Note: Please also use this writer as reference for the DQ as well.

Williamson, S. C. (2009). Assessment and evaluation 10/11. Boston: McGraw Hill.

Article 4 Are Standards Preventing Good Teaching?

Clair T. Berube

The National Standards movement seeks to raise the quality of the American educational system. According to one of its chief architects, Diane Ravitch (2000), national standards give clear expectations for students, teachers, parents, colleges, and employers that will result in improved student achievement. Forty-nine of the fifty states (save Iowa) have implemented such standards. In many cases, they have raised scores. In Virginia, for example, scores are going up on standardized tests (Virginia Department of Education 2001). But what, ultimately, does this mean? Is it possible for scores to go up without teacher creativity and student comprehension suffering?

For instance, I love the game show “Jeopardy.” The game is on a higher level than other game shows, and the contestants are all intelligent, well-read individuals. Although these contestants could perform at much higher levels, all they are asked to do on the show is to recall facts, the lowest level of thinking (Bloom 1956). Problem-solving activities, checks for comprehension, and evidence of analysis, synthesis, or evaluation are absent. But we view these contestants as the smartest of the smart just because they can recall facts at blazing speed. However, very intelligent people possessing this talent also can comprehend, analyze, synthesize, and evaluate.

In the same way, has the American educational system chosen an incomplete set of tools to measure the progress of the standards movement? Unlike the SAT, where a combination of multiple-choice and essay questions offer an understanding of a college-bound student's abilities, some states only use one type of test to measure student comprehension. In Virginia, it is the Virginia Standards of Learning (SOL). This particular test is used because (1) it is cheap, (2) it is easy to read, and (3) it is simple to grade. The test is also extremely objective, leaving no room for graders' opinions and differences of viewpoints. Although multiple-choice standardized tests claim to measure every level of learning, they really only test knowledge recall. And as educators, we use these multiple-choice “bubble” tests to convince ourselves that our students truly “understand” what we teach them, as evidenced by a passing test score.

But picture this scenario in any middle-school science classroom in America: A science teacher proudly explains to her principal how she has successfully taught Einstein's theory of relativity to her students (a very lofty concept indeed for middle schoolers). Her assessment of student discussions, projects, and papers has proven that they have “gotten” it. However, the principal looks over her bifocals in a disapproving manner and scolds, “Well that's fine, but can they pass the test?” Not understanding, the teacher reiterates that they have passed her assessments. But the principal is speaking of “The Test,” the high-stakes, multiple-choice standardized test that is given at the end of each year. This really happened to me; I was that sixth grade science teacher.

The problem is not the standards but, rather, how the standards are assessed. I would have been unable to teach without the Virginia Standards of Learning. They gave me a wonderful roadmap that I coordinated with my curriculum guide and the district's guidelines. But something went terribly wrong when my former state (Virginia) began measuring the standards with multiple-choice, high-stakes tests. These tests hold teachers' and administrators' creativity hostage and threaten job security and professional contentment. In addition, the tests hardly prove that students have learned anything other than rote memorization of facts. And although there are multiple-choice tests that try to focus on higher-order thinking skills (the SAT, for example), they also employ other forms of assessment, such as qualitative essays.

Toward the end of every school year, Virginia eighth graders are required to take the Standards of Learning (SOL) test, which is supposed to assess what they learned in eighth grade physical science class but also covers sixth and seventh grade science material as well. I wondered if the grades on this test were good indicators of the students' comprehension of the subject matter. So, I conducted a large study with middle-school science classes in Norfolk, Virginia. I hypothesized that teacher style affected SOL scores, such that the more constructivism-oriented teachers' classes would receive higher scores over those taught by lecture-oriented teachers, especially on the comprehension test.

16 17

Constructivism has been implemented in classrooms all across America and is recognized as a student-centered, discovery learning process where the teacher—while still teaching strong content and skill development—assists the students in problem-based learning. Given that children build knowledge from their own experience and beliefs, this epistemology emphasizes the construction of concepts rather than transmission and recording of information given by others (Applefield, Huber, and Moallem 2000/2001; Gatlin 1998).

I constructed my own version of the eighth grade SOL test that I called the “Comprehension Measurement.” It was the same SOL test they had taken one week earlier with one adjustment: After answering the multiple-choice question, I asked them to explain or defend their answers in short-answer/essay format. I created this instrument to find out if students taught by the more constructivist teachers also had higher comprehension scores, meaning that the students could adequately explain and defend their answers and, therefore, were not just memorizing answers supplied by their teachers.

I first had to find out what kind of teachers I had in my sample. I sent out a self-scoring Lickert scale taken from CLES: The Revised Constructivist Learning Environment Survey (Taylor, Fraser, and White 1994). The CLES measures teacher perception of constructivist attributes in the learning environment, namely their own classroom, and could determine if the teachers were constructivist, traditional, or a mix between the two styles. After returning and scoring the questionnaires, I visited each classroom to ensure that the teachers were indeed what they said they were. To do this, I used the subscales of the CLES, which detail behaviors to observe in constructivist classrooms. I coded the behaviors and developed cut-off points with which to categorize the teachers into either constructivist, traditional, or mixed. Then the students took the SOL test, and one week later, the same students took my comprehension test. I analyzed the SOL scores and compared them to my comprehension assessment scores. What I found was astonishing. Some students passed, some failed, which was to be expected, but I found that 71 percent of the students who passed the state mandated, multiple-choice test failed my comprehension test. They either could not explain their answers or gave bogus explanations. It seemed they could pass the SOL but did not understand the subject matter.

In teacher A's class, seventeen out of eighteen students passed the SOL test. Only three of those passing students failed the comprehension measurement test. Not bad. But still three students who passed the SOL test could not explain their answers. Teacher A was one of my mixed style teachers, teaching with some lecture and some discovery and student-centered instruction. But teacher A was the exception. Teacher B had eighteen students who passed the SOL. Fifteen of those failed my comprehension test. Only four students in her class passed both. Teacher C had six students out of nine pass the SOL. Five of those that passed failed the second test. And teacher D had twelve students out of eighteen pass the SOL. Every one of those students who passed the SOL failed the comprehension exam. Teacher E also had every one of her eight SOL-passing students fail my exam. Finally, teacher F had nineteen students pass the SOL, but fourteen of them failed the comprehension test (see Table 1).

Table 1 Comparison of Students Who Passed the SOL but Failed the Comprehensive Exam

Teacher Students Passing SOL Same Students Failing Comprehensive Exam Percentage

A (mixed)

17

3

17

B (constructivist)

18

15

83

C (mixed)

6

5

83

D (constructivist)

12

12

100

E (constructivist)

8

8

100

F (mixed)

19

14

73

Note. Mixed = teachers exhibiting mixture of constructivist and traditional teaching traits. Constructivist = teachers exhibiting constructivist teaching traits (as defined by Taylor and Fraser's CLES: The Revised Constructivist Learning Environment Survey).

One of the constructs that I measured during this study was teacher style. I hypothesized that the constructivist teachers would generate more student understanding because they employ more hands-on experiences, group work, and discovery learning. Some evidence supports this idea that constructivist teaching strategies that employ conceptual learning, those which do not isolate basic skills but incorporate them into skills required for completion of problemsolving tasks, does increase student comprehension (Applefield, Huber, and Moallem 2001). However, results from a multivariate analysis of variance (MANOVA) showed that scores were not higher for constructivist teachers; the more mixed and traditional, “drill and grill” teachers produced the students with the higher SOL scores. But many of those students still failed my comprehension test in high numbers. Surprisingly, the constructivist teachers, whom I 1718thought would produce the highest scores all-around, produced high pass rates on the SOL but the lowest pass rates on the comprehension test. Students could answer the simpler multiple-choice questions but could neither defend nor explain their answers, which requires higher-level skills than memorization. The constructivist teaching style (high-level teaching, according to Bloom's taxonomy) that some teachers employed during most of the year was essentially useless in preparing students for the high-stakes test.

The National Board on Educational Testing and Public Policy at Boston College conducted an extensive study to determine the effects of high-stakes testing on teacher practices. The study found that the influence of the test is greater as the stakes increase, with 40 percent of teachers in high-stakes states, such as Virginia, reporting that the tests influence their teaching on a daily basis (Lewis 2003). Teachers in high-stakes testing situations felt more pressure to have their students perform well, and therefore, more closely aligned their teaching to the test. These findings suggest that tests often affect instruction in ways that directly contradict the state educational reform policies' intent to raise standards (Schroeder 2003).

The most recent push for high-stakes testing grew out of the standards-based reform movement in the 1990s. This type of testing gave rise to accountability systems that are characterized by four components (Abrams and Madaus 2003):

• Content standards

• Tests designed to measure achievement of the content standards

• Performance targets

• Incentives, such as awarding diplomas upon passing the test

We rely on such systems for several reasons. First, public perceptions view test scores as conclusive proof that achievement is attained. Tests are seen as symbols of order, control, and attainment. However, if important decisions—jobs and governmental funding—rely on the outcome of high-stakes tests, then teachers will only teach to the test, allowing test content to define the curriculum (Abrams and Madaus 2003). School systems have gone as far as to “bribe” students to perform well. In Florida, Governor Jeb Bush implemented an educational accountability system called the “A + plan,” where financial rewards of $100 per student were offered to schools that receive an A on the basis of FCAT (Florida Comprehensive Assessment Test) scores (Myers and Curtiss 2003). In this case, it was not a reward but more like a bribe since the students were told before the test what they would receive if they passed.

Then there is the problem of high-stakes testing penalizing economically disadvantaged children. Gary Orfield, the co-director of the Civil Rights Project at Harvard University, found that high-stakes testing penalized low-income and ethnic minority students and is linked to high dropout rates in these groups. He warns that high-stakes tests are not standards but “punishment of innocent victims of unequal education” (Myers and Curtiss 2003).

Stuart S. Yeh (2001) argues that we should construct state-mandated tests that emphasize critical thinking. Conceptualizing critical thinking, beginning by looking at the term's meaning in the workplace, would be the first step. According to Yeh, critical thinking is frequently conceptualized as argumentation, a skill that the students failing the comprehension test could not display. By kindergarten, children know how to argue their side in a debate, and critical thinking could be introduced in that context very early. Yeh also argues that the current trove of critical-thinking tests consist of multiple-choice items that call for responses to artificial questions that have no bearing on real life. This, in his view, forces the tests to lack content validity. His solution is to construct standardized tests that emphasize argumentation, where students would have to use facts to defend their opinions instead of simply recalling them. Significant issues would be implemented that would tie real life problems into content.

Again, the problem is not standards or the standards movement. We, as a nation, should hold our teachers and students to high standards. But do high-stakes tests cancel out any form of higher-level teaching and learning? Do teachers who previously taught at high levels resort to “teaching to the test” during assessment crunch time at the end of the year?

As educators, we have to look closely at what material we assess and how we measure achievement. Low costs and ease in grading are hardly valid reasons to use high-stakes tests as indications of student achievement. Achievement should not be measured by how well we train our students to take multiple-choice tests. If we are not careful, we could become a nation of people who score high on standardized tests but who cannot understand, analyze, synthesize, and evaluate what we have truly learned.

References

Abrams, L. M., and G. F. Madaus. 2003. The lessons of high-stakes testing. Educational Leadership 61, no. 3 (November): 31–35.

Applefield, J. M., R. Huber, and M. Moallem. 2000/2001. Constructivism in theory and practice: Toward a better understanding. The High School Journal (December/January): 35–53.

Bloom, B., ed. 1956. Taxonomy of educational objectives: The classification of educational goals handbook 1: Cognitive domain. New York: David McKay.

Gatlin, L. S. 1998. The effect of pedagogy informed by constructivism: A comparison of student achievement across constructivist and traditional classroom environments. PhD diss. Univ. of New Orleans.

Lewis. A. C. 2003. Beyond testing. Education Digest 69, no. 1 (September): 70–71.

Myers, M. A., and D. Curtiss. 2003. Failing the equity test. Principal Leadership. 3, no. 2 (October): 70–73.

Ravitch, D. 2000. Personal communication with author. December 4: New York.

Schroeder, K. 2003. High-stakes horrors. Education Digest 68, no. 9 (May): 54–55.

18 19

Taylor, P. C., B. H. Fraser, and L. R. White. 1994. CLES: An instrument for monitoring the development of constructivist learning environment. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Virginia Department of Education. 2001. Statewide spring passing rates. http://www.pen.k12.va.us/VDOE/Assessment/StatePassRates01.html (accessed April 23, 2004).

Yeh, S. S. 2001. Tests worth teaching to: Constructing state mandated tests that emphasize critical thinking. Educational Researcher 30, no. 9 (December): 12–17.

Clair T. Berube is an assistant professor of education at Wagner College on Staten Island, New York.

From The Clearing House, July/August 2004. Reprinted by permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 Eighteenth St., NW, Washington, DC 20036-1802. Copyright © 2004. www.heldref.org

Article 5 Schools, Poverty, and the Achievement Gap

Ben Levin

Last June, representatives from more than 20 countries and several international agencies came together under the midnight sun in Trondheim, Norway, to discuss the challenge of creating greater equity in the outcomes of education. This meeting, sponsored by the Organisation for Economic Co-operation and Development (OECD) and the Norwegian Ministry of Education, was the culmination of several years of work on the theme of “equity in education.” The OECD will shortly issue a report titled No More Failures, replete with analysis and recommendations on how to improve equity in educational outcomes.

The concern for “raising the bar and closing the gap” in educational outcomes is now widespread around the world. Kappan readers will be familiar with the debate on the achievement gap, especially in the context of No Child Left Behind. In Europe, the results of PISA (Programme for International Student Assessment—www.pisa.oecd.org) brought the issue into stark relief as well. PISA, a large, carefully designed study now involving more than 40 countries, tests 15-year-olds in reading, science, and mathematics. There have been two rounds of results so far, in 2000 and 2003, with a third due to be released this December.

The findings of PISA have been striking and consistent. Some countries that thought they were doing well educationally found that they had not only poor overall results but also very large gaps between their highest- and lowest-achieving students. In Germany, the phrase “PISA Schock” has come into the language as a sign of how serious the problem is. In contrast, some other countries, such as Finland, Korea, and—yes—Canada, showed very high overall results and much smaller gaps in their achievement distribution.

The reality, in PISA and in every other assessment of student outcomes, is that socioeconomic status remains the most powerful single influence on students' educational and other life outcomes. This is true in Finland and Canada as well as in the U.S. and everywhere else. Where you are born and grow up matters enormously to what you are able to be and do. A recent study in my home town of Winnipeg, using a database of all children born in the city in 1984, showed that, whereas 89% of all students writing the grade-12 language exam passed, only 12% of students whose families had received social assistance in the previous two years passed the exam.1 Indeed, a large proportion of this group was either a year or more behind or out of school entirely.

Although the achievement gap in Canada is smaller than in the United States, it is far from trivial. UNICEF's Innocenti Research Centre recently released a report with the fascinating title of Child Poverty in Perspective: An Overview of Child Well-Being in Rich Countries (www.unicef-irc.org/publications). Using a rich array of data, it compares the situation of children in Canada, the U.S., the United Kingdom, and 18 other European countries on six dimensions, including material well-being, health and safety, education, peer and family relationships, behaviors and risks, and young people's subjective sense of well-being.

No country ranks high on all six dimensions. The Netherlands gets the best overall score. Canada's average ranking across the six areas is 12th, while the U.S. and the U.K. are at the bottom. And the kicker is that the report concludes: “Variation in government policy appears to account for most of the variation in child poverty levels between OECD countries. No OECD country devoting 10% or more of GDP to social transfers has a child poverty rate higher than 10%. No country devoting less than 5% of GDP to social transfers has a child poverty rate of less than 15%.”2

About 15% of Canadian children live in poverty, defined as living in a household with income less than 50% of the national median. What makes the Canadian situation galling is that in 1989 the Parliament of Canada passed a unanimous motion to end child poverty in the country by the year 2000. Surprise! When 2000 rolled around, the child poverty rate was higher than in 1989. In Canadian schools students in special education, recent immigrants, some visible minority groups, and Aboriginal youths lag behind national averages of educational achievement.

All of this makes one skeptical of new pronouncements by governments on their commitment to greater equity. Yet equity in educational outcomes is high on the international policy 2021agenda for powerful reasons. First, it is widely claimed that better educational outcomes are essential for national economic and social success, though as Gerald Bracey has pointed out in these pages, the claim is not necessarily well supported by the empirical evidence. In public policy and politics, though, evidence matters only if it affects beliefs, and this does not happen so quickly.

We should not lose faith in evidence entirely, though, because another factor driving the current interest in narrowing the achievement gap is research showing that reduced inequities in income and education are connected to better economic performance. Countries with less inequality in income and education actually show better economic performance, calling into question the long-standing belief that countries face an inevitable choice between equity and efficiency.3

Governments and international agencies are all considering what steps they can take. Some take the view that schools could do much more to reduce, if not eliminate, the effects of poverty. This has led to some very dubious strategies, such as various kinds of takeovers or reconstitution of so-called failing schools, as if the problem were simply one of working a little harder.

Others argue that socioeconomic status is too powerful and that schools alone will not be able to mitigate its impact. Richard Rothstein, another Kappan contributor, has made this argument particularly well, including pointing to alternative policy measures that might have more impact on outcomes than some school programs.4

For educators working in high-poverty communities, finding an appropriate stance toward poverty and the achievement gap can be difficult. Educators see the daily challenges in the lives of their students, including poor housing, inadequate income, and the effects of discrimination. Schools did not create these problems, and on their own they cannot solve them.

But folks in schools are not always clear on what they can or should do about the problems of poverty and inequality, and the steps we take are not always the best ones.5 We know that poor children often get teachers who are less qualified and instruction that is less challenging, when they need the best we can give them. The OECD report No More Failures lists such sensible steps as reducing retention in grade, stressing early intervention to address reading problems, reducing early tracking in secondary schools, directing additional resources to the highest-need schools, reaching out more to parents, and managing school choice to make sure it does not exacerbate inequities.

There are some good practices in Canada that address equity issues. First Nations are slowly increasing their high school graduation rates. British Columbia is working hard to improve outcomes for Aboriginal students in provincial schools, while Alberta has its Initiative for School Improvement, and Quebec has a strategy for student success. Ontario has developed ambitious strategies that are yielding improved literacy and numeracy skills in elementary schools and higher high school graduation rates. Most of these efforts rightly stress working with teachers to improve their ability to support success in diverse student populations.

A Personal Introduction

I am delighted to be able to write the In Canada column for the Kappan, a magazine I have read and admired for many years. Heather-Jane Robertson's knowledge, intelligence, and writing skills make her a tough act to follow! I thought it would be useful to introduce myself briefly to readers.

Though I now live in Toronto, I was born and have spent most of my life in Winnipeg. My career in education has moved back and forth between government and academe. I have been a civil servant in a department/ministry of education four times, including serving as deputy minister (chief civil servant) in Manitoba from 1999 to 2002 and in Ontario from 2004 to early 2007. In academe, I have been a researcher and professor and currently occupy a Canada Research Chair in Education Leadership and Policy at the Ontario Institute for Studies in Education, which is part of the University of Toronto. I have been a member of PDK since 1984. My three daughters are all graduates of Manitoba public schools.

I hope to write about a broad array of issues from a Canadian perspective while connecting them to the experience of Kappan readers in the U.S. and elsewhere. I welcome comments from readers or ideas for content for future columns. I can be reached at blevin@oise.utoronto.ca.

References

1. Noralou P. Roos et al., “The Complete Story: A Population-Based Perspective on School Performance and Educational Testing,” Canadian Journal of Education, vol. 29, 2006, pp. 684–705.

2. UNICEF, Child Poverty in Perspective: An Overview of Child Well-Being in Rich Countries (Florence: Innocenti Research Centre, Report Card 7, 2007), p. 7.

3. World Development Report 2006: Equity and Development (Washington, D.C.: World Bank, 2005).

4. Richard Rothstein, Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap (New York: Teachers College Press, 2004).

5. Benjamin Levin and J. Anthony Riffel, “Current and Potential School System Responses to Poverty,” Canadian Public Policy, vol. 26, 2000, pp. 183–96.

Ben Levin is Canada Research Chair in Education Leadership and Policy at the Ontario Institute for the Study of Education, University of Toronto (blevin@oise.utoronto.ca).

From Phi Delta Kappan, September 2007, pp. 75–76. Copyright © 2007 by Phi Delta Kappan. Reprinted by permission of Phi Delta Kappan and Ben Levin.

Article 6 Making Benchmark Testing Work

Joan L. Herman

Eva L. Baker

Six criteria can help educators use benchmark tests to judge student skills and to target areas for improvement.

Many schools are developing assessment systems to monitor their students' progress toward state standards throughout the academic year. Educators in these schools wisely recognize that information from annual state tests is often too little, too late. State tests can be powerful motivators, communicating expectations and focusing curriculum and instruction. But they rarely provide the ongoing information that schools need to guide instructional programs and address the learning problems of students who might otherwise be left behind.

Vendors and service providers have jumped in to fill this gap with a variety of products and services, known by such names as benchmark tests, progress monitoring systems, and formative assessments. These vendor-developed products and locally developed testing systems are designed to coordinate with state standards and assessments and are administered regularly—often quarterly—to gauge student progress. Available options include customized testing programs, off-the-shelf products aligned with existing state tests, and CDs and Web portals containing item banks from which educators can construct their own tests. Services include rapid, automated scoring and elaborate reporting systems for multiple audiences and purposes. Not uncommon, for example, are separate reports for administrators, teachers, parents, and students, providing information on class, group, and individual performance on specific grade-level standards and for overall grade-level proficiency.

Despite the glitz and gee-whiz appeal of such products, information about their effectiveness in improving student learning is generally hard to come by. Yet the quality of the assessment is essential: There is little sense in spending the time and money for elaborate testing systems if the tests do not yield accurate, useful information. If the information is flawed or erroneous, it is unlikely to provide good guidance for instruction or to support better decision making. The whole rationale for conducting the assessment falls apart; it merely creates the illusion that something is being done and people are paying attention.

The validity, or quality, of an assessment is derived from an array of evidence showing the extent to which that assessment provides sound information for particular purposes. The purpose of benchmark testing is to provide both accurate information about how well students are progressing toward mastery of standards and useful diagnostic feedback to guide instruction and improve learning. Here we discuss six criteria that determine the validity of benchmark tests: alignment, diagnostic value, fairness, technical quality, utility, and feasibility (Linn, Baker, & Dunbar, 1991). “Recommendations for Benchmark Tests” summarizes the implications of these criteria for educators.

Alignment

Alignment is the linchpin of standards-based reform. Unless benchmark tests reflect state standards and assessments, their results tell us little about whether students are making adequate progress toward achieving the standards and performing well on the assessment. The term alignment, however, has many potential meanings.

Aligning benchmark tests with state standards does not mean creating formative tests that mimic the content and format of annual state tests as specifically as possible. Although a strategy of strict test preparation may boost state test scores in the short term, available evidence suggests that early gains achieved in this way are not sustained in the long run (Herman, 2005; Hoff, 2000; Linn, 2000).

Aligning benchmark tests with state standards does not mean mimicking the content and format of annual state tests.

Further, aligning benchmark tests too closely with a state's tests gives short shrift to the state's standards. Annual tests of an hour or two's duration cannot address all the curriculum standards that a state has deemed essential knowledge for students. Because educators and students tend to focus on what will be tested, benchmark testing that covers only what appears on the state tests may accelerate curriculum narrowing. In contrast, 2526good benchmark testing can encourage instruction on the full depth and breadth of the standards and give students opportunities to apply their knowledge and skills in a variety of contexts and formats.

Recommendations for Benchmark Tests

1. Align standards and benchmark assessments from the beginning of test development. Decide what specific content to assess and at what level of intellectual demand. Include the application of complex learning. To create benchmark tests that enrich student learning opportunities, focus on the big ideas of a content area and counteract curriculum narrowing by designing benchmark tests that allow students to apply their knowledge and skills in a variety of contexts and formats.

2. Enhance the diagnostic value of assessment results through initial item and test structure design. Use extended-response items to reveal student thinking and potential misconceptions. Build distracters into multiple-choice items that reveal common student misunderstandings.

3. Ensure the fairness of benchmark assessments for all students, including English language learners and students with disabilities. Avoid unnecessarily complex language or specific contexts that could unfairly confound some students' ability to show what they know.

4. Insist on data showing tests' technical quality. Study psychometric indices to determine the reliability of assessments.

5. Build in utility. Design reports of test results to be user-friendly and to provide guidance on how to appropriately interpret and use the results.

6. Hold benchmark testing accountable for meeting its purposes. Crafting good benchmark tests and ensuring their wise use for improving student learning requires systematic design and continual evaluation.

Good benchmark testing can encourage instruction on the full depth and breadth of the standards.

For example, knowledge of Newton's laws is included in most states' physics standards. The typical state test may address this knowledge with one or two multiple-choice items or perhaps a short-answer item. In contrast, well-developed benchmark tests use not only multiple-choice and open-ended items but also performance tasks and laboratory experiments to delve deeper into students' understanding of Newton's laws. A test might ask students to explain the underlying principles of force and motion at work in a car crash, for instance, or to design a roller coaster that makes optimal use of physics principles.

Mapping Content

The alternative to aligning benchmark tests with the specific content and format of state assessments is to align them with priority content and performance expectations implicit in state standards. Alignment researchers and learning theorists have suggested that in establishing such priorities, educators must define both the major knowledge and skills to be addressed and the expected intellectual level of the performance (Porter, 2002).

The matrix in Figure 1 shows how a school might map expectations for a state's grade 6 mathematics standards. This matrix lays out the substance of standards in terms of the specific content knowledge that students need to acquire and includes four cognitive levels suggested by Norman Webb (1997): recall (knowledge of specific facts, definitions, simple procedures, formulas, and so on); conceptual understanding (knowledge of principles and the ability to apply them in relatively routine situations); problem solving (the ability to reason, plan, use evidence, and apply abstract thinking in novel situations); and extended and strategic thinking (the ability to apply significant conceptual understanding to extended, novel tasks requiring complex reasoning; to make connections among different content areas; and to devise creative solutions).

Figure 1 A sample matrix to map expectations.

Figure 1 provides a starting place for identifying what content the school should teach and assess. But even when educators develop such a matrix, they have not yet resolved an important tension: Like the annual state tests, benchmark tests cannot possibly address all of a state's standards. Imagine a test that included items for every cell in Figure 1, with every standard implying a myriad of important objectives and topics that could be assessed at every cognitive level. Testing time would be endless. Instead, educators need to decide in advance what content is most important to assess, and at what levels.

Focusing on Big Ideas

By incorporating the key principles that underlie state or district standards into benchmark assessments, educators have a reasonable strategy for addressing the breadth of these standards. Cognitive research across many different subject areas suggests the power of focusing on the key principles underlying a content domain rather than on the specific topics within the domain (Ball & Bass, 2001; Carpenter & Franke, 2001; diSessa & Minstrell, 1998; Ericsson, 2002). Encompassing specific topics, the principles help students to organize and use their knowledge. For example, understanding the principle of equivalence can help students balance mathematical equations.

Research also demonstrates the power of engaging students in applying and explaining key principles (see, for example, Chi, 2000; VanLehn, 1996). Incorporating this idea into assessments can help increase their learning value. Therefore, despite the ease of scoring multiple-choice items, benchmark tests should 2627employ many different formats to enable students to reveal the depth of their understanding.

For example, the following two test items call for students to give short answers, choose multiple-choice options, offer extended explanations, and draw pictures to demonstrate their understanding of fractions:

Six people are going to share five chocolate bars. Write the fraction that shows how much chocolate each person gets: \_\_\_\_\_\_\_\_

Then, explain what you did to find this answer. You can draw a picture of the chocolate bars to help explain your answer.

Which of the following fractions is between 2 ½ and 2 ¾?

A. 2¼

B. 2⅚

C. 2⅔

D. 2⅓

Explain how you found the answer to this problem. Draw a picture that shows your answer is correct.

Diagnostic Value

A test has diagnostic value to the extent that it provides useful feedback for instructional planning for individuals and groups. A test with high diagnostic value will tell us not only whether students are performing well but also why students are performing at certain levels and what to do about it.

Open-ended test items that ask students to explain their answers increase the diagnostic value of benchmark tests. Students' responses reveal their thinking, helping teachers to refine their instructional strategies and design targeted instruction for individual students.

27 28

Multiple-choice items can also yield important diagnostic information, particularly when they are purposely designed so that distracters—the incorrect answer options—reflect common student misunderstandings. Consider the science test item shown in Figure 2, which is designed to assess students' understanding of electricity principles. Number 4 is the correct choice; it has the highest current because the voltage is the largest and the resistance is the smallest of all the circuits. Students who incorrectly choose number 2 show the weakest understanding of current because the voltage is the smallest and the resistance is the largest of all four circuits. We may infer that students who circle number 1 realize that smaller resistance is often associated with larger current but do not understand the role of voltage; conversely, students who circle number 3 may understand that larger voltage is often associated with larger current but do not understand the importance of resistance. Our inferences would be stronger, of course, if the assessment also asked students to explain their reasoning for their choices.

Figure 2 Sample Test Item, Electricity Principles. This test item is intentionally designed so that each incorrect answer suggests a different misconception or common student error.

To provide good diagnostic information on where and how students are experiencing difficulties, benchmark tests must include enough items on each potential topic to render a reliable diagnosis. Drawing inferences from performance on only one or two items or from unreliable subscales may result in faulty conclusions.

Fairness

Fair benchmark tests provide an accurate assessment of diverse subgroups. Biased test items, in contrast, introduce unnecessary complexities that systematically hamper access for particular subgroups. For example, when test items use complex language to assess students' science knowledge, English language learners or poor readers may suffer an unfair disadvantage because they are unable to demonstrate their actual skill in science. Similarly, setting problems in contexts that are less familiar to some subgroups can impede those groups' ability to apply their knowledge and skill. A mathematics problem that asks students to compute the best route for a subway trip may be clear to students from New York City but may confuse students who have never been on a subway, even if they know what it is.

Fairness also is a prime issue in testing students with disabilities. For example, students with specific reading disabilities may need more time to process text. Although details on accommodations are beyond our scope here, two points are worth underscoring: (1) Accommodations offered to students in benchmark testing should mirror those documented in their individualized education plans and offered to them on annual state tests, and (2) the design of benchmark tests should make these tests accessible for as many students as possible. As with any standardized assessment, benchmark test items should be thoroughly reviewed for possible bias by representatives of diverse communities, as well as tested empirically to identify any items that have aberrant results for particular subgroups.

Technical Quality

Tests with high technical quality provide accurate and reliable information about student performance. Those of us who are not psychometricians tend to zone out when talk turns to technical indices of test quality. But item and test quality provide important information about whether we can trust the scores. If a test is weak on this characteristic, plans and decisions made on the basis of test data are likely to be faulty.

For example, reliability—determined by internal consistency, item response theory, inter-rater agreement, and a number of other indices—stands for the consistency of a measure and the extent to which scores on that measure represent some stable and coherent core. When a measure is highly reliable, the items within it operate similarly. Reliability problems arise if a student's performance varies significantly across items, within 2829a short period of time, or under a whole host of other conditions (during the stress of an exam, when the student is tired, when the testing room is uncomfortable, and so on).

Reliability problems arise if a student's performance varies significantly across test items.

Imagine, for instance, a test of archery skill in which hitting the bull's-eye represents high levels of performance. One individual shoots five arrows that all hit the bull's-eye or very close to it. A second individual also shoots five arrows, but they all land in the outer ring of the target. Although the first archer's performance is more accurate, both performances are reliable. Looking at the results, we can be confident that we have an accurate measure of each individual's archery skills—at least on that particular day, under those conditions. A third archer's performance, however, is inconsistent, or unreliable: One arrow hits the bull's-eye; one lands near it; another hits the outermost ring; and two miss the target altogether. Assessing that performance, we would find it difficult to judge the archer's skill level, predict performance on the next shot, or devise corrective action—all inferences we want to draw from benchmark testing.

Inter-rater reliability enters the picture for open-ended and performance items—for example, a district writing assessment—which must be scored by human judges. Low inter-rater reliability often means that raters have not been trained well enough to agree on the meaning of high-quality performance. Districts should carefully monitor inter-rater reliability and take action to improve it if needed.

For benchmark tests to have diagnostic value, we must ensure the reliability of the diagnosis. We can easily create subscales that look useful—for example, aggregating the results from four items that appear to measure students' understanding of rational numbers. But if those items do not result in a reliable scale—if a student's performance on them varies widely—then the results do not provide good information for our instructional decisions.

Reliability and accuracy are necessary but not sufficient prerequisites to validity (that is, the extent to which a test accomplishes its intended purposes). A prime purpose of benchmark testing is to show whether students are progressing toward achieving proficiency on state tests; therefore, if the benchmark tests are doing their job, there should be a strong predictive relationship between students' performance on the benchmark tests and students' performance on the state assessments.

Educators should plan to document the reliability and validity of their benchmark tests on an ongoing basis. Good tests aren't magically created by simply assembling test items that seem reasonable—even if the tests are aligned with priority standards and teachers and psychometricians developed them collaboratively. Schools need to pilot-test and revise their items and their test forms on the basis of the technical data, ideally before a test becomes operational. Devoting sufficient time for development will yield better information from a benchmark test in the long run.

Utility

Utility represents the extent to which intended users find the test results meaningful and are able to use them to improve teaching and learning. Benchmark tests with high utility provide information that administrators, teachers, and students can use to monitor student progress and take appropriate action. District administrators, for example, may use the data to identify schools that need immediate help in particular subjects. School principals may use the data to identify students for special after-school tutoring. And teachers may use the information to modify their teaching and to regroup students for supplementary instruction.

To make benchmark tests useful, schools must put the results in intended users' hands quickly and train them to interpret the information correctly. In addition, schools must administer assessments and provide feedback when such guidance can be most useful—that is, around the time when teachers address the test content in classroom instruction. If teachers in different classrooms or schools use different curriculum materials or take more or less time teaching the topic, then finding common testing times may be an issue. For example, if one school covers Newton's laws in the fall and another covers this topic in the spring, a fall benchmark test on the topic will be of little use to the second school. To address the problem, some school districts give teachers flexibility in determining what content to assess during each testing period.

Schools can also increase the effectiveness of benchmark tests by helping teachers use the results. Teachers who lack such support may not know what to do when assessment results show that students are struggling; they may hesitate to go back and reteach because they feel pressure to move on and “cover” the curriculum. Even if they do go back, they may replicate the same strategies that were unsuccessful in the first place.

Schools can increase the effectiveness of benchmark tests by helping teachers use the results.

In addition to giving teachers the data, schools must ensure that they have the pedagogical knowledge and access to alternative materials that they need to bridge identified learning gaps. Some districts and schools help teachers by establishing grade-level or subject-matter teams, including content and curriculum experts, to meet regularly to analyze student work, discuss strengths and weaknesses in learning, and formulate next steps for individual students and subgroups representing various learning needs.

Some of the benchmark testing products available address this need by producing test score reports that explicitly identify student strengths and weaknesses, make suggestions for teaching, and direct teachers to useful materials. However, the value of such approaches depends on the validity of score interpretations and the actual learning benefits of the suggested next steps. Educators considering the purchase of such products should look for evidence to support both of these components.

29 30

Feasibility

Benchmark testing should be worth the time and money that schools invest in it. Well-designed benchmark tests can contribute to, as well as measure, student learning. But if such tests are not well designed, they can waste students' and teachers' valuable time and energy, ultimately detracting from good teaching and meaningful learning.

Of course, to determine whether benchmark testing is worth the effort, educators ultimately need to look at the results. Are benchmark tests focusing attention on student performance and providing solid information on which to base improvement efforts? Are they actually improving student learning? The history of testing is fraught with good intentions that have gone awry. Like state assessments, benchmark tests will fulfill their promise only if we monitor their consequences and continually improve their quality.

References

Ball, D. L., & Bass, H. (2001). What mathematical knowledge is entailed in teaching children to reason mathematically? In National Research Council, Knowing and learning mathematics for teaching: Proceedings of a workshop (pp. 26–34). Washington, DC: National Academy Press.

Carpenter, T., & Franke, M. (2001). Developing algebraic reasoning in the elementary school. In H. Chick, K. Stacey, J. Vincent, & J. Vincent (Eds.), Proceedings of the 12th ICMI Study Conference (Vol. 1, pp. 155–162). Melbourne, Australia: University of Melbourne.

Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), Advances in instruction psychology (Vol. 5, pp. 161–238). Mahwah, NJ: Erlbaum.

diSessa, A., & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. Goldman (Eds.), Thinking practices in learning and teaching science and mathematics (pp. 155–187). Mahwah, NJ: Erlbaum.

Ericsson, K. A. (2002). Attaining excellence through deliberate practice: Insights from the study of expert performance. In M. Ferrari (Ed.), The pursuit of excellence in education (pp. 21–55). Hillsdale, NJ: Erlbaum.

Herman, J. (2005). Making accountability work to improve student learning. (CSE Technical Report #649). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Hoff, D. (2000, Jan. 26). Testing ups and downs predictable. Education Week, pp. 1, 12–13.

Linn, R. (2000). Assessment and accountability. Educational Research, 29, 2.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15–21. (ERIC Document Reproduction Service No. EJ 436 999)

Porter, A. (2002). Measuring the content of instruction: Uses in research and practice. Educational Researcher, 31, 3–14.

VanLehn, K. (1996). Cognitive skill acquisition. In J. Spence, J. Darly, & D. J. Foss (Eds.), Annual review of psychology (Vol. 42, pp. 513–539). Palo Alto, CA: Annual Reviews.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. Madison, WI: University of Wisconsin, National Institute for Science Education.

Joan L. Herman (herman@cse.ucla.edu) and Eva L. Baker (eva@ucla.edu) are Codirectors of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California–Los Angeles, 300 Charles Young Dr. North, Los Angeles, CA 90095.

From Educational Leadership, November 2005. Copyright © 2005 by ASCD. Reprinted by permission. The Association for Supervision and Curriculum Development is a worldwide community of educators advocating sound policies and sharing best practices to achieve the success of each learner. To learn more, visit ASCD at www.ascd.org

Article 7 Mapping the Road to Proficiency

Thomas R. Guskey

A table of specifications provides a travel guide to help teachers move students toward mastery of standards.

When the standards movement began in the United States more than 15 years ago, most educators welcomed the idea. The enthusiasm that greeted the first set of clearly articulated student learning goals, published by the National Council of Teachers of Mathematics in 1989, led other professional organizations to follow suit. During the next decade, the National Council for the Social Studies, the National Academy of Sciences, and the National Council of Teachers of English all developed standards in their respective disciplines. States also took up the task, with Kentucky leading the way in 1990. Today, 49 of the 50 states have established standards for student learning.

Thoughtfully constructed standards guide education reform initiatives by providing consensus about what students should learn and what skills they should acquire. Standards also bring much-needed focus to curriculum development efforts and provide the impetus for fashioning new forms of student assessment.

To bring about significant improvement in education, we must link standards to what takes place in classrooms.

But to bring about significant improvement in education, we must link standards to what takes place in classrooms. For that to happen, teachers need to do two important things: (1) translate the standards into specific classroom experiences that facilitate student learning and (2) ensure that classroom assessments effectively measure that learning (Guskey, 1999).

Teachers need to translate standards into experiences that facilitate student learning.

Some states, school districts, and commercial publishers have developed teaching guides that identify instructional materials and classroom activities to help teachers meet the first challenge. Rarely, however, do teachers get help in meeting the second challenge—developing classroom assessments that not only address standards accurately, but also help identify instructional weaknesses and diagnose individual student learning problems.

Translating Standards into Instruction and Assessments

Large-scale assessments provide evidence of students' proficiency with regard to the standards developed by states and professional organizations. These assessments are well suited to measure the final results of instruction and, thus, to serve the purposes of summative evaluation and accountability.

But teachers cannot be concerned only with final results. Their primary concern lies in the process of helping students reach proficiency. Large-scale assessments just don't offer teachers much help in that respect. They tend to be too broad and are administered too infrequently. In addition, teachers often don't receive their results until several weeks or months after students take the assessment.

To understand the difference between assessing the final product and supporting progress toward that product, we might consider a youngster learning to play tennis. If you were concerned only with summative evaluation and accountability, you would need to have a clear mental picture of a “proficient” tennis player—the standard that you wanted the student to attain at the end of the learning process. Your mental picture might include approaching the ball, positioning the racket correctly, swinging smoothly, returning the ball to the other side of the court, and following the rules of the game. You would then need to identify specific criteria for judging the student's performance and finally develop a rubric describing various levels of proficiency on each of these steps.

31 32

If you were a tennis coach, forming a clear mental picture of a “proficient” tennis player would be only your starting point.

If you were a tennis coach, however, that mental picture would be only your starting point. From there, you would go on to divide the aspects of your desired final performance into various components. You would probably think about matching the racket to the student's size and strength; adjusting the student's grip for backhand and forehand returns; explaining the importance of watching the ball; and demonstrating the backswing, return, and follow-through. You would introduce important terms, such as service line, backcourt, and volley. You would also need to explain the rules and describe how to keep score.

Building on this analysis, you would consider an appropriate sequence of learning steps, perhaps ordered in terms of difficulty or complexity. You would present basic elements, such as watching the ball, before such advanced elements as achieving appropriate follow-through and recovery. As you taught, you would check for any special problems the student may experience and correct them when they appeared. You would also need to become aware of individual differences among players and adapt your teaching to those differences. For instance, some players do well using a traditional closed stance; others do better with a more open stance. In addition, you would probably make a point of complimenting the student whenever progress was evident and providing reassurance during challenging times. And, of course, you would emphasize the enjoyable aspects of the game and give the student opportunities to experience these.

This example illustrates the complex process that takes place in effective standards-based teaching and learning. To organize instructional units and plan appropriate classroom activities, teachers must unpack the standards—that is, determine the various components of each standard that students must learn and then organize and arrange these components in a meaningful sequence of learning steps. Teachers must make adaptations for individual learning differences to ensure that all students understand, practice, and master each component as they progress toward the final goal. As part of this process, teachers need to develop procedures to formatively assess learning progress, identify learning problems, and determine the effectiveness of their instructional activities.

A Tool to Link Assessments to Standards

One tool to analyze standards for instruction and assessments is a table of specifications: a simple table that describes the various kinds of knowledge and abilities that students must master to meet a particular standard. Growing numbers of teachers are discovering how this strategy, described years ago in the work of Ralph Tyler (1949) and Benjamin Bloom (Bloom, Hastings, & Madaus, 1971), can help them align their classroom instruction and assessments with curriculum standards.

As a planning tool, a table of specifications serves two important functions. First, it adds precision and clarity to teaching. The information in the table helps teachers break down standards into meaningful components that exactly convey the purpose of the instruction. It also clarifies for students the learning goals of a course or unit so that students understand what they are expected to learn. In fact, many teachers use tables of specifications as teaching guides, sharing their tables with students to reinforce students' understanding and learning progress.

Second, a table of specifications serves as a guide for consistency among standards, the steps needed to help students attain them, and procedures for checking on students' learning progress. Although this alignment is essential in standards-based teaching and learning, teachers often neglect it in their planning (Guskey, 1997). For example, many teachers stress that they want their students to develop higher-level cognitive skills—such as the ability to apply knowledge to new situations—but administer quizzes and classroom assessments that tap mainly the skills that are easiest to assess, particularly knowledge of facts and definitions of terms.

Developing Tables of Specifications

To develop tables of specifications, teachers must address two essential questions regarding the standard or set of standards in question. The first question is, What must students learn to be proficient at this standard? In other words, what new concepts, content, or material are students expected to learn? Teachers often use textbooks and other learning resources as guides in addressing this question. But textbooks should not be the only guide. Teachers should feel free to add to or delete from what the textbook and other learning materials provide to better match the standards and better fit students' learning needs.

The second essential question is, What must students he able to do with what they learn? In answering this question, teachers must determine what particular skills, abilities, or capacities must pair up with the new concepts and material. For example, will students simply be required to know the steps of the scientific method of investigation, or should they be able to apply those steps in a classroom scientific experiment?

Teachers generally find it helpful to outline their answers to these two questions using some of the categories in the Taxonomy of Educational Objectives (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956), These categories represent a hierarchy of levels, moving from the simplest kinds of learning to more advanced cognitive skills. Figure 1 shows the categories that teachers in a wide variety of subject areas find most useful:

Figure 1 General format for table of specifications

• Knowledge of terms. Terms include new vocabulary, such as names, expressions, and symbols. Students may be expected to know the definitions of these terms, recognize illustrations of them, determine when they are used correctly, or recognize synonyms. Examples 3233include the terms factor and product for a mathematics standard dealing with multiplication and photosynthesis for a science standard related to plant life.

• Knowledge of facts. Facts include details that are important in their own right and those that are essential for other kinds of learning. Examples of facts are “The U.S. Senate has 100 members, two elected from each of the 50 states,” and “Wealthy families or church officials commissioned many well-known works of art and music produced during the Renaissance.”

• Knowledge of rules and principles. These generally bring together or describe the relationships among a number of facts. Typically, they concern patterns or schemas used to organize major concepts. Other terms for rules and principles include organizers, scaffolds, guidelines, and organizational cues. Examples include the commutative principle related to a mathematics standard and the rules for subject/verb agreement incorporated in a language arts standard.

• Knowledge of processes and procedures. To demonstrate their proficiency on some standards, students must know the steps involved in a certain process or procedure. Frequently, they must recall these steps in a specific sequence. For example, students may be expected to know the specific patterns of character development used in a novel, the appropriate order of steps in a mathematics problem, or the sequence of events necessary to enact legislation.

• Ability to make translations. Translation requires students to express particular ideas or concepts in a new way or to take phenomena or events in one form and represent them in another, equivalent form. It implies the ability to identify, distinguish, describe, or compute. In general, students employ translation when they put an idea in their own words or recognize new examples of general principles they have learned. Examples include having students identify the grammatical errors in sentences or convert temperatures from Fahrenheit to Celsius.

• Ability to make applications. Making applications means using terms, facts, principles, or procedures to solve problems in new or unfamiliar situations. To make applications, students first must determine what facts, rules, and procedures are relevant and essential to the problem and then use these to solve the problem. The ability to make applications involves fairly complex behavior and often represents the highest level of learning needed to be proficient on a particular standard. For example, writing a persuasive letter using appropriate elements of argument and correct grammatical forms requires the student to make applications.

• Skill in analyzing and synthesizing. Because of the complexity of analyses and syntheses, these skills typically are involved in standards for more advanced grade levels. Some teachers, however, believe that students at all levels should engage in tasks involving analysis and synthesis. Analyses typically require students to break down concepts into their constituent parts and detect the relationships among those parts by explaining, inferring, or comparing/contrasting. Examples of analyses include distinguishing facts from opinions in editorials published in the newspaper or comparing and contrasting George Washington and Ho Chi Minh, each considered the “father” of his country. Syntheses, on the other hand, involve putting together elements or concepts to develop a meaningful pattern or structure. Syntheses often call for students to develop creative solutions within the limits of a particular problem or methodological framework. They may require students to combine, construct, or integrate what they have learned. The assignment “Write a paragraph explaining how knowledge of mathematics and science helped Napoleon's armies improve the accuracy of their cannons” would require synthesis.

Once they become familiar with the format of a table of specifications, most teachers have little difficulty breaking down standards in terms of these categories. Those who use textbooks or other learning materials in developing tables usually find these 3334resources to be helpful in answering the first essential question (What must students learn to show their proficiency with regard to this standard?) but less helpful in addressing the second question (What must students be able to do with what they learn?). And because tables clarify the learning structures that underlie standards, many teachers use them both as teaching guides to help plan lessons and as study guides for students.

Advantages of Tables of Specifications

Although developing tables of specifications can be challenging at first, teachers generally find that doing so offers several advantages. First, analyzing standards in this way helps teachers link instructional activities more meaningfully to standards. If faced with several narrowly prescribed standards, for example, teachers can use the table as a framework for combining those standards and developing relationships among them in effective instructional units. On the other hand, if confronted with a very broad or general standard, developing a table can help teachers clarify the individual components that students must master to demonstrate their proficiency.

Tables of specifications also bring precision to teaching. By analyzing standards according to the categories in the table, teachers identify the different subskills that students may be required to learn and bring attention to the relationships among those subskills. Students may need to know the definition of a term, for example, to understand a fact pertaining to that term. Knowing two or three facts may be essential to understanding a particular procedure. Similarly, knowing a procedure will probably be a prerequisite to being able to apply that procedure in solving a complex problem. Clarifying these relationships makes instructional tasks more obvious and improves the diagnostic properties of classroom assessments.

Although this kind of analysis may guide teachers in choosing classroom activities, it does not dictate specific instructional practices. Teachers may address the “what” questions in developing a table of specifications in exactly the same way, and yet teach to that standard very differently. One teacher, for example, may use a discovery approach by introducing a complex problem or application to students and then helping students determine the facts, rules, or processes needed to solve the problem. Another teacher may use an advanced organizer approach by first explaining important rules or procedures to students and then posing complex problems to which students must apply those rules and procedures. In other words, precision does not prescribe method. Clarifying our goals does not dictate how we will reach them.

Tables of specifications bring added validity and utility to classroom assessments.

Finally, and perhaps most important, tables of specifications bring added validity and utility to classroom assessments. They help teachers ensure that their assessments provide honest evidence of students' learning progress, accurately identify learning problems, and provide useful information about the effectiveness of instructional activities.

Linking Classroom Assessments to Tables of Specifications

To serve formative evaluation and instructional purposes well, classroom assessments must include items or prompts for each important concept or subskill related to the standard being measured. By matching assessment items or prompts to the elements outlined in the table of specifications, teachers can ensure that their assessments measure all these important skills and abilities.

Consider, for example, the table of specifications shown in Figure 2, developed for an elementary school social studies standard related to the use and interpretation of maps. Although a large-scale assessment may include only one or two problems asking students to use or interpret maps, a classroom assessment designed for formative evaluation purposes would look very different. It would include items that assess students' knowledge of relevant terms, facts, principles, and procedures related to maps, as well as other items that measure their skill in translating that information into new forms. It would also include constructed or extended-response items that require students to apply their knowledge in using or interpreting maps. (Note that this particular elementary standard does not require analysis and synthesis skills.)

Figure 2 Table of specifications for a social studies unit on maps.

Incorporating items that draw on this wide range of cognitive skills enhances an assessment's diagnostic properties and makes it more useful as a learning tool. Suppose students are unable to answer a complex, high-level assessment item that asks them to look at a map showing various geographic features (two major rivers and their intersection, mountain ranges, flat and steeply sloped areas); to identify the location on the map where a major settlement is likely to develop; and then to explain their reasons for selecting that location.

A closer look may reveal that some students correctly answered earlier items in the formative assessment demonstrating their knowledge of the necessary facts and principles, but could not apply that knowledge in this practical, problem-solving situation. Such students clearly need additional guidance and practice in making applications. Other students may answer this high-level item incorrectly because they did not know the requisite facts and principles, as evidenced by their incorrect answers to those items appearing earlier on in the assessment. These students need to return to activities that help them gain this basic knowledge. Although such a distinction in students' learning needs matters little to those concerned only with summative evaluations of students' proficiency, it matters greatly to teachers concerned with helping students attain proficiency.

Linking classroom assessments to tables of specifications also guarantees consistency and thoroughness. In analyzing their formative classroom assessments, teachers often find 3435items they cannot locate on the table of specifications. Such items usually tap trivial aspects of learning that are unrelated to the standard, and they can be revised or eliminated from the assessment. At other times, teachers find essential learning elements included in the table that are not tapped in their classroom assessment. In such instances, teachers must expand the assessment to include measures of these vital aspects of learning. As a result, classroom assessments become more thorough, complete, and effective at serving their formative purposes.

Destination: High Achievement for All

In developing tables of specifications, teachers identify the signposts that students must reach on the way to demonstrating their proficiency on standards. Although some teachers initially find the process challenging, most soon discover that it not only improves the quality of their classroom assessments but also enhances the quality of their teaching. Analyzing standards in this way clarifies what students need to learn and be able to do. With that focus established, teachers can concentrate more fully on how best to present new concepts and engage students in valuable learning experiences.

A table of specifications is much like a travel guide. Although it never limits the pathways available, it enhances traveling efficiency, enjoyment of the journey, and the likelihood of successfully reaching the intended destination.

References

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives, handbook 1: Cognitive domain. New York: McKay.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York; McGraw-Hill.

Guskey, T. R. (1997). Implementing mastery learning (2nd ed.). Belmont, CA: Wadsworth.

Guskey, T. R. (1999). Making standards work. The School Administrator, 56(9), 44.

Tyler, R. W. (1949). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.

Thomas R. Guskey is Professor of Education Policy Studies and Evaluation, College of Education, University of Kentucky, Taylor Education Building, Lexington, KY 40506; 859-257-8666; guskey@uky.edu. He is author of the upcoming book Benjamin S. Bloom: Portraits of an Educator (Rowman & Littlefield Education, 2006).

From Educational Leadership, November 2005. Copyright © 2005 Thomas R. Guskey. Reprinted by permission of the author.

Article 8 Curriculum Mapping: Building Collaboration and Communication

Angela Koppang

This article explores the application and use of curriculum mapping as a tool to assist teachers in communicating the content, skills, and assessments used in their classrooms. The process of curriculum mapping is explained, and the adaptation of the process for special education teachers is detailed. Finally, examples are given of how curriculum mapping can assist both special and general education teachers in meeting the needs of students in the classroom. Although this article will apply the use of curriculum mapping data at the middle school level, the process of mapping is equally effective at the elementary and high school levels.

Mrs. Anderson, a seventh-grade life science teacher, has 24 students in her class, 4 of whom are students with disabilities. Jesse has a mild learning disability and needs little assistance in the class, although he has some difficulties with organization. Jenny and Brian have mild cognitive disabilities and have some modified expectations for vocabulary and lengthier written assignments. John has multiple disabilities, including moderate cognitive impairment and physical disabilities. He is responsible only for a small part of the vocabulary and content, and his work is primarily designed to parallel the classroom content. In addition to these students, Mrs. Anderson has Marina, an English language learner, and 19 other students of differing abilities. Twice each week, Mr. Jones, a special education teacher, joins Mrs. Anderson, and they share teaching responsibilities and group students for instruction in a variety of ways. On the remaining days, Mrs. Smith, a teaching assistant, is available to assist Mrs. Anderson and individual students in the classroom. How do these teachers effectively communicate about the content and skills that will be used in the classroom? They base all instructional planning—as well as decisions about curriculum adaptations and modifications—on the content, skills, and assessments found in curriculum maps developed by the teachers in the school.

What Is Curriculum Mapping?

Curriculum mapping is a method of collecting data about what is really being taught in schools. It has been advocated as a method of aligning the written and taught curriculum since the early 1970s. More recent advances in technology have expanded the use of curriculum mapping as a tool for improving communication among teachers about the content, skills, and assessments that are a part of the instructional process. This new application of curriculum mapping holds great promise for enhancing the collaboration between general and special education teachers to benefit all learners.

Curriculum mapping is a process used to gather a database of the operational curriculum of a school (Hayes-Jacobs, 1997). Although most schools have well-developed curriculum guides, information is often limited about how the standards set forth in those guides directly relate to what is actually happening in the classroom. Most curriculum guides identify what students should know and be able to do but give little insight into how students accomplish this learning or what assessment methods are used by teachers. In combination with traditional curriculum guides, curriculum maps can provide information about content and skills used for instruction, as well as the length of time devoted to various aspects of the curriculum. Including assessment methods on the maps provides a link to the expectations for the manner in which students will be expected to demonstrate their knowledge. The details included in the curriculum maps give a clearer picture of what actually occurs in each classroom.

In the curriculum mapping process, teachers use a calendar-based system (see Table 1) to map the skills, content, and assessments used in their classrooms (Hayes-Jacobs, 1997). Because each teacher takes an individual approach to meeting the curriculum standards, the individual 3637maps will reflect the differences in approaches for achieving curricular goals. The completed maps may be used for many purposes, including

Table 1 Life Science Curriculum Map

Life Science Content Skills Assessment

January

• Characteristics of plants

• Seedless plants

• Seed plants

• Complex plants

• Plant reproduction

• Rain forest

• List characteristics of plants

• Compare vascular & nonvascular plants

• Describe & illustrate structures of roots, leaves, & stems

• List characteristics of seed plants & find seeds in plant lab

• Describe & label the functions of the flower in flower dissection lab

• Describe methods of seed dispersal

• Understand the environmental impact of the rain forest

• Plant worksheet

• Vascular plant art

• Plant drawings

• Oral presentation of group work in plant lab

• Flower lab report & labels

• Rain forest essay

• aligning instruction to the written standards;

• developing integrated curriculum units;

• providing a baseline for the curriculum review and renewal process;

• identifying staff development needs; and

• most important, providing communication among teachers.

One of the most powerful outcomes of the curriculum mapping process is using the maps as a communication tool among teachers within a school.

Hayes-Jacobs (1997) said, “Curriculum mapping amplifies the possibilities for long-range planning, short-term preparation, and clear communication” (p. 5). This focus on planning, preparation, and communication facilitates a higher level of collaboration between general education teachers and special education staff. This process can involve general and special educators on many different levels to enhance effective collaboration within a school.

Curriculum Mapping Process

While mapping is most effective when the entire school staff is involved, many school staff members have started this process by mapping one or two grade levels at an elementary school or one interdisciplinary team or department in middle or secondary schools. The process is easily accomplished by both novice and veteran teachers. The key to the success of the process is staff discussions and how data are used after the maps are completed.

Each teacher begins the process of mapping by recording his or her content, skills, and assessments. Using a computer program enhances the process of mapping by allowing for revision of the maps, as well as the ability to share the maps throughout a school by posting them on a server or school Web site. Several excellent software programs are specifically designed for curriculum mapping; however, it is not necessary to purchase software to complete the mapping process. Many schools have started the process with a simple computer template created in a word-processing program resembling the one found in Table 1. This enables teachers to benefit from the use of technology in the mapping process, even if they do not have access to curriculum mapping software.

Mapping Content, Skills, and Assessment

Teachers begin by recording the content for the course or subject area. A curriculum map does not represent a daily lesson plan but reflects the major concepts and content that will be covered during that period. In facilitating the process with teachers in a variety of settings, I have found that on average, a teacher can map the content for one course or subject for the entire school year in 30 to 45 minutes.

After completing the content, the teacher identifies the key skills that will be used. The list of skills is often significantly longer than the list of content, and as a result, the skills portion of the map takes the greatest amount of time for teachers. I have found that it takes most teachers 1 to 1½ hours to complete the skills portion of the map for one course or subject area for the school year.

It is critical to identify the new skills that will be used and to be specific enough in that description and identification that it is clear to other readers. For example, instead of indicating that the students will be identifying the animals found in the rain forest, you would indicate that they would classify the animals by kingdom, phylum, and genus. When mapping skills, it is important to identify the new skill or the new context in which the skill will be applied. The more clearly the skill is identified, the more useful the map will be 3738to other teachers. Clarity regarding skills will enable special education teachers to prepare a learner for the skills that will be used and help the learner compensate for deficits in the skills so he or she can fully participate in the classroom.

The final element of the curriculum map is assessment—both formal and informal. Assessment strategies should be identified for all content and skills on the map. These could include informal assessments, such as teacher observation and student self-assessments, as well as formal assessments, such as student projects, presentations, quizzes, and traditional tests. The process of mapping assessments takes about 30 to 45 minutes to complete for one course or subject area for the year.

Mapping Time Frame

Mapping one course or subject area for the year will take about 2 to 3 hours and can be accomplished in several ways. Mapping can be completed in advance of teaching by projecting ahead for a month, a semester, or an entire year. Mapping can be done at the completion of a school year in preparation for the next year, or it can be completed month-by-month as you progress through the school year. Many teachers find it easiest to map as they go through the course of the school year and generally find that it takes only about 15 to 20 minutes a month to complete the map in this manner. Using a software program or computer template for mapping allows teachers to refine and realign their maps in an ongoing process and facilitates sharing the maps with other teachers in the building.

After all teachers complete their maps, copies of all the maps are given to all teachers in the building. Everyone reads the maps to gain an understanding of the content, skills, and assessments that will be covered in each grade level or course. Sharing maps allows teachers to gain information and identify repetitions, gaps, and potential areas for integration. Teachers then come together in mixed groups to discuss the maps and compare their findings. They determine any immediate revision points and identify any areas that require research and planning. Subcommittees are then formed to research these issues and make recommendations to the staff regarding curriculum alignment. The powerful impact of this process is that it puts decisions about curriculum alignment in the hands of the teachers who deliver the instruction.

Increased collaboration and communication among teachers ultimately benefits the students. As the curriculum alignment is achieved, students' educational experiences are enhanced. The curriculum is more coherent and clear for building knowledge and skills. In addition, instruction becomes more closely aligned to the state and district standards on which students will be tested. Finally, as teachers share information about what they teach, they begin to dialogue and share effective instructional strategies. General and special education teachers learn from each other and build strong partnerships that provide instruction to best meet the needs of their students.

Curriculum Mapping for Special Education Teachers

Special education teachers use curriculum maps to get a clear picture of the content, skills, and assessments used in the general education classroom so they can assist students with disabilities in inclusive classroom settings. The information the map provides is critical in helping special education teachers understand the instructional processes students will experience in the general education classroom. For those students with more severe disabilities, instruction is often so highly individualized that maps would have to be specific for each student to give a clear picture of the instruction. To truly communicate the appropriate information, traditional maps as completed by general education teachers would need to be created for each individual student. Because this is already done as a part of Individualized Education Programs (IEPs) the process would only increase the paperwork load for special education teachers. A different process must be used to develop communication among special education staff members.

In working in schools with special education teachers involved in curriculum mapping, I adapted a process that has been used by library media specialists for special education staff. The special education staff began to compile a list of curriculum-based resources that supported the content, skills, and assessments outlined in general education teachers' curriculum maps. These resources were entered into a searchable database that was accessible by all staff in the building (see Table 2). The database included information about the content and skills contained in the materials, along with information such as an approximation of reading level and/or the grade-level equivalency of the materials. It included any other specialized adaptive information that would assist anyone searching the database in understanding how the material might support classroom instruction. The database indicated where in the building these materials were located and the contact staff person in charge of these materials.

Thus began a process of sharing curriculum materials and other supportive resources among special education staff members, as well as between special and general education teachers. Any staff member can access these materials to support the learning needs of students who are not identified for any type of special service programs, but who may have specialized learning needs. 3839Curriculum materials that parallel the classroom content to a lower grade-level equivalency reading level could be used to support English language learners (ELL) or students with other learning delays. Teachers searching for materials to assist students in their classrooms can determine if materials that may fit their purposes are available. In addition, they know whom to contact about these materials. This often began a dialogue about strategies and materials that support learning needs of students and created a situation in which the special education teachers were able to share their specialized skills in teaching strategies with general education teachers. As teachers borrow and adapt these materials for students in the classroom, they gain more knowledge and skills in working with specialized learning needs of students with disabilities. They are better prepared to serve not only students with disabilities in their classrooms but all students in their classrooms.

After general education teachers complete their maps, special education staff code the resource database to the classroom teachers' maps, indicating those resources that specifically support the content, skills, and assessments used by the general education teachers. Not only does this facilitate the sharing of resources, it also clearly identifies those areas in which the school does not currently have many resources to support the classroom curriculum. Available budget moneys can then be directed toward the purchase or development of materials in those areas. Rather than having each special education staff member create his or her own adapted materials, educators can pool resources and expertise to find or develop appropriate materials.

Sharing this information helps all educators better direct limited budget resources and gives educators time to acquire and develop materials that best support the actual general education classroom curriculum and curriculum standards. Sharing is facilitated not only between general education teachers and special education teachers but also among program areas within and outside of special education.

Benefits of Curriculum Mapping

Although curriculum maps facilitate communication among teachers, the key benefit is improving the learning needs of all students, especially individuals with disabilities. Special education teachers are able to develop a clearer understanding of the general education classroom curriculum, along with knowledge of the skills and assessments that will be used. This information is vital for general and special education teachers who collaborate to support learning in the general education classroom. The maps also provide a strong basis for making decisions about inclusion and acquiring knowledge about the necessary level of classroom adaptation and modification to assist students with disabilities to participate in the general education classroom and curriculum. Beneficial information gained from mapping includes preteaching skills, correlating community-based outings with upcoming curriculum-based content, and using alternative assessments.

Maps give more detail about the skills and processes that will be used in the general education classroom than do traditional content-based lesson plans. Knowing the skills that will be used in upcoming lessons, special education teachers can begin to preteach skills to students before the skills are introduced in the general education classroom. This gives students more time and repetition to learn skills. When the skill is introduced in the general education classroom, these students will be able to participate at a level more comparable to their peers and will gain confidence in the ability to more fully participate in the general classroom.

Students in Mrs. Anderson's science class will be working on a rain forest project that will culminate in an essay about the rain forest. Mr. Jones, the special education teacher, works with Mrs. Anderson's curriculum map to identify the key concepts of the lesson. He prepares a graphic organizer or concept map for the students to use in class. This concept map is organized in a manner that reflects the structure and relationship of the concepts that will be highlighted in Mrs. Anderson's instruction about the rain forest. This is a type of content-enhancement routine that improves the organization of the instruction by presenting it in a learner-friendly format that emphasizes the “big picture” ideas (Boudah, Lenz, Bulgren, Schu-maker, & Deshler, 2000).

Mrs. Anderson and Mr. Jones model using the concept map for organizing instruction while students take notes or create their own concept maps. Students with disabilities receive a partially completed concept map that contains the key ideas and issues from the instruction (see Figure 1). The students add details to the concept map in each of the identified key categories during the instruction. Mr. Jones and Mrs. Anderson model how to appropriately use the concept map by adding information to a template of the map on an overhead projector. Having students fill in the information on this concept map not only helps them stay organized but provides them the multisensory approach of seeing the key concepts on the graphic organizer, hearing concepts from the teacher, and writing concepts on the map. All of this helps them retain information while focusing on the most important concepts (Friend & Bursuck, 2001).

Figure 1 Rain forest concept map.

At the end of the lesson, students review the concepts on the map and prepare questions for review, which they can then use in class or at home to review and prepare for a test. Students can use another template of the map as an organizer to outline the key ideas from their reading assignment. Finally, concept maps can become the framework for the information students will use to write their essays on the rain forest.

To assist students in writing these essays, Mr. Jones proposes to Mrs. Anderson that he teach a composition strategy called DEFENDS (Ellis & Lenz, 1987) to the science class. Mrs. Anderson is not familiar with this strategy but recognizes that the DEFENDS strategy will assist students as they write a paper defending their position on the destruction of the rain forest (see Figure 1). The strategy uses the following steps:

D Decide on an exact position

E Examine the reasons for the positions

F Form a list of points that explain each reason

E Expose the position in the first sentence

N Note each reason and supporting points

D Drive home the point in the last sentence

D Search for errors and correct

After Mr. Jones teaches the strategy to the class, he works with a small group of students who need additional assistance in the use of the strategy. When students have completed their essays, all students are asked to use the steps of the strategy to self-assess and refine them.

Curriculum maps also give special education teachers more time to develop appropriate classroom activities that parallel the classroom content for those students who may need significant modifications to participate in the general education classroom. Knowledge of the content, skills, and assessments used in the classroom will help special education teachers identify activities that will parallel general education activities and reinforce the same skills at a different level. Teachers can analyze the skills involved and determine if the student can perform the same task as other students, perform the same task with an easier step, perform the same task with different materials, or perform a different task with the same theme (Lowell-York, Doyle, & Kronberg, 1995).

In Mrs. Anderson's science class, students are classifying types of animals by kingdom, phylum, and genus. A student who is able to do the same task with an easier step may be classifying an animal only by kingdom. A student who needs to undertake the same task with different materials may be using picture cards with the name and pictures of animals. A student who needs to tackle a different task with the same theme may be naming animals or determining if they live on land or water.

Knowledge of the content, skills, and assessments that are part of the general education curriculum assists special education teachers in planning community-based learning experiences that support the content being taught in inclusive settings. Using the community-based experiences to support inclusive classroom learning can also provide opportunities for special education students to share what they have gained with the general education students. If 4041the science class is studying reptiles, a community-based learning experience might include a trip to a local pet store or zoo. Students may take along an instant picture or digital camera to record the reptiles they see on the outing, or they may gather information about the reptiles to share with their classmates when they return to school. The photos and information gathered can become a part of the curriculum materials for the special education students as well as supporting materials for the general education teachers and students.

Finally, assessment information included on the curriculum maps will help special education teachers understand how general education teachers will be assessing students' accomplishment in terms of the knowledge, skills, and processes in the curriculum. Special education teachers can assist students in developing study guides and equip students with test-taking strategies that fit the assessments used by general education teachers.

Special education teachers can use samples of classroom projects and assessments to build a portfolio that will demonstrate the attainment of IEP goals. In addition, information on the curriculum map offers general and special educators the opportunity to collaborate on alternative methods of assessing student knowledge. Because of the needs of their students, many special education teachers have a great deal to offer general education teachers in the development of assessment methods that do not rely solely on traditional tests and quizzes. As general education teachers collaborate on designing these alternate assessments, they improve their own skills in using multiple assessment methods.

The greatest benefit of using curriculum maps is the improved communication among all teachers in the school. As special and general education teachers have a better level of understanding of the content, skills, and assessments used in classrooms, they can build stronger collaborations to assist all students with special learning needs. General education teachers can gain a wealth of knowledge about strategies and structures that support learning from special education teachers. Special education teachers benefit from curriculum mapping by gaining a deeper understanding of the general classroom curriculum and how they can create meaningful curricular connections for students. Improved communication among all teachers in the school provides professional educators with another tool for effectively enhancing the learning of all students in the classroom, especially students with disabilities.

41 42

References

Boudah, D. J., Lenz, B. K., Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (2000). Don't water down! Enhance content learning through the unit organizer routine. Teaching Exceptional Children, 32(3), 48–56.

Ellis, E., & Lenz, K. (1987). A component analysis of effective learning strategies for LD students. Learning Disabilities Focus, 2, 94–107.

Friend, M., & Bursuck, W. D. (2001). Including students with special needs: A practical guide for classroom teachers (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Hayes-Jacobs, H. (1997). Mapping the big picture: Integrating curriculum & assessment K–12. Alexandria, VA: Association for Supervision and Curriculum Development.

Lowell-York, J., Doyle, M. E., & Kronberg, R. (1995). Curriculum as everything students learn in school: Individualizing learning opportunities. Baltimore: Brookes.

Angela Koppang, PhD, is an assistant professor in the Educational Leadership Department at the University of North Dakota and specializes in the areas of curriculum, instruction, assessment, and school leadership. She is a former elementary and middle school teacher and administrator and serves as a consultant in the areas of curriculum development and alignment. Address: Angela Koppang, Department of Educational Leadership, Box 7189, Grand Forks, ND 58201; e-mail: angela\_koppang@und.nodak.edu.

From Intervention in School and Clinic, January 2004. Copyright © 2004 by Pro-Ed, Inc. Reprinted by permission.