EDUCATIONSECTOR REPORTS

November 2008

MEASURING SKILLS FOR THE 21ST CENTURY

By Elena Silva

ACKNOWLEDGEMENTS

Thanks to my Education Sector colleagues Kelly Bathgate, Kevin Carey, Stacey Jordan, Rob Manwaring, Renée Rybak, and Bill Tucker for their help in thinking about the issues of assessment and student learning outcomes, and to Thomas Toch and Andrew Rotherham for their support in the writing and editing of this paper. Thanks also to Abdul Kargbo for his support in designing graphics and formatting the report. Robin Smiles deserves special thanks for her careful editing and diligence in managing the production of this report. I would also like to thank Cecilia Le, who contributed substantial time and effort to this project as a Fordham Fellow for Education Sector, and Claira Williams, who also provided valuable research assistance. My sincere appreciation to the many people who were kind enough to read and comment on an earlier draft of this paper. I am particularly grateful to the students, teachers, and directors at St. Andrew's School in Middletown, Delaware, and at New Technology High School in Napa, California, who graciously hosted our visits to their schools and classrooms and spent hours collectively speaking to us about their teaching and assessment practices. Thanks also to An-Me Chung of the Charles Stewart Mott Foundati on for her ongoing support of this work.

This report was funded by the Charles Stewart Mott Foundation. Education Sector thanks the foundation for their support. The views expressed in the paper are those of the author alone.

ABOUT THE AUTHOR

ELENA SILVA is a senior policy analyst at Education Sector, where she oversees the organization's teacher quality work. She can be reached at esilva@educationsector.org.

ABOUT EDUCATION SECTOR

Education Sector is an independent think tank that challenges conventional thinking in education policy. We are a nonprofit, nonpartisan organization committed to achieving measurable impact in education, both by improving existing reform initiatives and by developing new, innovative solutions to our nation's most pressing education problems.

1201 Connecticut Ave., N.W., Suite 850, Washington, D.C. 20036
202.552.2840 • www.educationsector.org

ABOUT THIS SERIES

This report is a product of Education Sector's Next Generation of Accountability initiative. The initiative seeks to strengthen public education by examining key elements of accountability, for instance, who should be responsible for student success and how they should be held responsible. Our work seeks to build on the strengths of current school accountability systems, more fully and effectively measure the depth and breadth of students' educational experiences, and encourage educators, parents, policymakers, and the larger public to pursue educational equity and excellence for all students.

When ninth-graders at St. Andrew's School, a private boarding school in Middletown, Delaware, sat down last year to take the school's College Work and Readiness Assessment (CWRA), they faced the sort of problems that often stump city officials and administrators, but rarely show up on standardized tests, such as how to manage traffic congestion caused by population growth. "I proposed a new transportation system for the city," said one student describing his answer. "It's expensive, but it will cut pollution."[1]

Students were given research reports, budgets, and other documents to help draft their answers, and they were expected to demonstrate proficiency in subjects like reading and math as well as mastery of broader and more sophisticated skills like evaluating and analyzing information and thinking creatively about how to apply information to real-world problems.

Not many public school students take assessments like the CWRA. Instead, most students take tests that are primarily multiple-choice measures of lower-level skills in reading and math, such as the ability to recall or restate facts from reading passages and to handle arithmetic-based questions in math. These types of tests are useful for meeting the proficiency goals of the federal No Child Left Behind Act (NCLB) and state accountability systems. But leaders in business, government, and higher education are increasingly emphatic in saying that such tests don't do enough. The intellectual demands of 21st century work, today's leaders say, require assessments that measure more advanced skills, 21st century skills. Today, they say, college students, workers, and citizens must be able to solve multifaceted problems by thinking creatively and generating original ideas from multiple sources of information—and tests must measure students' capacity to do such work.

While many policymakers, including Secretary of Education Margaret Spellings, have emphasized the need for schools to, first and foremost, teach the basics, learning science—an interdisciplinary field that includes cognitive science, educational psychology, information science, and neuroscience—suggests that the best learning occurs when basic skills are taught in combination with complex thinking skills. Decades of research reveals that there is, in fact, no reason to separate the acquisition of learning core content and basic skills like reading and computation from more advanced analytical and thinking skills, even in the earliest grades.

But standing in the way of incorporating 21st century skills into teaching and learning are widespread concerns about measurement. The cost, time demands, and difficulty in scoring tests of these less easily quantified skills have slowed the adoption of such tests, as have concerns among civil rights advocates that these tests would erode progress toward ensuring common standards of learning for all students. Collectively, these concerns derailed efforts in the late 1990s to move toward the use of performance-based assessments such as portfolios, exhibitions, and projects.

New assessments like the CWRA, however, illustrate that the skills that really matter for the 21st century—the ability to think creatively and to evaluate and analyze information—can be measured accurately and in a common and comparable way. These emergent models also demonstrate the potential to measure these complex thinking skills at the same time that we measure a student's mastery of core content or basic skills and knowledge. There is, then, no need for *more* tests to measure advanced skills. Rather, there is a need for *better* tests that measure more of the skills students' need to succeed today.

Unmet Challenges

The idea that schools should focus on more than just the basics is not new. A century ago, leaders of the progressive education movement, spearheaded by American philosopher and educator John Dewey,

argued for an education system that teaches more than just the basics of core academic subjects. Such calls, however, have intensified in the past two decades as the nature of the economy and work has changed. Several major reports in the 1990s prompted renewed attention to critical thinking in education. One, issued by the U.S. Department of Labor Secretary's Commission on Achieving Necessary Skills, challenged schools to teach not only basic skills but also the ability to think creatively and acquire and analyze information.[2]

More recently, the New Commission on the Skills of the American Workforce—a group of business leaders, governors, school chancellors, and former secretaries of labor and education—released a sequel to its 1990 report on the nation's educational and economic challenges. The message of the 2006 report, *Tough Choices or Tough Times*, is clear: Basic skills are necessary but not sufficient.

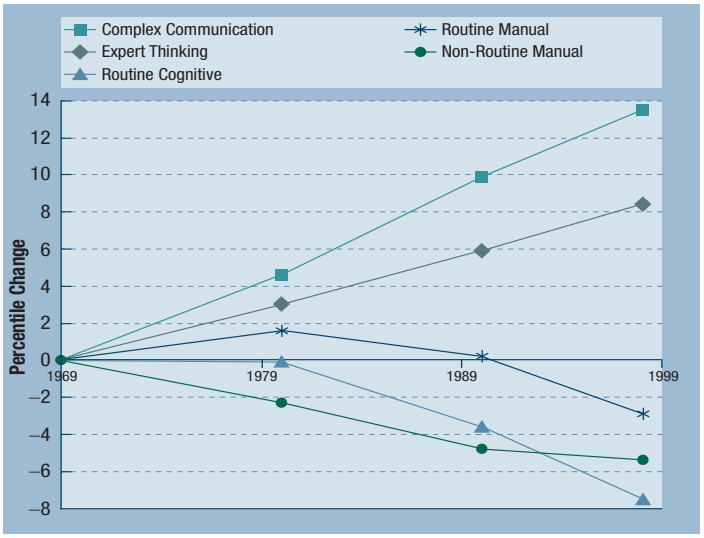
The commission's report describes how new technology and global competition have changed the game for American workers. Students need a strong foundation of basic skills, the commission asserts, but that alone is no longer enough for economic and job security. "It is a world in which comfort with ideas and abstractions is the passport to a good job, in which creativity and innovation are the key to the good life, in which high levels of education—a very different kind of education than most of us have had—are going to be the only security there is."[3]

This new reality applies to all children in the United States, not just an elite class of students. Nearly every segment of the workforce now requires employees to know how to do more than simple procedures—they look for workers who can recognize what kind of information matters, why it matters, and how it connects and applies to other information.

Richard Murnane and Frank Levy, both economists and professors at Harvard and MIT, respectively, have been researching and writing about workforce skills for more than a decade. They agree that basic skills, once in high demand for workers, are no longer what matter most. There are fewer tasks requiring only routine skills, they explain, and they are often done by computers.[4] (See Figure 1.

Concerns that the United States is losing its global competitive edge are heightened by the nation's

Figure 1. Skills for a New Economy



Source: Frank Levy and Richard Murnane, *The New Division of Labor: How Computers Are Creating the Next Job Market* (Princeton, NJ: Princeton University Press, 2004).

performance on the most recent international tests. The Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), two of the largest education surveys in the world, measure how well early adolescent students (PISA tests 15-year-olds and TIMSS tests the rough equivalent of eighth-graders) are faring in their abilities to problem-solve in math and science.[5] TIMSS found U.S. eighth-graders to be above average performers among participating nations and found substantial improvement in performance, particularly in science, from the 1999 to 2003 tests. But the PISA, designed to test students' application of math and science to real-world scenarios, found U.S. students to be among the worst performers. (See Table 1.) Taken together, these results reveal that U.S. students may be performing well in their mastery of instructional material but that this performance is not carrying over to the application of material to real-world problems.

The 'Must Have' Skills

It is an emphasis on what students can do with knowledge, rather than what units of knowledge they have, that best describes the essence of 21st century skills.

But that core notion is often lost in the welter of terms used to describe 21st century skills and in the many

Table 1. PISA 2003, 15-Year-Olds, Problem-Solving

Top 10 nations and the United States	
Country	**PISA Score**
Korea	550
Finland	548
Japan	547
New Zealand	533
Australia	530
Canada	529
Belgium	525
Switzerland	521
Netherlands	520
France	519
United States	**477**

Note: The OECD average is 500. Of 29 OECD countries participating in PISA 2003, only three countries—Greece, Turkey, and Mexico—scored below the United States.

Source: Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., Jocelyn, L. (2004). International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003. Results From the U.S. Perspective. (NCES 2005–003). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

sub-skills often included under the concept. Workforce and management-training groups often call 21st century skills "soft" or "interpersonal" skills. Vocational education programs call them "applied" skills or "workforce" skills. Many youth development programs refer to them as "life and career" skills. And researchers often use the term "non-cognitive" skills. "Technology literacy" is a frequently mentioned 21st century sub-skill. But it's defined in myriad ways. To various educational organizations and businesses, it's information-science skills, digital media fluency, advanced computer and internet communications, and "technacy," a newer term used to describe a deep knowledge of technological systems.

A number of organizations have developed frameworks that attempt to identify the individual skills and sets of skills students need to succeed and to help educators integrate 21st century skills into existing education programs. The *enGauge* 21st Century Skills framework, for instance, developed by the North Central Regional Educational Laboratory, includes "digital-age literacy," "inventive thinking," "effective communication," and "high productivity" as the most important skill sets.[6]

The Partnership for 21st Century Skills, a network of nearly 30 major businesses and education groups and one of the strongest advocates for infusing 21st century skills into education, has developed a framework for "21st century learning" with the intent to help states, districts, and schools integrate core subject learning with 21st century skills. Similarly, the U.S. Conference of Mayors passed a policy resolution in 2005 supporting a framework for 21st century skills that encourages citywide policies and programs aimed at preparing students with a more comprehensive set of skills.[7]

The framework with potentially the widest reach is that of the Definition and Selection of Competencies Project, created by the Organization for Economic Cooperation and Development (OECD), an organization of 30 industrialized nations. This framework describes a set of key competencies—for instance, the ability to consider the wider context of decisions and actions—that marry the need for basic literacy with essential deep conceptual understanding.[8] This framework helped to define OECD's long-term strategy for assessing competencies of young people, including its development of the PISA.

A Learning Imperative

At the same time, studies by national and international research organizations, including the National Research Council, OECD, and the International Society for Technology in Education, have shown that complex thinking and analytical skills are an integral part of learning at every stage of development.[9]

For decades, educators have relied on the principles of Bloom's Taxonomy of Educational Objectives to outline teaching practices, develop curriculum, and create testing standards.[10] The widely used rubric sequences learning on a continuum from lower-level to higher-level skills based on the belief that learning is a linear process—that the ability to develop a particular skill, such as constructing a flow chart that describes how and why certain historical events led to others, is necessarily preceded by the development of another particular skill, such as recalling a timeline of historical dates. Teachers have been trained to "move students up" through this continuum of skills, beginning with the acquisition of knowledge and eventually getting to skills like analytical thinking.

An After-School Opportunity

With a long history of imparting skills like problem-solving, inquiry, and critical thinking, and connecting these skills to academic goals, the after-school or out-of-school field seems fertile territory for finding ways to measure 21st century skills.

But the push for accountability in education by way of NCLB has been a mixed blessing for the after-school world. On the one hand, the demand for results has compelled the field to make significant strides in the development of better and more appropriate evaluation and assessment tools. The National Institute on Out-of-School Time, for example, recently worked with the Massachusetts Department of Education to develop the Afterschool Program Assessment System (APAS), which is designed to help after-school programs improve their quality and reach their desired outcomes. APAS includes a tool, the Survey of After-school Youth Outcomes (SAYO), which measures changes in youth outcomes over time. The SAYO, first used by Massachusetts in 2004 and now expanding to Georgia and North Carolina, uses pre- and post-participation surveys of teachers and after-school staff to measure changes in youth behaviors that are aligned with the outcome goals of the program. While the tool is not intended to be used as a diagnostic instrument for individual youth, it enables the collection of a huge amount of demographic information and outcome data and is a step in the right direction for the after- and out-of-school field.

Toolfind is another Massachusetts-based development. The United Way of Massachusetts Bay developed the Toolfind database in 2006 to identify and share psychometrically sound tools that can be used by average practitioners. It now sits as a database of 46 tested tools in 11 outcomes areas including problem-solving, positive behavior (self-control, cooperation, conduct in school, responsibility), leadership, learning orientation (motivation, persistence, study habits), and academic skills.

On the other hand, the after-school field is not resistant to the pressures of school-based accountability. Many programs align their work and assess their success based principally on academic outcomes.* Citizen Schools, a national network of after-school education programs for middle school students, provides a useful illustration of this. Citizen Schools was designed to impart skills like critical thinking, teamwork, and communication. But they assess their performance strictly in terms of school-based academic measures, such as reading and math scores.

Liz Reisner, principal at Policy Studies Associates, conducted a recent evaluation of Citizen Schools. She explains that while there is strong interest within Citizen Schools to show how the program teaches a broader set of skills, their work is still driven by stakeholders and funders who expect to see success measured by academic gains. There are ways to measure some of these skills, Reisner says. For example, it might be possible to assess decision-making skills by analyzing the middle school participants' selection of high-quality college preparatory high schools. "We're working on it," Reisner says, "but without incentives to measure these skills, there's not a lot of attention [to measuring the stated goals of critical thinking, teamwork, and communication]."

Measuring success by school-based academic outcomes is understandable for two reasons. First, the call for measures of impact in after-school programs that are more rigorous than commonly used satisfaction surveys is not only an added task but a daunting one. Most programs, operating independently with varying purposes and goals, do not have staff with expertise, or time to develop expertise, in assessment and evaluation. The simplest approach for these programs, then, is to adopt existing acceptable indicators of success. In most cases, school-based data on achievement, often in the form of test scores, serve this purpose.

Second, the after-school world is highly dependent on funding from the 21st Century Community Learning Centers (CCLC), a federal funding stream of nearly $1 billion that targets students at high-poverty and low-performing schools.** "When the program began in 1998," says Jen Rinehart, vice president for research and policy at the Afterschool Alliance, "there wasn't much of a focus on school at all. NCLB changed that, and CCLC moved from a community learning center model to a focus almost entirely on academic outcomes, although notably didn't change its name to reflect this difference. That's had a major impact on the focus of after-school programs."

This shift to adopt school-based academic goals is a problem for the after-school world, which has defined itself and its relevance by its ability to support not just the academic but also the social, emotional, and physical development of children. In effect, the after-school field risks being relegated to a large-scale tutorial program, supporting narrow proficiency goals of school at the expense of its broader youth development goals.

The pressures on the after-school field to align with the goals of school-based accountability are substantial. But they also present an opportunity for the field. By combining its deep roots in youth development with promising advances in assessment, the after-school field is poised to play a powerful role in informing an educational assessment system that measures a more comprehensive set of skills. As schools look to teach more than the basics, and to evaluate their success in doing so, they will need more than supplemental help. A recent report by Heather Weiss and Priscilla Little, both researchers at the Harvard Family Research Project, points to a growing body of research that shows that quality after-school programs can improve a wide range of outcomes for youth. Weiss concludes that the best scenario for students "is the one where out-of-school programs are recognized for the range of learning they provide and work with, not against, schools to provide them."†

*Survey of published goals and outcomes of after-school and out-of-school programs, including 21st Century Community Learning Centers grantees, Spring, 2007.

**The Afterschool Alliance estimates the total funding of the after-school industry to be $3.5 billion. 21st CCLC represents the largest pot of funding for after- and out-of-school programming.

†Interview with Heather Weiss, July 2008.

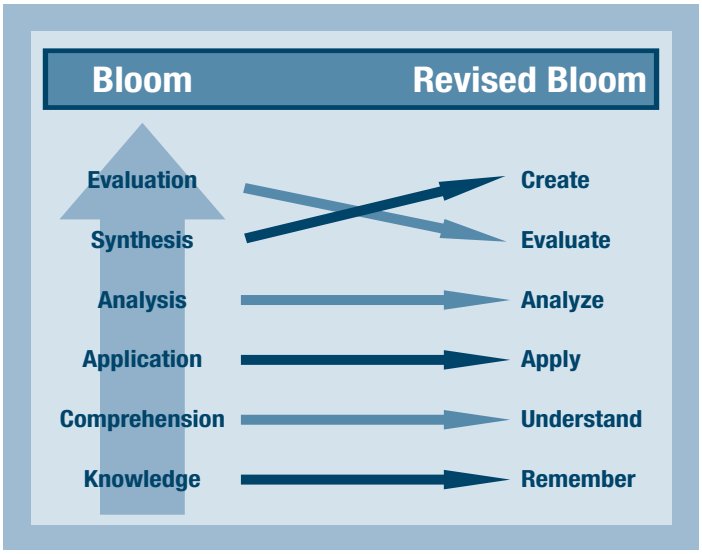
But two former students of Benjamin Bloom, the researcher who developed the taxonomy in the early 1950s, published a new version of the taxonomy in 2001 based on the new findings that most skills can be gained and employed simultaneously or out of order. "This is different from the old taxonomy, which said, for example, that you cannot apply until you comprehend, or that you must understand before you can analyze," explains co-author Lorin Anderson, who with David Krathwohl convened a working group that spent five years revising Bloom's taxonomy.[11] "We now know that, in many instances, these processes can be learned at the same time, or even in reverse order." (See Figure 2.)

The notion that basic and advanced skills are best learned together is one of the major findings of a recent report on mathematics education, funded and released by the U.S. Department of Education. The best learning happens, the report asserts, when students learn basic content and processes, such as the rules and procedures of arithmetic, at the same time that they learn how to think and solve problems.

The mathematics report also concluded that there is no set age or developmental stage when children are ready to gain complex thinking skills. This is in sharp contrast to the previously held notion that very young children are concrete and simplistic thinkers who cannot think abstractly or gain deep understanding of concepts. Thus, while there are building blocks of knowledge—students must master addition and subtraction before they multiply or divide—the idea that students should be taught facts and simple procedures before they get to problem-solving or critical thinking no longer makes sense. "The common idea that we can teach thinking without a solid foundation of knowledge must be abandoned. So must the idea that we can teach knowledge without engaging students in thinking. Knowledge and thinking must be intimately joined," says Lauren Resnick, a professor of psychology at the University of Pittsburgh and a leading expert on cognitive science.[12]

Teaching children basic facts and simple procedures in a way that helps them also learn how to apply and use this knowledge and these skills mirrors the natural process of learning. So the integration of advanced thinking and analytical skills into teaching and learning makes it easier for students to acquire even the most basic skills and core knowledge.

Figure 2. A New Taxonomy



Source: A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, 2001.

The belief that there should be a solid, specific, and shared core curriculum, an idea advanced most notably by the nonprofit Core Knowledge Foundation, founded and led by former professor and literary scholar, E.D. Hirsch Jr., is not at odds with this approach. The Core Knowledge curriculum supports the point that learning factual knowledge and the ability to apply, analyze, and solve problems go hand-in-hand. Teachers using the Core Knowledge approach do not stress rote memorization of facts; they use an array of strategies including workshops, research projects, dramatizations, and collaborative learning groups because they know that students will learn best if they are exposed to both subject knowledge and ways to apply this knowledge at the same time.

The attributes that business and higher education leaders are calling for in young people—that they be independent thinkers, problem-solvers, and decision-makers—are captured by the advanced skills in the revised Bloom's taxonomy, the ability to analyze, evaluate, and create.[13]

"What is unique to 'create,' 'evaluate,' and 'analyze' is that the content is not explicit in the process or product being created," write Anderson and his colleagues. This is particularly true for "create," they explain, which requires a student to use existing information to come up with something entirely original—a new idea, a unique product, an alternative solution—tied to a specific purpose.[14] Without these processes, Anderson says, "people,

when faced with a problem or a challenge, will either call someone for help or just quit."

Integrating 21st century skills into teaching and assessment, then, is not only an economic imperative, driven by changes in the workforce, but a vital aspect of improving student learning.

Creative Measures

Yet, there remains an assumption that 21st century skills cannot be fairly or reliably measured. Most existing tests measure only whether a student possesses a particular piece of knowledge, not whether the student can analyze this information, evaluate its utility, or create new knowledge from it—the core of 21st century skills. But new models of assessment that measure both basic skills and more advanced skills are emerging to challenge the assumption that such skills can not be measured and to move us toward an assessment system that is more aligned with what students now need to know.

The CWRA, used by St. Andrew's School in Delaware, offers one example. It consists of a single 90-minute task that students must respond to using a library of online documents, from one-page newspaper editorials to 20-page research reports. Facing problems like a city beset by pollution from a now-defunct factory or a community health clinic struggling to serve a growing immigrant population, students must grapple with real-world dilemmas; make judgments that have economic, social, and environmental implications; and articulate a solution in writing.

The CWRA grew out of the Collegiate Learning Assessment (CLA), developed by the Council for Aid to Education and the RAND Corporation. (See "Learning From Higher Education" sidebar on Page 7.) Like the CWRA, the CLA is a single test that measures analysis and writing skills. But, while the CLA is used by more than 175 higher education institutions, the CWRA is in use by only a handful of private schools (like St. Andrew's, which began using it a few years ago) and a single Long Island, N.Y., public school.

The CWRA is intended as a tool for school improvement, not necessarily to measure individual student gains. But those who use it affirm its value as an essential metric for student learning: "Are we teaching our students to think intelligently and critically, to do more than just follow or even lead, but to find new paths to go down? That's what we learn from [the CWRA]," says John Austin, the academic dean of St. Andrew's.[15]

New technologies are making it easier to measure individual student mastery of 21st century skills. River City, for example, is a "virtual world" that simultaneously teaches and assesses middle school science students. Like other simulated learning programs in education and a range of other industries, River City presents students with a problem and asks them to develop a hypothesis and procedure, test it—virtually—and then describe their findings and make recommendations in a report.[16]

From a technical standpoint, these "multi-user virtual environment" tools are among the most advanced performance assessments that now exist. They can keep detailed records of the moment-by-moment movements and decisions of each participant in the environment and provide a log for each student in each session. Teachers, then, can track the progress of individual students.

Programs like River City are good for tracking student gains at the classroom or school level, but the true test for accountability will be if assessments like this can work on a larger scale. Several promising examples move toward this goal, taking the necessary step of linking these types of assessments to existing state or national standards.

Researchers at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California-Los Angeles have developed an assessment that measures complex thinking and judgment skills within the existing framework of state math assessments. The system, called PowerSource, is funded by a U.S. Department of Education Institute of Education Sciences grant and is now being piloted in nearly 70 schools. Designed for middle school pre-Algebra, it consists of multiple interim assessments that are formatted as narrative themes or graphic novels. "It's still an experiment," explains Eva Baker, who directs the center. "But it has real promise for improving instruction and for demonstrating mastery of a broad set of skills."[17]

PowerSource measures advanced skills in the context of measuring content proficiency, which means it can demonstrate student learning for specific subject matter

Learning From Higher Education

Colleges and universities, while free from the accountability pressures faced by K–12 schools and after-school institutions, have good reason to focus on 21st century skills.

The American Association of Colleges and Universities stressed in its 2002 report, *Greater Expectations: A New Vision for Learning as a Nation Goes to College*, that the current emphasis on "factual recall" is a major barrier to success in college.* Today's college students, the report concludes, need to be "integrative thinkers who can see connections in seemingly disparate information and draw on a wide range of knowledge to make decisions." Colleges and universities are thusly focused on ensuring that their students and prospective students have a set of strong creative and analytical skills.

New models for assessing these skills have emerged in the last few years. The Collegiate Learning Assessment (CLA), the parent test of the CWRA, was recently highlighted by the U.S. Secretary of Education's Commission on the Future of Higher Education as a promising approach for the assessment of student learning at colleges and universities.** Developed for the Council for Aid to Education by former Rand researchers, Steven Klein and Richard Shavelson, the CLA is comprised of a single test that measures writing skills and analysis and problem-solving skills at the same time. It is now in use by more than 175 higher education institutions.

"We are testing a broader range of outcomes for college students that are relevant to their future as workers and citizens," says Shavelson. "What we ultimately want is for people to behave intelligently—to think critically, reason analytically, and make decisions that are justifiable."***

Robert Sternberg, dean of the School of Arts and Sciences and psychology professor at Tufts University, had a similar goal in mind—to measure more relevant skills—when he developed a series of alternative assessments designed to measure student "creativity, practicality, and wisdom." The tests, administered in written form and by video, ask students to write stories, form captions for untitled cartoons, and solve everyday problems such as moving a large bed up a winding staircase. Sternberg administered his tests to more than 1,000 college freshmen and high school seniors from 15 schools as part of what he named the Rainbow Project. He found that his tests predicted student grades as college freshmen twice as well as SAT scores and high school grade point averages, and that the tests reduced ethnic-group difference in scores compared to the SAT.†

On the 2006–07 application for Tufts undergraduates, slightly more than half the applicants or about 8,000 chose to participate in the assessment.‡ Sternberg said that, as predicted by the earlier Rainbow Project, Tufts admitted better-qualified applicants as defined by the traditional measures of SAT score and grade-point average and also increased diversity, admitting 30 percent more blacks and 15 percent more Hispanics than in the previous year.

Sternberg encourages other colleges and high schools to assess this broader set of skills and freely shares his materials. The selective independent high schools Choate Rosemary Hall in Wallingford, Conn., and Phillips Academy Andover in Massachusetts are using similar measures in their admissions processes. Sternberg laments that many of the high schools making efforts to teach and assess these skills educate an already advanced student body, while the primary focus with underserved students is on basic proficiency. "Those students are going to be the ones most likely to capitalize on their practical skills," he said.

*Association of American Colleges and Universities (AAC&U), *Greater Expectations: A New Vision for Learning as a Nation Goes to College*, 2002.
**U.S. Department of Education, Washington, D.C., 2006. http://www.ed.gov/about/bdscomm/list/hiedfuture/index.html.
***Personal interview with Richard Shavelson, March, 2008.
†Interview and correspondence with Robert Sternberg, February 2008.
‡Correspondence with Tufts University Admissions Office, March, 2008.

while also testing students' development of higher-order skills. Students are asked, for example, to apply algebraic principles as well as explain *why* they chose the principles. PowerSource was designed to measure a broader set of outcomes by focusing on a handful of big ideas rather than a heap of discrete facts.[18]

The United Kingdom recently developed an innovative national assessment that aligns with its existing national standards. The Key Stage 3 (ages 12–13) Information Communications Technology (ICT) Literacy Assessment, created by the British government's Qualifications and Curriculum Authority, measures a set of technical skills as well as a student's ability to use those skills to solve a set of complex problems. Students are provided a toolkit of applications to use to complete tasks that measure learning skills such as "finding things out," "developing ideas," and "exchanging and sharing information." Student actions are tracked and mapped against expected abilities for that level of education and test results provide both national scores for students and detailed feedback about student performance that can be used to inform teaching and learning at the classroom level.[19]

The closest thing in the United States is the 2009 NAEP Science Assessment, administered by the federally funded National Assessment of Educational Progress (NAEP). The test will for the first time measure not just students' knowledge of science principles, but also

whether they can apply their knowledge. This is a big change for NAEP and represents a potential move toward national assessments that provide a richer picture of student mastery of science content and the scientific inquiry process.

The International Baccalaureate (IB) Diploma Programme, a rigorous two-year high school course of study taught in more than 2,000 public and private schools in 130 countries under the auspices of the nonprofit International Baccalaureate Organization, serves as evidence that the assessment of core content and advanced skills, aligned with a program of standards and curriculum, can happen at a large, even international, scale. The 40-year-old program is built on the principle that students can and should master both basic subject matter and higher-order skills. The program has developed common curricula, standards, and assessments, which are used throughout its school network and has developed strategies for ensuring the standardization of both teaching and teacher-graded testing.[20]

The program assesses student performance using a range of techniques, both internal (classroom-based, teacher-led) and external. All courses, for example, have three or four separate assessment components, none worth less than 20 percent or more than 50 percent of the overall assessment. Each component includes a range of performance tasks in various formats appropriate to the subject matter, which could include multiple-choice questions, short-response questions, structured and open-ended problem-solving questions, data analysis questions, case studies, and essay questions.

The testing industry, keenly aware that the call for 21st century skills means more demand for tests that measure these skills, has also been working hard to develop assessments that measure more than the basics.

The Educational Testing Service (ETS), a private nonprofit organization and one of the world's largest test developers, has several initiatives underway to measure a broader set of skills, including critical thinking, communication, and a host of socio-emotional skills like adaptability and agreeableness. Researchers at the organization's Center for New Constructs are studying how to assess skills that are not measured by the SAT, ACT, and other traditional standardized tests. Pilot projects of more than 20 individual assessments that measure analytical and a host of other skills like negotiation and teamwork are taking place on college campuses and in several school districts.[21]

Richard Roberts, one of the center's principal research scientists, says that these tests, by describing a problem-solving scenario such as deciding what to do when a group project that needs weeks of work is due in days, challenge students to think in ways that most tests do not. "We are not just asking students 'what should you do in this situation?' but also 'what would you do in this situation?' Students have to think differently—they have to think deeper—to answer the second question because there is simply no right answer," explains Roberts.[22]

Cost, Time, Quality, and More

Measuring 21st century skills on a large scale is not going to be an easy task. For one, assessments like the CWRA are expensive. The Government Accounting Office estimated in a 2003 report that the cost to score North Carolina's multiple choice, machine-scored assessments was approximately 60 cents per test.[23] The cost for Massachusetts, with its combination of multiple-choice and open-ended questions, was approximately $7 per test. In contrast, the cost to score the CWRA's performance task is more than $40 per test, although this is still a small portion of the roughly $8000 spent on education per student. Even this is considered by those in the testing industry to be inexpensive compared to the cost of large-scale performance-based assessments requiring human observation or scoring.

Using people to grade a wide range of open-ended and performance-based assessments of 21st century skills raises concerns about the reliability of results. People may be able to assess in more depth and with more nuance than a computer program, but human scorers inevitably introduce a level of subjectivity into the assessment process. So-called "inter-rater reliability" is a challenge; no matter how clear scoring standards are it's difficult to expect human scorers to grade tests with perfect consistency. Training and monitoring scorers can be time-consuming and costly, but it can help. The IB Diploma Programme, for instance, has nearly 5,000 test examiners worldwide. The program ensures a high level of consistency among its examiners, most of whom are experienced Diploma Programme teachers, by providing

A Certain Type of Teacher?

The New Technology High School model was founded in 1996 when local businesses in Napa, Calif., began complaining that students were unprepared for high-tech jobs. The model incorporates project-based learning, small school size, one computer for every student, and an environment in which students are responsible for their own learning. Within three years of the school's founding, the Bill and Melinda Gates Foundation stepped in to help replicate the model, which costs approximately $800 more per student per year than a traditional school. There are now 40 New Technology schools in nine states; five are charter schools, and 21 are small schools within a larger high school.

Student learning at New Technology is designed to simulate real life and real work. Instead of completing traditional worksheets and daily assignments, students are assigned periodic projects, often as teams, and must complete a combination of products, including written essays and practical demonstrations. Each project assigned to students is accompanied by a set of rubrics that measure a student's performance on fundamental skills, like writing, as well as criteria such as critical thinking, application, and originality. Students receive multiple grades, one for each criterion, for each project.

"In a standard school," explains a world studies teacher at the Napa-based New Technology High, "you would read a paper and say, 'Wow. This student is not a strong writer but has some good ideas.' But there's no way to communicate that in the grade. Here, I could give them a low C for written communications, but a higher grade in another area. So I know … and the student knows what areas we need to really work on."

This type of project-based learning is made easier by a suite of Web-based tools used to track and improve student performance. Teachers, students, and parents have constant access to an online system that provides detailed information about how students are progressing and how they can improve.

Teachers acknowledge that it can be a challenge to teach this way. "We take it for granted that students know how to learn, but the reality is most of them don't," said a Spanish teacher. "It's our job to teach them how to learn—but it's not always easy, particularly with the students who struggle with poor time management and initiative." It requires a lot of work and perhaps a certain type of teacher. A recent survey of teachers conducted by the Buck Institute for Education compared teachers working in newer reform models using project-based learning, with teachers working in traditional high schools using conventional instructional approaches. Teachers in the reform model schools, including a sample of 71 New Technology teachers, tended to be newer and younger teachers with more training in and exposure to project-based learning and the technologies necessary to support this type of instruction.

New Technology is expanding quickly (it plans to open 10-12 additional schools next year). But to take models like New Technology truly to scale in public education is no small task. It means finding teachers who are prepared to use new methods and new tools to teach not only the high-performing students who come ready for challenging real-world projects but also low-performing students.

detailed instructions on scoring for each assignment, requiring each examiner to submit samples of their scoring, and employing a cadre of senior examiners who determine scoring standards and monitor examiner work.

The cost and time required to write, administer, and score open-ended questions has led many state policymakers to increase the number of multiple-choice questions in testing. As Thomas Toch has written in the Education Sector report *Margins of Error: The Testing Industry in the No Child Left Behind Era*, the percentage of public school students taking statewide tests with no open-ended questions reached 42 percent two years ago, a figure that has continued to climb.[24] Earlier this year, state policymakers in Kentucky looked to save as much as $10 million by removing open-response questions from the state's accountability system. "Is this open-response test worth the [money and] six to eight weeks of resources when we could be spending time on task and then be getting more valuable information from the norm-referenced test?" asked state Senate Majority Leader Dan Kelly, who led the state's push to eliminate portions of the assessment system.[25] Kelly's question is a fair one, given the scope of statewide assessment. Giving tests with fewer, less standardized items invariably reduces the reliability of results.[26]

The research community is grappling with another reliability question: whether 21st century skills can be coached or "faked" on a test. A student, for example, might answer in ways that suggest she is an analytical thinker when in fact she is merely demonstrating that she has learned what types of answers make her seem that way. This potential problem is a focal point of research on measurement and testing. The Center for New Constructs at ETS, for example, sponsored an entire conference in 2006 on the issue of faking.[27] With general agreement that faking can happen and that it can skew results, participants discussed the utility of new types of questions, including more subtle questions or more complex "forced choice" questions, which can detect and correct faking.

Advancements in assessment technology can answer many of these concerns. Simulation-based assessments, for example, like River City, are able to assess students' understanding of complex problems using multiple-choice formats that are automatically scored, making them not only cheaper and more efficient, but also more reliable.

And computer-adaptive tests, which adjust the difficulty of questions based on students' performance on previous questions, can not only be scored immediately but also make faking far more difficult since the test changes with the individual test-taker. Delaware, one of several states experimenting with computer-adaptive tests, recently completed a pilot of 30,000 students in four districts, and found that the adaptive tests were actually better at identifying student growth than existing grade-level tests.[28]

Still, better assessment is only one part of delivering better learning to all students. The basic principle that there is no real choice between basic and 21st century skills—that both are essential learning outcomes for students—must also apply to standards and curriculum. Even more important, delivering better learning hinges on preparing and supporting quality teachers who can deliver the "must have" combination of basic and advanced learning to all students. (See "A Certain Type of Teacher?" sidebar on Page 9.)

In the long run, new forms of assessment, as well as other yet-to-be-developed measures, will be critical for making assessment effective both for educational purposes—to ensure that teachers and students can monitor and improve the learning process—and for accountability purposes—to ensure that schools are giving all students what they need to succeed. This will require a larger investment in the development and design of assessments and assessment systems. It will also mean more coordination between policymakers, educators, researchers, and test developers, who too often work in isolation of one another.

Endnotes

[1] In-person student interviews at St. Andrew's School, March 2008.

[2] SCANS reports, available online at http://wdr.doleta.gov/SCANS/whatwork/.

[3] The New Commission on the Skills of the American Workforce, *Tough Choices or Tough Times* (Washington, DC: National Center on Education and the Economy, 2006).

[4] Richard Murnane and Frank Levy, *The New Division of Labor: How Computers Are Creating the Next Job Market* (Princeton, NJ: Princeton University Press, 2004).

[5] Trends in International Mathematics and Science Study (TIMSS) 2003; The Programme for International Student Assessment (PISA) 2003.

[6] Ed Coughlin, Cheryl Lemke, et al. *enGauge 21st Century Skills: Literacy in the Digital Age* (Los Angeles, CA: The Metiri Group, 2003).

[7] *Framework for 21st Century Learning* (Tucson, AZ: The Partnership for 21st Century Skills, July 23, 2007).

[8] *Definition and Selection of Key Competencies* (Paris, France: OECD, May 27, 2005).

[9] J.D. Bransford, A.L. Brown, and R.R. Cocking (eds), *How People Learn*, 2000; OECD, *Innovation in the Knowledge Economy: Implications for Education and Learning*, 2004; and R. Kozma, *Technology, Innovation, and Educational Change: A Global Perspective* (Eugene, OR: International Society for Technology in Education, 2003).

[10] Benjamin S. Bloom and David R. Krathwohl, *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain* (New York: Longman, Green, 1956). Also D. Krathwohl, B.S. Bloom, and B.B. Masia, *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain* (New York: David McKay Co., Inc., 1964).

[11] L.W. Anderson, D. R. Krathwohl, eds, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (New York: Allyn-Bacon Longman, 2001).

[12] L. B. Resnick, "Making America Smarter," *Education Week Century Series* 18(40) (June 16, 1999), 38–40.

[13] See diagram, Figure 2.

[14] L. W. Anderson, D. R. Krathwohl, eds, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*.

[15] Interview with John Austin, March 2008.

[16] D. J. Ketelhut, "The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment," *The Journal of Science Education and Technology* 16 no. 1 (2007): 99–111. For more information about River City, also see http://muve.gse.harvard.edu/rivercityproject/.

[17] Interview with Eva Baker, May 2008.

[18] Eva Baker, *Moving to the Next Generation System Design: Integrating Cognition, Assessment, and Learning* (Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, February, 2007).

[19] Key Stage 3 ICT Assessment Tasks, National Assessment Agency, United Kingdom, available online at http://www.naa.org.uk/naa_18908.aspx.

[20] International Baccalaureate Organization. "Diploma Programme Assessment Principles and Practice." (United Kingdom: Antony Rowe Ltd., September 2004).

[21] ETS uses the term "non-cognitive" because it is widely used in psychology and measurement. It is, however, a misnomer for describing these skills, as many are actually cognitive in nature.

[22] Interview with Richard Roberts, March 2008.

[23] "Title I: Characteristics of tests will influence expenses; information sharing may help states realize efficiencies." Report to Congressional requesters. (Washington, DC: Government Accounting Office, May, 2003).

[24] Thomas Toch, *Margins of Error: The Testing Industry in the No Child Left Behind Era* (Washington, DC: Education Sector, 2006).

[25] Linda Jacobson, "Kentucky Lawmakers Take Aim at State Tests," *Education Week*, March, 2008.

[26] A bill was submitted during the Spring 2008 session that called for the replacement of the KY CATS system with a norm-referenced test and an augmented test. It did not move out of committee. According to the state's office of accountability and assessment, the state has established a task force on assessment to consider future changes to state tests.

[27] "The Technical Advisory Committee on Faking on NonCognitive Assessments," October 13, 2006. Sponsored by the Educational Testing Service.

[28] The Proposed Delaware Comprehensive Assessment System developed by the Delaware Statewide Academic Growth Assessment Pilot, Spring 2008, available online at http://www.mapuser.k12.de.us/future/index.html#delaware.