that is being taught to the test.

First of all, teachers may not have the time or assessment-literacy skills to create classroom assessments that generate reliable student progress data to inform their lesson planning. Assessment-literacy skills include skills to create assessments, analyze and reflect on student progress data as well as provide meaningful feedback to learners. Also, teachers may not have timely access to relevant student data, both progress data and summative data, from high-stakes tests.

The main pitfall, however, is teaching to the test. Many states are undergoing the process of designing or identifying new tests that align with content-area standards as part of their efforts to implement their NCLB plan, which provides an opportunity to make high-stakes tests more meaningful. However, it is quite common for schools to focus on tests so intensely that the tendency then is to teach students to take those tests rather than learn beyond the test. As Reich (2001) warns:

"They're best at measuring the ability to regurgitate facts and apply standard modes of analyses. … [But] it's far more important to learn how to identify and solve new problems, think critically, and challenge assumptions. … Standardized tests can help measure whether children have achieved an adequate level of communication skills and numeracy, and even help pinpoint where children need more guidance. … It is training a generation of young people to become exquisitely competent at taking standardized tests, and a generation of teachers to become exceedingly good at teaching how to take them. Neither of these competences has much to do with preparing young people for what they will encounter when they leave our schools."

Furthermore, limiting the curriculum to teach to a test that only covers basic scientific facts and mathematical formulas because of short class periods, lack of lesson planning time, or history of low student performance on high-stakes tests violates existing guidelines for fair use of tests.

"Any narrowing of the curriculum, along with the confusion of training to pass a test with broader notions of learning and education are especially problematic side effects of high-stakes testing for low-income students. The poor, more than their advantaged peers, need not only the skills that training provides but need the more important benefits of learning and education that allow for full economic and social integration in our society" (Amrein & Berliner, 2002).

More importantly, such limitation on curriculum delivery has become a major concern not only for teachers but also for parents. They fear their children are spending time to learn facts and complete multiple drills, "rather than spending time on problem solving and the development of critical and analytical thinking skills. Teachers at the grade levels at which the test is given are particularly vulnerable to the pressure of teaching to the test" (Domenech, 2000).

A recent National Academy of Science/National Research Council report on school learning makes clear that schooling that too closely resembles training, as in preparation for testing, cannot accomplish the task the nation has set for itself, namely, the development of adaptive and educated citizens for the new millennium (Heubert & Hauser, 1999). Some researchers even question:

Is there "evidence of student learning, *beyond the training that prepared them for the tests they take*, in those states that depend on high-stakes tests to improve student achievement? … Although states may demonstrate increases in scores on their own high-stakes tests, it appears that transfer of learning is not a typical outcome of their high-stakes testing policy" (Amrein & Berliner, 2002).

Such transfer of learning, however, may not necessarily happen in hands-on learning experiences. So if teachers have been teaching science using a hands-on, discovery-oriented approach, they may have good reason to be anxious when the students in their class are expected to display a command of science on standardized tests. After all, teaching science in a hands-on fashion may not provide the background in science knowledge that students from more traditional, textbook-oriented classrooms have. On the other hand, the students in a discovery-oriented class are very likely to acquire many science process skills and develop a favorable attitude toward science.

DIFFERENT POINTS OF VIEW:

When Amrein and Berliner (2002) examined 18 states to see if their high-stakes testing programs were affecting student learning, they summarized arguments (pp. 4–5) highlighted in research studies supporting high-stakes testing as follows:

- "Students and teachers need high-stakes tests to know what is important to learn and to teach.
- Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the laziest students to work harder.
- Students work harder and learn more when they have to take high-stakes tests.

- Students will be motivated to do their best and score well on high-stakes tests….
- Scoring well on the test will lead to feelings of success, while doing poorly on such tests will lead to increased efforts to learn."

Amrein and Berliner (2002) examined the validity of these statements through both quantitative and qualitative research as well as by interviewing teachers who work in high-stakes testing environments. They came to a conclusion that "these statements are true only some of the time, or for only a modest percent of the individuals who were studied" (p. 5) and mentioned the stress factor as one of the reasons causing such varied outcomes.

Stiggins (1999) describes the pressure from high-stakes tests and its rationale as "the way to spur greater effort … through intimidation by means of the threat of dire consequences for low test scores" (p. 192). Some school districts have adopted this attitude in order to motivate low-performing students to achieve. However, there are other schools that think it is actually possible to make better use of standardized science tests with less anxiety.

Learning the basics is important for future learning; however, methods for learning the basics can be embedded in lessons that foster deep content-area learning. Students should understand that the high-stakes tests will not measure all that they have learned. If, for example, your class has carried out a hands-on unit on the use and waste of water in your school, explain to the students that they should not expect to see questions about it on their tests. Point out that some of the science units they have studied have given them a lot of information that will not be measured but will still matter a lot. Emphasize that they should not feel bad if many of the things they have learned are not on the test.

To further reduce student anxiety regarding high-stakes tests, teachers can take some class time before the test to teach basic standardized test taking strategies. The students will likely take many standardized tests during their school years and when they pursue employment. Investing time and energy to improve test-taking skills may annoy teachers because of their own feelings about testing, but it may help students become more successful test takers and not be intimidated by the test itself.

ILLUSTRATIVE CASES:

The following links provide additional information on examples of formative and summative assessment:

- **Data-Driven Decision Making: Best Practices Case Studies**. See information about best practices and case studies of school districts that are using data-driven decision-making systems to support continuous improvement and achieve their educational goals.
- **Data-Driven Districts**. American Association of School Administrators published an article about four data-driven districts that have implemented different strategies to use data to inform decision making.
- **A Private Universe**, produced by the Harvard-Smithsonian Center for Astrophysics, is a video documentary bringing to focus learning dilemmas that even the brightest students have about some of the key science concepts.

Additional Links

Assessment Ideas for the Elementary Science Classroom

This Web site deals specifically with the needs of elementary and/or middle-level science teachers. In addition to discussing conferences, interviews, contracts, and portfolios, this site provides templates or guide sheets for creating and using various assessment techniques.

Assessment Standards for School Mathematics

This is the third book in the National Council of Teachers of Mathematics Standards series. It has been developed as a guide for examining current assessment practices and planning new assessment systems.

Classroom assessment and the *National Science Education Standards*

This 2001 book for prospective and practicing teachers, designed as a companion to *National Science Education Standards*, discusses how formal and informal assessments can guide and improve pedagogy.

Consumer Guide: Performance Assessment

This Consumer Guide, sponsored by the U.S. Department of Education, is an archived publication for teachers, parents, and others who are interested in alternative techniques for student assessment. It not only discusses techniques but also lists the addresses of organizations that can provide additional information about the topic.

Exemplars: Standards-Based Assessment and Instruction

This Internet site offers differentiated performance assessment tasks that meet the national standards for science and mathematics and are designed to improve both assessment and instruction within these disciplines. All of the performance tasks include teaching tips, suggestions for use, rubrics, and a list of possible solutions.

Family Education Network: Authentic Assessment

This Web site provides techniques and resources for authentic assessment.

Math Assessment Techniques on PBS Teacher Source

The PBS Teacher Source section on mathematics has a selection of mathematics techniques that teachers will find helpful and practical. The techniques include questioning, conferencing, interviewing, self-assessing, and multiple-choice testing.

National Center for Education Statistics: The Nation's Report Card

The Nation's Report Card is a continuing assessment of what American students know and can do in a variety of subjects, including science and mathematics. The testing program itself is called the National Assessment of Educational Progress (NAEP). It samples student achievement at Grades 4, 8, and 12.

Classroom Assessment and the National Science Education Standards

The National Research Councilhas produced a useful, accessible book on classroom assessment in science that contains many vignettes about how teachers can adjust their teaching based on their observations, questioning, and analysis of student work. While the anecdotes are specific to K–12 science teaching, the chapters have broad applicability to the documented value of formative assessment on classroom achievement as well as what it requires in terms of teacher development and how classroom assessment relates to summative assessment such as state tests.

National Science Education Standards

This Web site contains the complete text of the *National Science Education Standards*, which are premised on a conviction that all students deserve and must have the opportunity to become scientifically literate.

No Child Left Behind Act

The reauthorized Elementary and Secondary Education Act focuses on four ideals: stronger accountability for results, more freedom for states and communities, encouraging proven education methods, and more choices for parents.

Northwest Regional Educational Laboratory

The *It's Just Good Teaching* series focuses on using assessment to inform and improve instruction in mathematics and science. The training kit, Toolkit98, offers readings, overheads, exercises, and handouts to help groups of teachers think through assessment issues in their schools.

Principles and Standards for School Mathematics

The National Council of Teachers of Mathematics offers a list of standards by grade and a summary of principles.

Project 2061

The focus of Project 2061 is to foster scientific literacy. It is a reform initiative developed by the American Association for the Advancement of Science, which seeks to improve the quality, increase the relevance, and broaden the availability of science, mathematics, and technology education. The entire text of *Benchmarks for Science Literacy* is available online. The publication expands the science-literacy goals in Project 2061's publication *Science for All Americans* into specific goals that should be achieved by the end of Grades 2, 5, 8, and 12.

Project Zero Research Projects

The Project Zero Web site features Assessment Projects and Current Research Projects which offer alternative ways of assessing student progress in several disciplines, including science. Most of the projects have an interdisciplinary focus.

The Heart of the Matter: Using Standards and Assessment to Learn

This professional development book presents strategies for using standards and assessments to support meaningful learning for all students. It explores issues related to assessment, accountability, and standards-based reform in the context of teaching for understanding and critical inquiry.

Using Data, Getting Results: A Practical Guide for School Improvement in Mathematics and Science

This resource guidebook with CD-ROM is designed to help teachers, administrators, and community members institute reform in school science and mathematics curricula with a focus on improving student learning.

Western States Benchmarking Consortium

The Consortium's strategic framework can help districts conduct self-assessments in four strategic areas: student learning, data-driven decision making, capacity development, and community connectedness.

CONTACTS:

American Association for the Advancement of Science (AAAS)
http://www.aaas.org
1200 New York Avenue NW
Washington, DC 20005
Phone: 202-326-6400

American Educational Research Association (AERA)
http://www.aera.net
1230 Seventeenth Street NW
Washington , D.C. 20036-3078
Phone: 202-223-9485

Educational Testing Service (ETS)
http://www.ets.org
Rosedale Road
Princeton, NJ 08541
Phone: 609-921-9000

National Assessment of Educational Progress (NAEP)
http://nces.ed.gov/nationsreportcard/
National Center for Education Statistics
1990 K Street, NW
Washington , DC 20006
Phone: 202-502-7300

National Association for State Boards of Education (NASBE)
http://www.nasbe.org
277 South Washington Street, Suite 100
Alexandria , VA 22314
Phone: 703-684-4000

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
http://www.cse.ucla.edu/index6.htm

UCLA CSE/CRESST
300 Charles E. Young Drive North
GSE&IS Building Third Floor/Mailbox 951522
Los Angeles, CA 90095-1522
Phone: 310-206-1532

National Council of Teachers of Mathematics (NCTM)
http://www.nctm.org
1906 Association Drive
Reston, VA 20191-1502
Phone: 703-620-9840

National Science Foundation (NSF)
http://www.nsf.org
4201 Wilson Boulevard
Arlington , VA 22230
Phone: 703-292-5111

National Science Teachers Association (NSTA)
http://www.nsta.org
1840 Wilson Boulevard
Arlington VA 22201-3000
Phone: 703-243-7100

References

Opportunity to Learn Under NCLB

Regularly assessing students' understanding and progress can provide data to identify students who may need an additional or alternative opportunity to learn. With regard to assessment, Winfield (1987) notes that opportunity to learn relate to "the provision of adequate and timely instruction of specific content and skills prior to taking a test" (p. 438) and suggests that opportunity to learn can be measured by indicators such as "time spent in reviewing, practicing, or applying a particular concept or by the amount and depth of content covered with particular groups of students" (p. 439). Students who fail tests should be provided meaningful opportunities for remediation that focus on the knowledge and skills on the test, and provide enough time for them to remedy any weakness in that area before retaking the test (American Educational Research Association [AERA], 2000).

More than a decade-old concept, opportunities to learn were proposed to maximize fairness and equity for students (Porter, 1993). Opportunities to learn, historically, were defined as standards that “are to represent what schools and teachers must do if the new curriculum and achievement standards are to be met” (Porter, 1993, p. 1). Later, the *National Science Education Standards* (National Research Council, 1996) claimed that the most powerful indicators of opportunity to learn include teachers' content knowledge, pedagogical know-how, and understanding of students' progress and learning needs. The No Child Left Behind (NCLB) Act (NCLB, 2002) includes opportunity to learn in its accountability structure and specifies that lack of equity in the classroom can result in serious adverse consequences for schools and teachers.

An opportunity to learn for one student may not meet the learning needs of another student in the same classroom. Research shows that students learn using different modalities, at different paces, and with different levels of prior knowledge, motivation, and interest that all impact their learning achievement (APA Task Force on Psychology in Education, 1993; APA Work Group of the Board of Educational Affairs, 1997; Bransford, Brown, & Cocking, 1999; Tomlinson, 1999). In classrooms where teachers are attuned to the range in opportunities to learn that a group of students may need, teachers foster student growth and development on an individualized basis. Opportunities-to-learn requirements affect all students, especially those with established individualized educational programs (IEPs). Although students with IEPs may need more accommodations, a classroom that provides equitable opportunities to learn will provide enough flexibility in its scope and sequence to support all students' progress toward adequate yearly progress requirements.

References

American Educational Research Association. (2000, July). *AERA position statements: H igh-stakes testing in preK-12 education* . Retrieved September 6, 2005, from http://www.aera.net/policyandprograms/?id=378

APA Task Force on Psychology in Education. (1993, January). *Learner-centered psychological principles: Guidelines for school redesign and reform*. Washington, DC: American Psychological Association and Mid-continent Research for Education and Learning.

APA Work Group of the Board of Educational Affairs. (1997, November). *Learner-centered psychological principles: A framework for school reform and redesign.* Washington, DC: American Psychological Association. Retrieved September 6, 2005, from http://www.apa.org/ed/lcp.html

Bransford, J. D., Brown, A. L., Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press. Retrieved September 6, 2005, from http://www.nap.edu/html/howpeople1/

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press. Retrieved September 6, 2005, from http://books.nap.edu/html/nses/

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved September 6, 2005, from http://www.ed.gov/legislation/ESEA02/

Porter, A. (1993). *Brief to policymakers: Opportunity to learn*. Center on Organization and Restructuring of Schools. Retrieved September 6, 2005, from http://www.wcer.wisc.edu/archive/cors/brief_to_principals/BRIEF_NO_7_FALL_1993.pdf

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.

Winfield, L. F. (1987). Teachers' estimates of test content covered in class and first-grade students' reading achievement. Elementary School Journal, 87(4), 438–445.

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

North Central Regional Educational Laboratory LEARNING POINT ASSOCIATES™

GLOSSARY

alternative assessment: An assessment in which students originate a response to a task or question. Such responses could include demonstrations, exhibits, portfolios, oral presentations, or essays. (Compare to traditional assessment.)

authentic assessment: An assessment presenting tasks that reflect the kind of mastery demonstrated by experts. Authentic assessment of a student's ability to solve problems, for example, would assess how effectively a student solves a real problem.

benchmark: A standard or point of reference for measuring and analyzing existing or proposed curricula in the light of content specific goals/standards. Benchmarks often are used in conjunction with standards.

data-driven decision making: A process of making decisions about curriculum and instruction based on the analysis of classroom data and standardized test data. Data-driven decision making uses data on function, quantity, and quality of inputs, and how students learn to suggest educational solutions. It is based on the assumption that scientific methods used to solve complex problems in industry can effectively evaluate educational policy, programs, and methods.

high-stakes testing is "the term used for assessments that determine if a student is retained in a grade or allowed to receive a diploma and graduate" (Lynch, 2000, p. 216).

performance assessment: Direct, systematic observation of an actual student performance or examples of student performances and rating of that performance according to pre-established performance criteria. Students are assessed on the result as well as the process engaged in a complex task or creation of a product.

qualitative research: Collection of nonnumerical data using interviews, observations, and open-ended questions to gather meaning from nonquantified narrative information.

quantitative research: Collection of numerical data in order to describe, explain, predict, and/or control phenomena of interest.

reliability: The degree to which an instrument consistently measures in the same way on repeated trials (e.g., a math test given to a student one day would yield roughly the same score if given to the same student the next day). Reliable assessment is one in which the same answers receive the same score regardless of who performs the scoring or how or where the scoring takes place. The same person is likely to get approximately the same score across multiple test administrations.

standardized tests: A test that is administered under controlled conditions that specify where, when, how, and for how long students may respond to the questions or "prompts." Assessments are administered and scored in exactly the same way for all students. Traditional standardized tests are typically mass-produced and machine-scored; they are designed to measure skills and knowledge that are thought to be taught to all students in a fairly standardized way. Performance assessments also can be standardized if they are administered and scored in the same way for all students.

summative assessment: An assessment at the end of a program period determining the overall quality of a program's results.

traditional assessment: An assessment in which students select responses from a multiple-choice list, a true/false list, a matching list, or work out the full solution of an equation, set out the proof in a geometry problem, etc. (Compare to alternative assessment.)

validity of an instrument: The degree to which a measure accurately assesses the specific concept it is designed to measure, excluding extraneous features from such measurement (e.g., whether a reading-comprehension assessment focuses on students' understanding of a story or their ability to read the story).

Reference

Lynch, S. J. (2000). *Equity and science education reform*. Mahwah, NJ: Lawrence Erlbaum Associates.

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

References

American Association for the Advancement of Science [AAAS], Project 2061. (1993). *Benchmarks for science literacy*. New York: Oxford University Press. Retrieved September 6, 2005, from http://www.project2061.org/publications/bsl/default.htm

American Educational Research Association. (2000, July). AERA position statements: High-stakes testing in prek–12 education. Retrieved September 6, 2005 , from http://www.aera.net/policyandprograms/?id=378

Amrein, A. L., & Berliner, D. C. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved September 6, 2005, from http://epaa.asu.edu/epaa/v10n18/

Anderson-Inman, L., Ditson, L. A., & Ditson, M. T. (1999, April). Computer-based concept mapping: Promoting meaningful learning in science for students with disabilities. *Information Technology and Disabilities, 5*(1–2). Retrieved September 6, 2005, from http://www.rit.edu/~easi/itd/itdv5n12/article2.htm

Atkin, J. M., Black, P., & Coffey, J. (Eds.). (2001). *Classroom assessment and the National Science Education Standards*. Washington , DC : National Academies Press.

Barchfeld-Venet, P. (2005). Formative assessment: The basics. *Alliance Access, 9*(1), 2–3. Retrieved September 6, 2005, from http://ra.terc.edu/publications/Alliance_Access/Vol9-No1/w05.pdf

Bennett, R. E. (1998). *Reinventing assessment: Speculations on the future of large-scale educational testing*. Princeton, NJ: Educational Testing Service Policy Information Center.

Bennett, R. E. (2003). Inexorable and inevitable: The continuing story of technology and assessment. *The Journal of Technology, Learning, and Assessment, 1*(1). Retrieved September 6, 2005 , from http://www.bc.edu/research/intasc/jtla/journal/v1n1.shtml

Black, P., and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148. Retrieved September 6, 2005, December 8, 2003, from http://www.pdkintl.org/kappan/kbla9810.htm

Bransford, J. D., Brown, A. L., Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Braun, H. (2001). *Assessment and evaluation issues in the age of e-learning* (Preparing Tomorrow's Teachers to Use Technology Program Technical Report). Washington, DC: U.S. Department of Education.

Burmaster, E. (2003, October). *Grow Wisconsin: Invest in people initiative.* Panel discussion at the Wisconsin Economic Summit IV in Milwaukee, WI.

Burstein, J., Marcu, D., Andreyev, S., & Chodorow, M. (2001, July). *Towards automatic classification of discourse elements in essays.* Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France. Retrieved September 6, 2005, from http://acl.ldc.upenn.edu//P/P01/P01-1014.pdf

Bybee, R. W., & Kennedy, D. (2005). Math and science achievement. *Science, 307*(5709), 481.

CEO Forum on Education and Technology. (2001). *School technology and readiness report: Key building blocks for student achievement in the 21st century*. Retrieved September 6, 2005, from http://www.ceoforum.org/downloads/report4.pdf

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). NY: Academic Press.

Commission on Instructionally Supportive Assessment. (2001). *Building tests to support instruction and accountability.* Retrieved September 6, 2005, from http://www.nea.org/accountability/buildingtests.html

Consortium for School Networking. (2003). *Vision to know and do: The power of data as a tool in educational decision making.* Washington, DC: Author.

Domenech, D. A. (2000, December). My stakes well done: The issue isn't academic benchmarks, it's the misguided use of a single test. *School Administrator Web Edition*.

Duschl, R. D., & Gitomer, D. H. (1997). Strategies and challenges to change the focus of assessment and instruction in science classrooms. *Educational Assessment*, *4*(1): 37–73.

Foertsch, D. J. (1999). *Understanding assessment: An introduction to using published tests and developing classroom tests*. Unpublished manuscript, North Central Regional Educational Laboratory, Oak Brook, IL.

Foltz, P. W., Gilliam, S., & Kendall, S. A. (2000). Supporting content-based feedback in online writing evaluation with LSA. *Interactive Learning Environments*, *8*(2), 111–129. Retrieved September 6, 2005, from http://www.k-a-t.com/papers/ILE_foltz2000.pdf

Fontana , D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, *64*(3): 407–417.

Frederiksen, J. R., & White, B. J. (1997, March). *Reflective assessment of students' research within an inquiry-based middle school science curriculum.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Haertel, G., & Means, B. (2000). *Stronger designs for research on educational uses of technology: Conclusions and implications*. Menlo, CA: SRI International.

Herman, J. L. (1997, October). *Large-scale assessment in support of school reform: Lessons in the search for alternative measures.* (CSE Technical Report 446.) Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved September 6, 2005, from http://www.cse.ucla.edu/CRESST/Reports/TECH446.pdf

Heubert, J. P. & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation. Washington, DC: National Academy Press. Retrieved September 6, 2005, from http://www.nap.edu/html/highstakes/*

Hickey, D. T., Kindfield, A., Horwitz, P., & Christie, M. A. (2003). Integrating curriculum, instruction, assessment and evaluation in a technology-supported genetics learning environment. *American Educational Research Journal, 40*(2), 495–538.

Joint Committee on Testing Practices. (2004). C*ode of fair testing practices in education*. Washington, DC: Joint Committee on Testing Practices. Retrieved September 6, 2005, from http://www.apa.org/science/FinalCode.pdf

Kintsch, E., Steinhart, D., Stahl, G., & LSA Research Group. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments, 8*(2), 87–109. Retrieved September 6, 2005, from http://lsa.colorado.edu/papers/ekintschSummaryStreet.pdf

McAninch, A. R. (1993). *Teacher thinking and the case method: Theory and future directions*. New York: Teachers College Press.

McCurdy, B. L., & Shapiro, E. S. (1992). A comparison of teacher monitoring, peer monitoring, and self-monitoring with curriculum-based measurement in reading among students with learning disabilities. *Journal of Special Education*, *26*(2): 162–180.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academies Press.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academies Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved September 6, 2005, from http://www.ed.gov/legislation/ESEA02/

North Central Regional Educational Laboratory & Metiri Group. (2003). *enGauge 21st century skills for 21 st century learners: Literacy in the Digital Age*. Naperville , IL : North Central Regional Educational Laboratory. Retrieved September 6, 2005 , from http://www.ncrel.org/engauge/skills/skills.htm

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment, Board on Testing and Assessment, Center for Education, National Research Council. Washington , DC : National Academy Press. Retrieved September 6, 2005, from http://www.nap.edu/books/0309072727/html/

Reich, R. B. (2001, May). Drop your standards. *The American Prospect Online*. Retrieved September 6, 2005, from http://www.prospect.org/webfeatures/2001/05/reich-r-05-11.html

Rose, D. H. & Meyer, A. (2002). *Teaching every student in the Digital Age: Universal design for learning*. Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved September 6, 2005, from http://www.cast.org/teachingeverystudent/ideas/tes/

Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (May, 2002). *Benchmarking and alignment of standards and testing.* (CSE Technical Report 566). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved September 6, 2005, from http://www.cse.ucla.edu/CRESST/Reports/TR566.pdf

Russell, M. (2000). *It's time to upgrade: Tests and administration procedures for the new millennium.* Secretary's Conference on Educational Technology 2000. Retrieved September 6, 2005, from http://www.ed.gov/rschstat/eval/tech/techconf00/russell_paper.html

Russell, M., & Haney, W. (1997, January). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, *5*(3). Retrieved September 6, 2005 , from http://olam.ed.asu.edu/epaa/v5n3.html

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.

Sawyer, R. J., Graham, S., & Harris, K. R. (1992, September). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology*, *84*(3): 340-352.

Seltzer, M., Choi, K., & Thum, Y. M. (April, 2002). *Examining relationships between where students start and how rapidly they progress: Implications for constructing indicators that help illuminate the distribution of achievement within schools.* (CSE Technical Report 560). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved September 6, 2005, from http://www.cse.ucla.edu/CRESST/Reports/TR560.pdf

Senge, P., Cambron-McCabe, N., Lucas, T., Smith, B., Dutton, J., & Kleiner, A. (2000). *Schools that learn*. New York: Doubleday.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29(7)*, 1–14. Retrieved September 6, 2005 , from http://35.8.171.42/aera/pubs/er/arts/29-07/shep01.htm

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan 8*1(3), 191–198. Retrieved September 6, 2005, from http://www.pdkintl.org/kappan/k9911sti.htm

Stiggins, R. J. (2004, September). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86*(1), 22–27.

U. S. Department of Education, Office of Elementary and Secondary Education. (2002). *No Child Left Behind: A desktop reference*. Washington , DC : Author. Retrieved September 6, 2005 , from http://www.ed.gov/admins/lead/account/nclbreference/index.html

Vendlinski, T., & Stevens, R. (2003). Assessing student problem-solving skills with complex computer-based tasks. *The*

Journal of Technology, Learning, and Assessment, 1 (3). Retrieved September 6, 2005 , from http://www.bc.edu/research/intasc/jtla/journal/v1n3.shtml

Wolf, F. M. (1986) *Meta-analysis: Quantitative methods for research synthesis*. SAGE University series on quantitative applications in the social sciences, series no. 07-059. Newbury Park , CA : SAGE.

Wright, A. W. (2001, October). The ABCs of assessment. *The Science Teacher*, 60–64. Retrieved September 6, 2005, from http://science.nsta.org/enewsletter/2004-03/tst0110_60.pdf

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Get Adobe Reader   Adobe Reader FAQ

Content and general comments: info@ncrel.org   Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

Defining Adequate Yearly Progress (AYP)

The No Child Left Behind (NCLB) Act (NCLB, 2002) builds upon the accountability provisions in the Improving America's Schools Act (IASA) of 1994 (IASA, 1994), which required each state to establish challenging content and performance standards and to implement assessments that measure students' performance against those standards (Goertz, 2001). The IASA defined adequately yearly progress (AYP):

"In a manner that 1) results in continuous and substantial yearly improvement of each school and local education agency sufficient to achieve the goal of all children … meeting the state's proficient and advanced levels of achievement; [and] 2) is sufficiently rigorous to achieve the goal within an appropriate timeframe (as cited in Elmore & Rothman, 1999, p. 85)" (Goertz, 2001).

Goertz (2001) explains that the NCLB legislation makes several critical changes to the IASA definition for AYP and requires each state to create its own definition of AYP within the parameters set by Title I. NCLB states that each state is required to define AYP in a manner as follows:

"(i) Applies the same high standards of academic achievement to all public elementary school and secondary school students in the State; (ii) is statistically valid and reliable; (iii) results in continuous and substantial academic improvement for all students; (iv) measures the progress of public elementary schools, secondary schools and local educational agencies and the State based primarily on the academic assessments ... (v) includes separate measurable annual objectives for continuous and substantial improvement for each of the following: (I) The achievement of all public elementary school and secondary school students. (II) The achievement of—(aa) economically disadvantaged students; (bb) students from major racial and ethnic groups; (cc) students with disabilities; and (dd) students with limited English proficiency" (NCLB, 2002, Part A, Subpart 1, Sec. 1111, 2[c]).

As a result of NCLB, each state has developed a plan for the minimum levels of improvement in measurable terms of student performance that local educational agencies must achieve within the given time frames specified by the NCLB legislation.

Linn, Baker, and Betebenner (2002) explain experts' concerns about the NCLB definition and timelines for getting students and all subgroups of students to the "proficient" standard and for identifying and sanctioning low-performing schools under the federal definition of AYP. Of particular concern are the validity and reliability of the tests used for high-stakes decisions such as high school graduation, college entrance, or determining the consequences of NCLB's AYP requirement. Sample error and nonpersistent factors such as teacher turnover or strikes and disruptive student cohorts are large contributors to the variance in school-building scores from one year to the next (Cronbach, Linn, Brennan, & Haertel, 1997; Linn, Baker, & Betebenner, 2002). Hinging students' future opportunities on one standardized test score, when the test itself is unaligned to content standards, is a common example of inappropriate test use. High-stakes decisions need to be founded on a robust data set about students' achievements. Because the standards' movement, technology, and the NCLB legislation are spawning new tests at the state level, test precision is a vital concern.

The Education Commission of the States has created an interactive, online database that catalogs each state's status with regard to the NCLB requirements and allows for interactive searches and comparisons among states according to each of the accountability criteria including the definition of AYP.

References

Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement, 57*, 373–399.

Goertz , M.E. (2001, September). *The federal role in defining "adequate yearly progress:" The flexibility/accountability trade-off.* Consortium for Policy Research in Education. Retrieved September 6, 2005 , from http://www.cpre.org/Publications/cep01.pdf

Improving America's Schools Act of 1994, Pub. L. No. 103-382. (1994). Retrieved September 6, 2005, from http://www.ed.gov/legislation/ESEA/toc.html

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), 3–16. Retrieved September 6, 2005, from http://www.aera.net/uploadedFiles/Journals_and_Publications/Journals/Educational_Researcher/3106/3106_Linn.pdf

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425. (2002). Retrieved September 6, 2005, from http://www.ed.gov/legislation/ESEA02/

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

Barbara Campbell: Checking for Student Understanding

Barbara Campbell

Veteran Elementary Teacher

Video Transcript:
All right, children need to be able to explain relationships—verbally as well as in written form—well. And that could include a journal, whatever. Whether I'm teaching problem solving, number sense, number operations, measurement, algebra, data collecting analysis, or probability, it's important that the child understands the concepts well enough to explain them clearly. It could be verbally, be informal, it could be at an end of the unit test. It could be a quiz, lots and lots and lots of ways. I assess informally daily through modeling and checking for understanding. When I get verbal feedback that the children are not understanding of the concept, I thankfully have students who will tell me honestly: I didn't get that. That is a really good way to assess right at that point when they're in the process of learning and for them to willingly want to learn, it's my job to help them try to figure that out and come up with creative ways to teach it so that they do understand it.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Gilbert Valdez: Learning and Assessment

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
I think states need to recognize that learning is about big concepts and something that has meaning for a lifetime and not something that you memorize on Thursday and regurgitate back on Friday. What we're talking about is using assessments that test big ideas and … big understanding rather than something that allow students to guess which answer is least wrong through tests-taking skills and not having any knowledge of content.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Gilbert Valdez: 21st Century Skills Requirements

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
I think the main thing is that they have to understand 21st century skills are calling for how do you go about creating new knowledge, new understandings, not so much memorizing and knowing old knowledge. So it's about the process of learning and so the assessment has to be about finding solutions where there isn't one, answers where there aren't ready ones, where you have to get them from multiple sources. I think what teachers have to understand is that formative assessment for the 21st century requires a flexibility in what are the sources, what are the possibilities, possible answers that there are alternative solutions, alternative strategies, and alternative responses and they're accurate. It's … divergence rather than convergence in responses and so the assessments have to reflect that. They have to allow students to give you different ways of finding the right answer than the one that maybe you had thought for in the final analysis it makes sense and it's the correct answer. How, the pathways to getting there aren't nearly as important as what's the quality of the response.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Gilbert Valdez:
Summative Versus Formative Assessment

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
Well I think you have to understand that the two tests have two different, purposes and … intents. High-stakes testing's purpose is to prove that students have some knowledge over a given amount of time. Formative assessment is intended to improve learning and it's used to determine how you go about changing teaching and teaching strategies so that the learning is improved and made more efficient and more effective.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Gilbert Valdez:
Two Functions of Assessments

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
I think that it's the two functions of assessment that we have to really understand. They're both important. The community, the people who are paying the bills, need to have evidence that in fact there is learning. That's the high stakes, the, accountability to prove that learning is taking place. But if they're going to be really good at learning and teaching they're going to have to use formative assessment. One of the things that most people don't know is that formative assessment, classroom formative assessment, actually has some of the greatest effect sizes over any intervention that teachers can use. Effect size of .70, which is almost the highest when you figure that the education of the mother has got one of the highest levels of, and it's around .56 rather and .60. And so what we're saying is that a teacher who is really good at formative assessment can almost match some of the things that are taking place at home in terms of what they can do. So it's one of the few interventions where they can start to match what's taking place in the home in a classroom. It's something that every teacher needs to be able to do well.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Assessment and Science Standards

There are several purposes of assessment in science education, and national science standards help specify them. Of the 28 national science standards five are assessment oriented:

- **Standard A:** Assessments must be consistent with the decisions they are designed to inform (NRC, 1996, p. 78).
- **Standard B:** Achievement and opportunity to learn science must be assessed (NRC, 1996, p. 79).
- **Standard C:** The technical quality of the data collected is well matched to the decisions and actions taken on the basis of their interpretation (NRC, 1996, p. 83).
- **Standard D:** Assessment practices must be fair (NRC, 1996, p. 85).
- **Standard E:** The inferences made from assessments about student achievement and opportunity to learn must be sound (NRC, 1996, p. 86).

The data collected from assessments is meant to inform states, districts, and schools as well as help teachers improve instruction. In order to improve assessment of student learning, many types of assessment should be considered (Wright, 2001, p. 60). For example, a new approach to assessment is to base test design and content on a continuum of expertise that can measure levels of knowledge transfer from classroom science lessons to standards-based, high-stakes tests (Atkin, Black, & Coffey, 2001; Hickey, Kindfield, Horwitz, & Christie, 2003; Pellegrino, Chudowsky, & Glaser, 2001).

Normally, science content standards are meant to be curriculum guidelines that specify the central concepts and skills students need to master within the discipline of science. However, many state-level standards are either too vague or too detailed to be useful in this regard (Rothman, Stattery, Vranek, & Resnick, 2002). Assessment designers need to review these guidelines to identify indicators that can be used to acknowledge growth of student content expertise. More constructive than a fact-based approach to science, instruction and assessment that focuses on gateway science concepts, principles, and methods can provide students with opportunities to acquire a deep understanding of science and transfer this understanding from one science lesson to another and beyond.

Supporting such growth opportunities, appropriate assessment is critical in measuring, monitoring, and improving student progress. Assessment may be formative or summative (or a combination of both) or diagnostic. Formative assessment is a tool to monitor and map students' day-to-day progress in multiple problem solving. Summative assessment, on the other hand, provides a basis for reporting student grades or meeting academic standards. In other words, it is a snapshot of student learning collected on a given day to gather and report data. Diagnostic assessment is usually given before instruction to help determine students' prior understanding of topics before instruction (Wright, 2001, p. 60).

Among the National Science Education Standards for assessment in science is a call for a match between the technical quality of the student data collection processes and the consequences of the actions taken or decisions made on the basis of those data (National Research Council, 1996). Data-based decision making is a critical component of carrying out the requirements of NCLB. In science and in other disciplines, it can ultimately improve outcomes for all children when both formative and summative achievement measures are used to drive instruction, individualize opportunities to learn, and fine tune the organization of schooling (Consortium for School Networking, 2003).

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

Assessment and Mathematics Standards

The National Council of Teachers of Mathematics has outlined principles and standards as well as assessment standards for the discipline of mathematics. Their Principles and Standards for School Mathematics (NCTM, 2000) reflects a concern that students in the United States are not mathematically literate and often fail to see the relationship between mathematics learned in school and applications to real-life situations. The content and process standards are organized into grade bands: PK–2, 3–5, 6–8, and 9–12. Schools and communities, with guidance from the standards, must determine how to embed these principles into their schools' mathematics instruction. Forty-nine of the 50 states have adopted state-level standards in mathematics.

NCLB strengthens Title I accountability by requiring states to implement statewide accountability systems covering all public schools and students. These systems must be based on challenging state standards in reading and mathematics, annual testing for all students in Grades 3–8, and annual statewide progress objectives ensuring that all groups of students reach proficiency by 2014. National Assessment of Educational Progress results show the extent to which students can solve problems using their mathematical knowledge. Since 1990, NAEP mathematics assessments have placed increasing emphasis on fostering mathematical knowledge and power. The 1996, 2000, and 2003 nationwide mathematics assessments focused on reasoning and communication by requiring students to connect their learning across mathematical content strands. Results of the 2003 Mathematics National Center for Education Statistics show that overall mathematics scores are improving, especially in the fourth grade.

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org

Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

Fair Use of High-Stakes Testing

Recently, the Joint Committee on Testing Practices (2004) published a draft of the *Code of Fair Testing Practices in Education*. The draft provides guidance for test users and test developers in four critical areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers.

Experts' position statement on high-stakes testing (American Educational Research Association [AREA], 2000) warns that there is potential harm if high-stakes testing programs are carried out without adequate educational resources available to schools or without sufficient reliability and validity:

"Policy makers and the public may be misled by spurious test score increases unrelated to any fundamental educational improvement; students may be placed at increased risk of educational failure and dropping out; teachers may be blamed or punished for inequitable resources over which they have no control; and curriculum and instruction may be severely distorted if high test scores per se, rather than learning, become the overriding goal of classroom instruction."

Experts believe there are several conditions essential for ensuring sound implementation of high-stakes educational testing programs for students. The conditions include the following (AERA, 2000):

- Protection against high-stakes decisions based on a single test.
- Adequate resources and opportunity to learn the tested content and cognitive processes.
- Validation for each separate intended use of the test (e.g., individual certification, school evaluation, curricular improvement, increasing student motivation).
- Full disclosure of likely negative consequences of high-stakes testing programs when credible scientific evidence suggests potential side effects.
- Alignment between the test and the full breath and depth of the curriculum and related standards.
- Validity of passing scores and achievement levels must be explicitly demonstrated.
- Opportunities for meaningful remediation for examinees who fail high-stakes tests with sufficient time provided for learning in areas of weakness.
- Appropriate attention to language differences among examinees when interpreting test results and to accommodations for English language learners.
- Appropriate attention to students with disabilities when interpreting test results and accommodations according to special needs of the learner.

- Careful adherence to explicit rules for determining which students are to be tested according to district, state, and federal testing policies aligned with NCLB.

- Sufficient reliability for each intended use and interpretation of test scores.

- Ongoing evaluation and research of intended and unintended effects of high-stakes testing should be public knowledge.

References

American Educational Research Association. (2000, July). *AERA position statements: H igh-stakes testing in preK-12 education* . Retrieved September 6, 2005, from http://www.aera.net/policyandprograms/?id=378

Joint Committee on Testing Practices. (2004). C*ode of fair testing practices in education*. Washington, DC: Joint Committee on Testing Practices. Retrieved September 6, 2005, from http://www.apa.org/science/FinalCode.pdf

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

North Central Regional Educational Laboratory LEARNING POINT ASSOCIATES™

Using Progress Maps to Organize Student Assessment Data

When traditional assessment practices fail to provide the type of feedback students need to understand their mistakes or misconceptions so that they can continue to improve, the focus should be on classroom-based assessments that will yield the feedback students need to organize their learning goals. This can be accomplished by doing the following:

- Using benchmarks as the basis for rubric criteria.
- Identifying levels of transfer as the basis for levels of expertise on a rubric.
- Plotting scores over time on individual student's progress maps and providing customized feedback according to levels of expertise or development.
- Plotting aggregated student scores on progress maps and designing future opportunities to learn according to clusters and deviations in the scores.
- Embedding progress maps in students' electronic portfolios to inform their self-reflection and personal goal-setting.

Progress maps, a technique new to most educators, can incorporate rubrics and help organize longitudinal assessment data within students' electronic portfolios based on continuums of expertise in content-area domains. They "provide a description of skills, understandings, and knowledge in the sequence in which they typically develop—a picture of what it means to improve over time in an area of learning" (Pellegrino, Chudowsky, & Glaser, 2001, p. 190).

The basis for designing progress maps is a developmental continuum of increasing expertise in a content area. Progress maps are set up on an individualized student basis with data from an initial baseline assessment about that student's level of expertise or status of his/her knowledge and skills in relationship to a set of related standards, say at the beginning of the year. Formative assessments using benchmark criteria that follow a developmental continuum provide subsequent data throughout the year to plot on the progress map and provide a quick visual for monitoring student's growing expertise pertaining to the content-area standard(s).

The progress-mapping technique is most efficient when technology can be used to collect, organize, and store the data via electronic rubrics, graphing tools, and e-portfolios. Analysis of standards reveals that not all standards deserve attention on a developmental continuum, but rather benchmarks need to focus on the key concepts, knowledge, and skills that enable students to progress (Porter, 2002).

R

References

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment, Board on Testing and Assessment, Center for Education, National Research Council. Washington , DC : National Academy Press. Retrieved on September 6, 2005, from http://www.nap.edu/books/0309072727/html/

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31* (7), 3–14. Retrieved September 6, 2005 , from http://35.8.171.42/aera/pubs/er/pdf/vol31_07/AERA310702.pdf

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org

Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

North Central Regional Educational Laboratory LEARNING POINT ASSOCIATES™

Using Concept Mapping as an Assessment Method

Concept mapping is a powerful tool for linking knowledge and could be a key to developing strong performance assessments that ought to be designed to generate both an assessment of how students are applying concepts and to assess the deep understanding that students are gaining.

According to Plotnick (1997), researchers began to study the concept-mapping technique in the 1960s based on the theories of Ausubel (1968), who stressed the important role of prior knowledge in learning about new concepts. Since then, concept mapping has been gaining inroads as a tool to enhance problem solving in education.

Mintzes, Wandersee, and Novak (2000) define concept maps as "tools for organizing and representing knowledge." Pellegrino, Chudowsky, & Glaser (2001) identify concept mapping as a method for arranging conceptual nodes—and labels of nodes and links—to show relationships among multiple concepts in a domain . An important benefit of using concept mapping as an assessment method is its ability to detect or illustrate students' deep content understandings as well as their misconceptions when they create a personal explanation of content matter (Novak, Gowin, & Johansen, 1983; Novak, 1993).

In 1989, the American Association for the Advancement of Science began Project 2061 and issued a set of recommendations for mathematics, science, and technology through its 1990 publication of *Science for All Americans* and 1993 publication of *Benchmarks for Science Literacy*. Project 2061 is the group's long-term initiative to reform K–12 science, mathematics, and technology education nationwide. The team of Project 2061 and the National Science Teachers Association copublished *Atlas of Science Literacy*, a collection of conceptual strand maps that show how students' understanding of the ideas and skills that lead to literacy in science, mathematics, and technology might grow over time. Each map depicts how K–12 learning goals for a particular topic relate to each other and progress from one grade level to the next.

According to research cited by Novak (n.d.), besides being a learning tool, concept mapping is also one of the most powerful evaluation tools, "encouraging students to use meaningful-mode learning patterns" (Novak & Gowin, 1984; Novak, 1990; Mintzes, Wandersee, & Novak, 2000). According to Plotnick (1997), one-way, two-way, or nondirectional links can be established between the concepts, which may be categorized along with the links and may show temporal or causal relationships between concepts.

Specifications for concept-mapping assessments usually include (1) defining the task demands in terms of a given set of concepts or terms within a content area and electronic information links that are provided to the student in a finite database and (2) scoring the concept maps using the Herl metric for scoring semantic content, organizational structure, number of links used in the map, and comparison with expert teacher maps that serve as criterion maps (Herl, Baker, & Niemi, 1996; O'Neil & Schacter, 1997). Several computer software programs allow users the flexibility of moving concepts together with linking statements or moving concept groups and links on a specific concept map (Novak, n.d.). Printing capability of these software programs gives students an opportunity to produce a product and share it with classmates or the teacher.

Inspiration® and Kidspiration® are currently among the most popular concept mapping software programs. Zeitz and Anderson-Inman (1992) conducted a hallmark study of classroom use of Inspiration and found that it encourages students to reflect and revise their understanding of conceptual relationships representations, more than when compared to students' maps drawn with paper and pencil. In follow-up studies, Zeitz and Anderson-Inman (1993) found that Inspiration is useful in stimulating students' prior knowledge at the start of a unit of study and that teachers used "the concept maps and outlines generated by the students to assess the level of student comprehension and to correct misconceptions that became apparent as the students entered and linked new concepts as they learned" (Kight, 1998).

Anderson-Inman, Ditson, and Ditson (1999) cite considerable evidence that concept mapping promotes meaningful learning in science. In science education, concept mapping has been widely recommended and used in a variety of ways to observe change in students' understanding of concepts over time, to assess what the learner knows, and to reveal their unique thought processes. It has been used in evaluation of science curriculum and instructional activities for promoting conceptual understanding and positive learner attitudes toward science. When Anderson-Inman, Ditson, and Ditson (1999) investigated the use of concept mapping as an accommodation strategy for students with learning difficulties, they found that computer-based concept mapping assists students who are oriented toward visual learning (or who have difficulty reading and writing text) to graphically represent what they are learning. Students' graphic representation of science concepts can be tracked over time, which provides teachers with a method for monitoring conceptual growth as a function of instruction.

References

Anderson-Inman, L., Ditson, L. A., & Ditson, M. T. (1998, June). Computer-based concept mapping: Promoting meaningful learning in science for students with disabilities. *Information Technology and Disabilities, 5*(1–2). Retrieved September 6, 2005, from http://www.rit.edu/~easi/itd/itdv5n12/article2.htm

Ausubel, D. (1968). *Educational psychology: A cognitive view*. New York : Holt, Rinehart, and Winston.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive science structure of U.S. history knowledge. *Journal of Educational Research*, 89, 206–218.

Kight, K. S. (1998). *Using Inspiration to organize reading and writing*. Retrieved September 6, 2005, from http://www.strategictransitions.com/ldarticle.htm

Mintzes, J., Wandersee, J., & Novak, J. (Eds.). (1999). *Assessing science understanding*. San Diego: Academic Press.

Novak, J. D. (n.d.). The theory underlying concept maps and how to construct them. Retrieved September 6, 2005, from http://cmap.coginst.uwf.edu/info/printer.html

Novak, J. D. (1990). Concept maps and Vee diagrams: Two metacognitive tools for science and mathematics education. *Instructional Science,* 19, 29-52.

Novak, J. D. (1993, March). How do we learn our lesson?: Taking students through the process. *The Science Teacher, 60* (3), 51–55.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York and Cambridge, UK: Cambridge University

Press.

Novak, J. D., Gowin, D. B., & Johansen, G. T. (1983). The use of concept mapping and knowledge: Vee mapping with junior high school science students. *Science Education*, *67*(5), 625–645.

O'Neil, H., Jr., & Schacter, J. (1997). Test specifications for problem-solving assessment. *CSE Technical Report 463*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved September 6, 2005, from http://www.cse.ucla.edu/CRESST/Reports/TECH463.PDF

Pellegrino, J. W., Naomi Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment, Board on Testing and Assessment, Center for Education, National Research Council. Washington, DC: National Academy Press. Retrieved on September 6, 2005, from http://www.nap.edu/books/0309072727/html/

Plotnick, E. (1997, June). Concept mapping: A graphical system for understanding the relationship between concepts. ERIC Digest. Syracuse , NY : ERIC Clearinghouse on Information and Technology. (ERIC Document Reproduction Service No. ED407938). Retrieved September 6, 2005, from http://www.ericdigests.org/1998-1/concept.htm

Zeitz, L., & Anderson-Inman, L. (1992). *The effects of computer-based formative concept mapping on learning high school science.* Paper presented at the American Educational Research Association, San Francisco , CA .

Zeitz, L., & Anderson-Inman, L. (1993). *Computer-based concept mapping in a high school biology class: Effects of student characteristics.* Paper presented at the American Educational Research Association, Atlanta , GA.

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org

Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

North Central Regional Educational Laboratory LEARNING POINT ASSOCIATES™

Why Portfolios in Science and Mathematics?

The use of portfolios is one of the educational innovations that can help round out traditional assessment techniques. A portfolio (or a work folio) is an organized collection of a student's work that features the very best that he or she can do. It "documents conceptual understanding, problem solving, reasoning, and communication abilities" (Northwest Regional Educational Laboratory, 1997, p. 19). Most teachers who use portfolios have students organize and display work in language arts, mathematics, and art for inclusion in their portfolios. More and more teachers also are considering the use of portfolios in science as one way to get a fuller picture of students' progress, interests, and attitudes.

While each piece placed in a portfolio can be assessed with respect to the degree to which the student achieved specific unit objectives, the portfolio as a whole will illustrate the student's progress and specific interests. In the discipline of science, specific examples of a student's science work should go in his or her science portfolio, including the following:

- Written observations and science reports.
- Drawings, charts, and graphs that are the products of hands-on discovery activities.
- Reaction pieces such as prepared written responses to science software, videos, discovery experiences, and field trips.
- Media products such as student-produced science work in audio, video, or digitized form.

As an assessment tool, portfolios—either electronic or paper—can allow students to fully feature their learning over time. Mahood (n.d.) outlines several benefits of requesting and encouraging students to keep a portfolio. He states that portfolios do the following:

- Invite the revisiting of concepts through different learning modalities.
- Encourage interdependency and responsibility of all students in a group setting.
- Allow students to use fully their creative energies and potentials.
- Provide support for experimentation and risk taking.
- Improve critical thinking and evaluation skills.
- Provide for a successful subject-matter Web site says scientific not subject matter experience.

An example of an assessment portfolio can be found at http://assist.educ.msu.edu/ASSIST/classroom/assessment/section2/toolasstportfolios.htm

References

Mahood, J. (n.d.). *An introduction to science portfolios.* Retrieved September 6, 2005 , from http://www.accessexcellence.com/21st/TL/mahood_port.html

Northwest Regional Educational Laboratory. (1997). Assessment strategies to inform science and mathematics instruction. *It's Just Good Teaching* series. Retrieved September 6, 2005 , from http://www.nwrel.org/msec/images/resources/justgood/06.97.pdf

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

Gilbert Valdez:
Collaboratively Designed Classroom-Based Assessments

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
It's one of the greatest forms of professional development and … the reason for that is that it requires teaches to think together how do we, what meaning do we have for this concept, what is the content knowledge, how will we know if students know that content? And it also creates common understanding of the content among the teachers. Sometimes two teachers side by side will have similar language but a different understanding of what that means. When you start developing an assessment, it requires you to really explain and almost negotiate what content is. And that's not necessary at any other time except when you're trying to develop what is the real thing that's valuable and meaningful in this concept that we're trying to teach.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Gilbert Valdez: Gifted Students

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
Well … it's quite, quite interesting because gifted students often times are overlooked … in classrooms, especially with the intent now of achievement gaps and everything else. But if you look at the statistics, gifted students have a much higher drop-out rate and difficulty in, sometimes, behavior. And that's because we're not challenging them. I mean for them, in many cases, the norm that we've set for this standard is … easily accomplished on Monday and what do you do Tuesday, Wednesday, Thursday, and Friday? I think it's the enrichments, the ability to provide extensions which also have, a different kind of assessment, assessments that give you greater depth, allow you to test for a greater understanding for new learning, creation of knowledge by these very bright students.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

Gilbert Valdez:
Understanding Validity and Reliability in Assessment

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
Well, validity and reliability are, have, a special meaning in statistics. And often times the reason they are confused is because they have more similar meanings in common language. Validity is nothing more than: Does the test measure what you're intended to use that instrument for? If you're talking about reading comprehension, does it test reading comprehension? It doesn't matter, if it's math knowledge, whatever it is. That's validity. In other words, how good is that instrument to give you an explanation and knowledge about what students do and don't know? Reliability is basically: Can that instrument be used reliably in other settings by different people? In other words, if you go to a different classroom and I go to a different classroom and the students are "knowledge," would the fact that I'm administrating the test versus you give you a different thing? If a test shows the same results with different people administering it in different settings, then it's reliable. So the concept of interrelated reliability that you often hear implies that regardless of who is giving the test, it's got reliability between the settings. So validity is basically about: Is a test doing what it's intended to do? And reliability is: Will it do, be consistent, across different settings and different people administering the test?

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.

North Central Regional Educational Laboratory LEARNING POINT ASSOCIATES™

Using Assistive Technology for Assessment Accommodations

On a daily basis, teachers typically are the ones to determine when accommodations are needed to ensure fair and equitable testing for students with special needs. They also are often charged with selecting appropriate accommodations for individual students and need to have a basic understanding of assistive technologies available to provide accommodations.

Assistive technologies are those that support students with disabilities, of which a total of 6.5 million were being served in 2002 through the Individuals with Disabilities Education Act (IDEA) of 1997 (President's Commission on Excellence in Special Education, 2002). IDEA defines an assistive technology device as "any item, piece of equipment, or product system, whether acquired commercially off the shelf, modified, or customized, that is used to increase, maintain, or improve functional capabilities of a child with a disability" (IDEA, 1997).

Assistive technology may be virtually any device that increases, maintains, or improves a functional capability of a student with a disability. A list of assistive devices arranged in order from lower to higher technology is available at a Web site developed by the Technology and Media Division of the Council for Exceptional Children and the Wisconsin Assistive Technology Initiative. NCREL's *Pathways to School Improvement* Web site features a Critical Issue on enhancing system change and academic success through assistive technologies. Another research-based article explores how technology use can improve the literacy skills of students with disabilities.

When selecting or reviewing technology for assessment accommodation purposes, Rose and Meyer (2002) explain the importance of focusing on the goal of the assessment and separating out the tangential variables of the technology and its supports. They claim it is more accurate to include a learner's daily support as an accommodation during assessment when the goal of the assessment is not undermined by the technology (e.g., using a speech-to-text application when assessing students' conceptual science knowledge). Examples of types of specialized software that are available for assessment accommodation purposes are as follows:

- Screen reader
- Screen enlarger
- Text-to-speech
- Voice recognition
- Word prediction
- Conceptual mapping
- Word processor

References

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C. 1401. (1997). Retrieved September 6, 2005 , from http://www4.law.cornell.edu/uscode/html/uscode20/usc_sec_20_00001401----000-.html

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families.* Retrieved September 6, 2005 , from http://www.ed.gov/inits/commissionsboards/whspecialeducation/reports/index.html

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the Digital Age: Universal design for learning*. Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved September 6, 2005, from http://www.cast.org/teachingeverystudent/ideas/tes/

Return to "Multiple Dimensions of Assessment That Support Student Progress in Science and Mathematics."

Content and general comments: info@ncrel.org Technical information: pwaytech@contact.ncrel.org
Copyright © North Central Regional Educational Laboratory. All rights reserved.
Disclaimer and copyright information.

Gilbert Valdez: Technology in Assessment

Gilbert Valdez, Ph.D.

Director
North Central Regional Technology in Education Consortium
and Codirector of the North Central Eisenhower Mathematics and Science Consortium (NCEMSC)

Video Transcript:
Well, the value of technology is to add productivity to learning. If you think about what the tractor did for farming, this is what technology can do for learning. And it's become a necessary productivity tool for the future. If you look at 21st century skills, they're going to be requiring the use of technology very effectively. And some of the things that we most should be testing are those abilities to find, evaluate, and package in a different way more meaningful according to the audience that knowledge and not necessarily looking for what did somebody say before was the—because then there'd be no right answer. Or more often, nobody has ever answered that particular question because the knowledge is new. So I think what technology does is serve as a very powerful tool, which means that in assessment, if you're going to assess for it, you better have that tool as a part of the assessment instrument or assessment strategy. It's kind of ridiculous to think that we could assess very powerful uses of technology with paper-pencil or multiple-choice questions when in fact you really have to talk about: Can the student do the task? That's what you're trying to document.

info@ncrel.org
Copyright © North Central Regional Educational Laboratory. All Rights Reserved.
Disclaimer and copyright information.