Reform at a Crossroads

A Call for Balanced Systems of Assessment and Accountability

NASBE Study Group on Assessment Systems for the 21st Century Learner
Reform at a Crossroads: A Call for Balanced Systems of Assessment and Accountability

The Report of the NASBE Study Group on Assessment Systems for the 21st Century Learner
**Study Group on Assessment Systems for the 21st Century Learner**

**State Board Members***

Kerry Winterer (Chair) – Nebraska  
Dixie Allen – Utah  
Melissa Bartlett – North Carolina  
Sharon Carroll – Montana  
Margaret Cox – Hawaii  
Nancy Danhof – Michigan  
James C. Dawson – New York  
Blair Ewing – Maryland  
Anita Mendoza – Arizona  
Mary Ann Miller – Kentucky  
Jenny Phillips – West Virginia  
Rita Sablan – Northern Marianas  
Laura Slover – District of Columbia  
Diane Tatum – Arkansas  

* All were members of their state board of education at the time of the Study Group meetings, except Mary Ann Miller, who is Policy Advisor at the Kentucky Department of Education.

**Ex Officio Members**

Jan Barth – West Virginia  
Frances Evans – California  
Stuart Kahl – Measured Progress, Inc.  
Beth Miller – Nellie Mae Education Foundation  
Sherri Miller – ACT, Inc.  
Paul Nichols – Pearson Foundation  
Charlie Toulmin – Nellie Mae Education Foundation

**NASBE Staff**

Mariana Haynes, Project Director and lead author  
Jessie Levin, Project Associate  
David Kysilko, Editor

**Presenters to the Study Group**

**January**

Eva Baker, National Center for Research on Evaluation, Standards, and Student Testing  
Stuart Kahl, Measured Progress, Inc.  
Alissa Peltzman, Achieve, Inc.  
John Tanner, Council of Chief State School Officers

**March**

David Driscoll, National Assessment Governing Board

Laura Hamilton, RAND Corporation  
Doris Redfield, Edvantia, Inc.  
Stacy Scott, Associate Superintendent, Office of Shared Accountability, Montgomery County Public Schools

**June**

Sandra Boyd, Achieve, Inc.  
Wayne Camara, College Board  
Pasquale DeVito, Measured Progress, Inc., formerly National Research Council  
Stuart Kahl, Measured Progress, Inc.  
Sherri Miller, ACT, Inc.  
Chris Minnich, Council of Chief State School Officers  
Zollie Stevenson, Jr., U.S. Department of Education  
Charlie Toulmin, Nellie Mae Education Foundation  
Bi Vuong, Data Quality Campaign

Copyright © 2009  
National Association of State Boards of Education  
All rights reserved.

Additional copies of this report are available for $16 each plus 10% for shipping and handling ($4.50 minimum) from NASBE at 2121 Crystal Drive, Suite 350, Arlington, VA 22202. All orders under $50 must be prepaid. Call 800-220-5183 for more information or order online at [www.nasbe.org](http://www.nasbe.org). Volume discounts available.

For more information about NASBE’s work in assessment and accountability systems, contact Dr. Mariana Haynes at 703-684-4000 or marianah@nasbe.org.
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>I. The Need to Overhaul State Assessment and Accountability Systems</td>
<td>6</td>
</tr>
<tr>
<td>II. Linking Assessments with Teaching and Learning</td>
<td>19</td>
</tr>
<tr>
<td>III. Creating a Balanced Assessment System for the 21st Century</td>
<td>28</td>
</tr>
<tr>
<td>IV. Assessment for Learning</td>
<td>36</td>
</tr>
<tr>
<td>V. Conclusion</td>
<td>44</td>
</tr>
<tr>
<td>VI. Study Group Recommendations</td>
<td>45</td>
</tr>
<tr>
<td>Appendix A. Glossary of Assessment Terms</td>
<td>48</td>
</tr>
<tr>
<td>Appendix B. State Actions in Assessment and Accountability</td>
<td>52</td>
</tr>
<tr>
<td>Appendix C. Bibliography</td>
<td>56</td>
</tr>
<tr>
<td>Endnotes</td>
<td>58</td>
</tr>
</tbody>
</table>

NASBE would like to thank the Pearson Foundation, the Nellie Mae Education Foundation, and ACT, Inc. for their support of the activities and report of the Study Group on Assessment Systems for the 21st Century Learner. However, recommendations and substance of this report reflect the collective thinking of the Study Group members and do not necessarily represent the views of Pearson, Nellie Mae Education, or ACT.
Profound changes in the nation’s economy, demography, and technological landscape have prompted states to rethink what competencies students must have to succeed in the 21st century. In response to the increased expectations, states are taking the lead to reexamine the foundational elements of state education systems—the standards and assessments that are used to drive and measure attainment of essential knowledge and skills. Research shows that if 12th graders are to be successful in their next tier of study—whether in two- or four-year colleges or in technical or career coursework—they must be able to use sophisticated and specialized skills. Young adults need to be able to read and evaluate subject matter texts; write expository, descriptive, and persuasive text; and apply mathematics constructs to solve complex problems in new contexts. Technology will continue to have considerable impact on what and how people learn as well. High school graduates will need to use information technologies to apply knowledge and skills that are difficult to cultivate and measure through traditional classroom instruction and multiple-choice exams.

From a historical context, major school reform has rested with the states. They were at the forefront of standards-based education during the 1980s and ‘90s and are again taking action to ensure American students receive a world class education. The growing sense of urgency to transform public education comes not only from rising expectations, but in recognition that despite decades of reform efforts, profound inequities continue to exist in the social fabric of the United States. High level communication, technological, and cognitive skills are unevenly distributed across groups based on race/ethnicity, geography, and socioeconomic status. Large numbers of students do not graduate from high school on time or at all, and those who do find their diplomas do not necessarily signify their readiness for entry level jobs or readiness to enter and succeed in college. Studies show that many graduates require remediation their freshmen year and far too few go on to earn college degrees. Increasingly, entry level positions in today’s
A Call for Balanced Systems of Assessment and Accountability

National Association of State Boards of Education

5

economy call for at least some form of advanced training. In response, policy leaders and education stakeholders have recognized the need to challenge fundamental assumptions about how we educate and prepare students to meet the demands of employment, advanced training, and civic participation.

These demands fall hard on teachers and students. Growing concerns about how to significantly improve the quality of education for all students has generated healthy debate about the policies that shape education in states. With much at stake in a downturned economy, there has been greater scrutiny of our basic systems for evaluating students and holding schools accountable. Indeed, many critics of public education contend that these systems are inherently flawed and are major impediments to improving student achievement. Our assessment systems, critics charge, are out of step with policies adopted in the highest performing and most equitable nations, reflecting earlier theories of learning and measurement that impede progress in how we scale and sustain improvements in teaching and learning. As a result, a growing majority of testing experts and analysts now believes that education cannot be transformed under the constraints of current state assessment and accountability systems.

But this does not mean that the role of assessments will or should be diminished. The good news is that new learning and measurement theories have been developed that promise to reframe the foundations of assessments systems that can better support learning. These advances in science as to how people learn and how to measure learning and performance offer promising approaches to assessment that may provide clearer information about the progress of student learning that is more consistent with 21st century expectations.

State boards of education recognize the urgency to retool schooling in response to extraordinary economic, societal, and technological changes. The movement to strengthen the nation’s workforce by redefining what all students need to learn has profound implications for the design of state assessment systems. The confluence of events—the change in the national and international climate, acknowledgement of long-standing inequities in education opportunity and achievement, and the recognition of constraints inherent in current assessment and accountability systems—prompted NASBE to convene a year-long Study Group to examine assessment systems for the 21st century learner. The Study Group considered a number of key questions as to how to design assessment systems that would capture complex high-level skills and deep content knowledge while providing meaningful information about students’ progress and their specific learning and performance gaps.

These questions include:

- How do we create assessment and accountability systems that reflect what we really want students to know and be able to do and that incorporate the latest research about teaching and learning?
- How can we rethink the core standards in terms of how they contribute to students’ abilities to be successful?
- What can we learn from high-performing nations about the design of coherent assessment systems?
- How should we redesign our assessment systems and what are the implications of these changes for each level of the system (state, district, school, and classroom)?
- What are the implications of these new assessments for helping teachers develop highly effective instructional practices (which involve calibration and differentiation of teaching in response to the level of individual students’ learning)?
- What are the implications for higher education teacher and administrator development programs?
- What are the accountability issues that must be addressed as the purposes and expectations for PK-16 systems evolve?
Students and their teachers face extraordinary challenges to meet increased expectations for today’s high school graduates. As dramatic social, economic, and technological changes unfold, the American public has deep concerns regarding education in this country and our ability to prepare young people for college and careers in the 21st century. The U.S. economy is moving from manufacturing to services with many more jobs requiring specialized skills. In fact, jobs requiring limited skills—the blue-collar and clerical jobs that supported the middle class for most of the 20th century—are rapidly diminishing. In contrast, it is estimated that the number of jobs requiring some college-level education now accounts for 46 percent of job growth and that individuals with higher level cognitive skills will accrue far greater earnings and opportunities than their counterparts with only a high school diploma.

The Study Group identified the need to examine how states design accountability and assessment systems as central to achieving educational goals that have shifted in light of dramatic changes in the nation’s demography and economy. Fundamentally, the K-12 system only works for the 30 percent who graduate prepared for the future. This leaves behind the 40 percent who graduate, but do so without the knowledge and skills needed to succeed in college, career training, or entry-level jobs, as well as the 30 percent who drop out before high school graduation. While these percentages served the nation when it had a strong industrial and manufacturing base, they are not going to serve us well after we have been hit by what’s been called the “perfect storm”—the confluence of changing demographic and global economic trends that could derail the nation from providing real opportunity for all its citizens.

I. The Need to Overhaul State Assessment and Accountability Systems

A. U.S. Schools Face Growing Global Competition

In part, the press to internationally benchmark the performance of U.S. students with other countries results from two factors—1) the acknowledgement of the current technological, economic, and political trends that have combined to increase demand for higher skills and 2) the growing concern regarding the downward shift in the U.S. ranking on a number of important indicators in comparison with other advanced nations. For example, on recent Programme for International Student Assessment (PISA) tests, U.S. students ranked 25th in math, 21st in science, 15th in reading, and 24th in problem solving in a comparison of students from 30 industrialized countries.

Other research has found that the level of mathematics taught in the
United States is significantly lower than in Asian countries such as Chinese Taipei, South Korea, Singapore, Hong Kong SAR, and Japan, and while America had the best high school graduation rate in the world in the 1960s, by 2006 it had declined to 18th out of 24 industrialized nations.

Concern about these troubling numbers is not idle handwringing that can be equated to popular anxieties over which countries are racking up more gold medals at the Olympics. “[This] is another wake-up call that our students are treading the waters of academic achievement while other countries’ students are swimming faster and farther,” Secretary of Education Arne Duncan said of the international data released in August 2009. “Our students have stagnated educationally, putting our long-term economic security at risk.”

There is no doubt the United States faces stiff competition for jobs from skilled workers in other countries, as globalization has rapidly transformed industries throughout the world such as telecommunications, computer networking, aerospace, and biotechnology. States recognize that given the new economic realities, they must take steps to ensure that students receive a world-class education that positions them to succeed in the knowledge-fueled, global economy—and this is where the international assessments have more to tell leaders in the United States.

While researchers urge policymakers to be cautious about making too much of how any one nation’s education reforms might be affecting its test results, a look at the responses many high-performing countries have had to these international scores provides numerous illustrations of possible changes in policies and practices. For example, Schmidt reports that high-scoring countries such as Finland, Germany, Singapore, and Australia all shifted gears in the 1990s in reaction to their poor performance on international tests. These countries were experiencing large percentages of students failing basic tests, huge differences between states/regions, and achievement gaps related to students’ socio-economic status. In response, the countries created national standards, adopted school-based performance measures, focused on strengthening teacher development, and afforded schools greater flexibility in adopting a variety of instructional approaches.

Comparative studies of top performing nations show that they institute rigorous, coherent standards with an array of closely aligned instructional and assessment tools and commit substantial resources to developing highly effective teachers. Finland, which performs at high levels on the Trends in International Mathematics and Science Study (TIMSS) and PISA, adopted national standards and made substantial investments and improvements in teacher quality, including considerable strides to identify what’s essential by specifying curricula and rigorously training teachers on pedagogy and content. Hammond notes that the highest achieving nations focus on using in-school performance assessments that serve a number of purposes. Their systems include rich assessment tasks at the classroom level that are used to help teachers gauge students’ progress and are then combined with centralized assessments to evaluate school performance. Teachers in these countries play a more central role in developing and scoring assessments that serve as an important bridge between educational goals and classroom expectations.

State policymakers must adopt a global lens in reworking the assessment and accountability systems to revitalize and transform education in America. In the words of Andreas Schleicher, head of the Indicators and Analysis Division at the Organisation for Economic Co-operation and Development (OECD)’s Directorate for Education, “It is only through such benchmarking that countries can understand relative strengths and weaknesses of their education system and identify best practices and ways forward. The world is indifferent to tradition and past reputations, unforgiving of frailty and ignorant of custom or practice. Success will go to those individuals and countries which are swift to adapt, slow to complain, and open to change.”
B. The Persistence of Inequities Across All Levels of the System

The press to ensure equitable opportunities for all students will require a collective solution to not only raise the level of preparedness for all students, but to also address the huge achievement gaps that exist across all levels of the system. What’s disturbing is that national surveys continue to show that a large portion of our nation’s adults have not attained the literacy and numeracy skills needed to participate in today’s workplace. In the area of literacy, for example, data from the National Assessment of Education Progress (NAEP) reveal that for the past two decades reading scores for 13- and 17-year-olds have flat-lined. Besides low skill attainment throughout the United States, the skills needed to function effectively in a complex, knowledge-based society are unevenly distributed across groups defined by race/ethnicity, residence, and socioeconomic status. 20

These disparities show up in all measures of educational attainment:

- Although 70 percent of students graduate, many lack the knowledge and skills to succeed beyond high school; a majority of nation’s 12th graders read and do math below grade level;
- Only about 55 percent of black students and 58 percent of Hispanic students graduate on time with a regular diploma;
- To participate in college classes, 42 percent of public community college freshman and 20 percent of freshman in public four-year institutions need to first take remedial courses in basic skills such as reading, writing, and math;
- Only 40 percent of 9th graders enroll in college four years later; and
- Despite best intentions and massive investments by states in their design of standards-based accountability systems, our efforts have not succeeded in closing achievement gaps linked to students’ race/ethnicity, income, and place of residence. 22

A number of disturbing trends within American’s education system have come to the fore regarding the inequities in the level of support, instructional quality, and learning opportunities provided to students. Learning gaps related to socioeconomic status start early, but then widen throughout formal schooling. The effect of poverty on achievement doubles once students enter schools because schools in high-poverty areas typically deliver less demanding curricula. 23 Black, Hispanic, and Native American youth continue to lag far behind Asians and whites in both the quantity and rigor of courses taken and in levels of achievement. 24 Even more disconcerting, in contrast to high-performing countries, disadvantaged and low-achieving students get a watered-down curriculum, larger classes, and less qualified teachers, while advanced students have access to a high-quality, national curriculum with rigorous coursework and examinations as part of Advanced Placement programs.

A report issued in 2009 by McKinsey and Company highlights the serious economic consequences of the nation’s inability to address these enormous disparities, noting that “the persistence of these educational achievement gaps imposes on the United States the economic equivalent of a permanent national recession.” 25

What’s even more troubling is that the achievement gap between students from rich and poor families is much more pronounced in the United States than in other high-performing nations around the world. 26 Studies measuring the impact of family background on international assessments found that the United States ranks in the top quarter of the most unequal countries based on the performance gaps for students from different family backgrounds. For example, American 15-year-olds rank in the bottom third of OECD countries on overall achievement in math and science, and they rank in the top-third when it comes to gaps between students from different family backgrounds. 27 In science, the difference in scores on PISA between students from different socioeconomic backgrounds is greater in the United States than in almost any other country. 28 In other words, the United States fares poorly on a key indicator of equal opportunity in a society: the degree to which economic status predicts student achievement. By every measure of
educational achievement, poor and minority students in this country fare worse than their other American and international peers.

Other countries around the world such as Japan, Korea, Finland, and Canada do a much better job of leveling the educational opportunities for students from lower-income families. In responding to the question of what’s more important for a nation’s sustained growth—producing a group of high performers or bringing everyone up to a basic level of performance—economists have concluded that increasing the average level of achievement will have a greater effect on growth than increasing the percentage of individuals who achieve at high levels. There is now broad agreement that if we are unable to close the existing skills gaps among different demographic groups and substantially boost achievement levels overall, we will not be able to meet the standards of a world-class education system.

These gaps have serious economic and moral consequences for individuals and society at large, especially considering the U.S. population is becoming increasingly diverse. Reducing the inequality in education is as important to our well-being as a society as it is to retaining our competitive edge. States need to design assessment systems that permit comparisons across education systems in order to calibrate good versus bad schooling, to gauge acceptable levels of student performance, and to cull important lessons about the policies and practices that leverage systemic improvements. Even though states are deeply engaged in efforts to raise standards, improve teaching effectiveness, and improve low-performing schools, new solutions must be found to develop systems that can inform and drive improvements at all levels. Of increasing concern are the enormous variations in quality and outcomes that persist across all levels of the system—classroom, schools, districts, states, and nations around the world—and the inherent problem of scaling effective instructional practice across all levels of the system.

It is urgent and essential for states to achieve two goals: first, improving rigor and the relevance of standards to what’s needed in the 21st century and second, ensuring comparability across states in order to reduce disparities in what students learn. In summary, the Study Group concurred that not only do we need better standards and assessments, but we need to retool systems to bring substantial improvements in curriculum, instruction, and the effectiveness of teachers in diagnosing and tailoring improvements to meet students’ needs.

C. The Need for “Fewer, Clearer, Higher” Standards

Over the past several decades, standards-setting has served as the centerpiece of state education improvement efforts as states used a range of processes to define what students should know and be able to do in core subjects. Policymakers also anticipated that the standards would not only set what students needed to learn, they would improve how teachers taught.

On the positive side, many of the standards developed by states and professional groups such as the National Council of Teachers of Mathematics and the science standards developed by the National Research Council emphasize that students not only need to attain a deep understanding of the content of various subjects but also to develop the sophisticated thinking skills necessary to perform competently across these different domains. For example, the best math and science standards clearly define the content to be learned and emphasize the need for students to be actively engaged in learning in order to build coherent structures of knowledge and apply that knowledge.

On the negative side, as the purpose of assessment has increasingly shifted toward using tests for holding students and educators accountable for learning outcomes, the resulting patchwork of state content standards has raised fundamental questions about their quality, the
type of tests used to assess their attainment, and their utility in driving instructional improvements in teaching and learning. For the most part, standards have been derived using methods whereby a panel of experts arrives at a consensus as to what an “educated” student needs to know and be able to do, with the result vetted by an array of constituencies before finally being adopted by the state board of education. This process—replicated 50 times for each subject—has generated sets of standards with little vertical or horizontal articulation resulting in many incongruities and discontinuities. Most state standards are voluminous; they are too broad and vague, requiring students to learn a wide range of content without sufficient depth. Others are so specific that the big idea is lost or buried.34

Moreover, while there have been pockets of improvement under the No Child Left Behind Act (NCLB), in general the federal law has not been able to reduce the wide variations in educational quality found in the United States. Critics contend that NCLB allows too much flexibility in setting standards and is far too prescriptive in how to intervene when underperformance occurs. Concerns have surfaced about what’s been included in states’ content standards, how they are written, and how acceptable performance levels (e.g., proficient, basic, etc.) are defined. Furthermore, the variability in curriculum and testing has made it hard for states to know how students across the nation compare to one another, much less how well they are keeping pace with other industrialized nations.

Laura Hamilton from RAND addressed the Study Group about research on the impact of standards-based reform and cited a number of concerns with the design of standards and large-scale assessments. The theory of action behind standards-based reform posits that by: 1) measuring students’ attainment of standards; 2) holding students, teachers, and administrators accountable for improving student achievement; and 3) applying consequences and incentives on the basis of test results, the quality of education for all students will improve over time. Throughout the past two decades, states have applied high-stakes decisions to test results, including retention and promotion determination, high school graduation, teacher and administrator evaluation, how resources are allocated, and whether schools are rewarded or sanctioned.35

She noted that the mismatch between standards and assessments poses a serious challenge to broadly improving student learning and performance. The original idea behind standards-based reform was that standards would guide instruction and then we would measure progress toward meeting standards. The test was supposed to serve as a neutral instrument. Unwittingly, standards-based accountability has become “test-based” accountability.36 In fact, as Hamilton reported to the Study Group, this has been a major disappointment in standards-based reform. Standards that emphasize complex reasoning and problem-solving skills are often ignored in the classroom because they are not measured by most state accountability tests, while increased instructional time has been devoted to those skills and topics that are included in the tests. In addition,

“Learning science is something that students do, not something that is done to them. In learning science, students describe objects and events, ask questions, organize knowledge, construct explanations of natural phenomena, test those explanations in many different ways, and communicate ideas to others…Students establish connections between knowledge of science and the scientific knowledge found in many sources; they apply science content to new questions; they engage in problem solving, planning, and group discussions; and they experience assessments that are consistent with an active approach to learning.”

standards-based reform does not appear to have transformed instruction or improved instructional coherence in the fundamental ways envisioned by some early reform advocates, as demonstrated by the finding that variability in teaching is enormous even within the same school.

Part of the problem with our standards, William Schmidt noted in an interview, is the sheer number of topics they typically cover. “Standards need to focus on a small enough number of topics so that teachers can spend months, not days, on them,” he said. “[In the early grades, in teaching math] top-achieving countries usually cover about four to six topics related to basic numeracy, measurement, and arithmetic operations. That’s all. In contrast, in the U.S., state and district standards, as well as textbooks, often cram 20 topics into the first and second grades. That’s much more than any child could possibly absorb.”

In addition, the large number of topics that must be covered within one subject for one grade level restricts specifying what constitutes quality instruction, consequently limiting teacher development and diminishing the capacity of states, districts, and schools to bring to scale what works in improving teaching and learning. How should we define quality third-grade math instruction, for example, when third grade math is so broadly defined?

Finally, the high rate of student mobility poses an enormous challenge to standardizing students’ opportunity to learn, considering that individual teachers and their delivery of curricula trumps all else.

The Study Group concluded that states should:

1. Create fewer, clearer, and higher standards (common standards) to meet expectations for entry into college and the workplace, to set standards against international benchmarks, and to increase standardization and comparability across states.

2. Ensure that decisions are based on up-to-date valid, reliable research to identify the knowledge and skills that predict college and career readiness and incorporate advances in the sciences of learning and thinking (e.g., developmental pathways, learning progressions).

3. Focus on the implementation of standards by leveraging systemic improvements and harnessing the research community as critical partners in determining what works to improve teaching and learning so that all students have the opportunity to attain common goals.

D. The Need for 21st Century Expectations

Recognizing the shift in the nation’s economy and demography, states have not remained idle in exercising their authority to begin benchmarking their education systems at levels commensurate with the best in the world. The first step is recalibrating expectations and increasing the rigor of standards and curricula content. States have initiated a number of joint efforts to begin crafting a system that prepares students for college and careers, including the American Diploma Project (see text box on page 12).

Other states have adopted a range of approaches to defining benchmarks and assessing student progress towards attaining them. Ten states administer tests to high school students to assess college and career readiness (e.g., setting proficiency scores on state exam to align with university placement exams, adopting a test that will be used by postsecondary system for placement; or requiring students to take a national college admissions exam). About half the states require students to take comprehensive high school exams, most of which are based on standards for 10th grade or higher. As states redouble their efforts to define readiness levels in relation to what’s needed at college or job entry, more states are phasing out the comprehensive high school exams (currently administered in 24 states) in favor of end-of-course exams that can serve a number of purposes—to align coursework, measure student mastery of course content, and signal the need to intervene to close gaps in students’ performance and preparation for postsecondary training and jobs.

What we know is that coursework alone is not predictive of college entrance or success and that
much closer monitoring of what each individual student is learning is needed to ensure graduation equates with preparedness for jobs and advanced training. ACT has conducted extensive research to link mastery of specific subject matter content to success after high school. The findings show that only 1 in 10 students at the eighth grade level and 1 in 5 students at the 10th grade level are on target for entry-level college courses in English composition, algebra, social science, and biology. Furthermore, these studies show that increasing graduation requirements in the form of academic coursework is necessary but insufficient to ensure a rigorous curriculum. Subject area classes often lack adequate depth and rigor; their grading is not aligned to challenging expectations, and classes are not targeted toward key student outcomes that employers and professors expect of high school graduates.

One example of states developing more rigorous, world-benchmarked standards is the American Diploma Project (ADP), launched by Achieve, Inc. at the National Education Summit on High Schools in 2005. Research conducted during the initial phase of ADP produced college- and career-ready benchmarks in English and mathematics focused on remedying two major deficiencies in graduates’ preparation identified by college faculty and employers—one, the overemphasis on narrow foundation skills in middle school and two, the lack of complex and conceptual competencies acquired late in high school that requires several years to develop.

The ADP benchmarks integrate a strong content foundation with cross-disciplinary proficiencies including research and evidence gathering, critical thinking and decision making, communication and teamwork, and media and technology. In 2008, Achieve reported that many of the 33 states in the network share a common core—a subset of the benchmarks that represents essential expectations for college- and career-readiness. ADP states committed to working together to address difficult policy issues directed to increasing the rigor of education by working on:

- Aligning high school academic content standards in English and mathematics with the demands of college and careers;
- Requiring students to complete a college- and career-ready curriculum so that earning a diploma ensures that a student is ready for postsecondary opportunities;
- Administering statewide high school assessments anchored to college- and career-ready expectations; and
- Creating comprehensive accountability and reporting systems that promote college and career readiness for all students.

By early 2009, as a number of new states joined the network, Achieve identified 23 states that have revised their high school academic standards in English and/or mathematics to align them with college and career expectations; another 21 states and the District of Columbia reported that they are in the process of or have plans to similarly align their English and mathematics standards. States have also begun to redesign graduation requirements: 14 states require students to automatically enroll in the “default” college- and career-ready curriculum but allow them to opt out of the requirements if their parents sign a waiver. Another six states and the District of Columbia have set mandatory course requirements without opt-out provisions.

Source: Achieve, Inc.42

American Diploma Project

One example of states developing more rigorous, world-benchmarked standards is the American Diploma Project (ADP), launched by Achieve, Inc. at the National Education Summit on High Schools in 2005. Research conducted during the initial phase of ADP produced college- and career-ready benchmarks in English and mathematics focused on remedying two major deficiencies in graduates’ preparation identified by college faculty and employers—one, the overemphasis on narrow foundation skills in middle school and two, the lack of complex and conceptual competencies acquired late in high school that requires several years to develop.

The ADP benchmarks integrate a strong content foundation with cross-disciplinary proficiencies including research and evidence gathering, critical thinking and decision making, communication and teamwork, and media and technology. In 2008, Achieve reported that many of the 33 states in the network share a common core—a subset of the benchmarks that represents essential expectations for college- and career-readiness. ADP states committed to working together to address difficult policy issues directed to increasing the rigor of education by working on:

- Aligning high school academic content standards in English and mathematics with the demands of college and careers;
- Requiring students to complete a college- and career-ready curriculum so that earning a diploma ensures that a student is ready for postsecondary opportunities;
- Administering statewide high school assessments anchored to college- and career-ready expectations; and
- Creating comprehensive accountability and reporting systems that promote college and career readiness for all students.

By early 2009, as a number of new states joined the network, Achieve identified 23 states that have revised their high school academic standards in English and/or mathematics to align them with college and career expectations; another 21 states and the District of Columbia reported that they are in the process of or have plans to similarly align their English and mathematics standards. States have also begun to redesign graduation requirements: 14 states require students to automatically enroll in the “default” college- and career-ready curriculum but allow them to opt out of the requirements if their parents sign a waiver. Another six states and the District of Columbia have set mandatory course requirements without opt-out provisions.

Source: Achieve, Inc.42
To bridge the disconnect between students’ achievement level and what they need to succeed following high school, states must attend to how to improve the academic quality and intensity of curricula. Policy leaders have looked for ways within the current framework to increase the responsiveness to the individual student while at the same time improving the rigor of course content. States have taken action to focus on clearer and higher standards; define essential course outcomes, and effectively prepare and support teachers in integrating a variety of approaches to measure students’ progress and respond to their individual learning needs.43

But the speed of change and the challenge of framing a coherent response have made the process difficult. States have recognized that despite considerable investments—fiscal and human—in the design of standards and assessments, the enterprise is complex and extremely costly. To meet rising expectations, address the limitations of current assessments and accountability systems, and maximize efficiencies, states are opting to no longer go it alone.44 In the last several years, states have made strides through cross-state collaboratives to develop common assessments in such efforts as the Algebra II Assessment Consortium and the New England Common Assessment Program (NECAP). (See text box on NECAP on page 14). Perhaps most dramatically, in the summer of 2009, 48 states and three territories joined in a state-led initiative to create a more meaningful, 21st-century-focused set of common standards in English language arts and mathematics—an effort that has already imprinted the phrase “fewer, clearer, higher” standards on the policy language of education leaders across the country. (See text box on common core standards, page 16.)

As described in From No Child Left Behind to Every Child a Graduate, “We must fundamentally change the way we educate students as a whole by summoning our resources and expertise to solving these complex and challenging issues, or we will experience further educational decline and a lower standard of living.”45

E. The Limitations of Current Assessment Systems

In addition to addressing the need for fewer, clearer, and higher standards, significant problems have arisen with how current standardized assessments have been designed and applied. State standards, the tests that measure them, and the accountability systems built upon them have had a profound impact on the nation’s education system, increasing the amount of testing in K-12 schools—and the stakes that go with them—particularly under NCLB. But while we have depended heavily on using educational assessment for high-stakes individual and programmatic decisions, it is largely recognized that current assessment and accountability systems will not achieve our fundamental aims of improving the quality of teaching and learning for all students.

John Tanner, Director of Innovative Measures at the Council of Chief State Schools Officers, pinpointed for the Study Group a number of constraints associated with assessment systems that are used largely for accountability purposes. First, he said, large-scale assessments tend to measure a narrow range of knowledge and skills drawn from the lower end of the standards through the use of multiple-choice formats. Skills that represent higher levels of cognitive demand are difficult to measure with multiple-choice tests. As a consequence, teachers and administrators shape curriculum and instruction to target a more narrow range of content skills.

Second, state tests adhere to older testing and learning theories that characterize achievement as reflecting individual student’s innate abilities that remain relatively stable over time and context. Third, state tests tend to measure discrete facts in a few core content areas rather than capturing the complex knowledge and skills considered requisite for the 21st century learner. Finally, current assessments are limited in improving teaching and learning for all students—which has been the major goal of education reform. Accountability tests provide a single snapshot of achievement based on how well the student performs relative to his/her peers or in terms of a cut score that assigns the individual to a performance band (e.g., “below basic” or “above basic” in reading). These measures offer little diagnostic information as to what specific instruction or supports students need relative to their level of
Collaboration across state lines in standards, assessments, and other areas is one strategy policymakers are approaching with less hesitancy than in the past. However, such efforts are not happening for the first time in today’s context. In response to NCLB, in 2002 New Hampshire, Vermont, Rhode Island, and later Maine, leveraged resources and political will to develop the New England Common Assessment Program (NECAP). The state’s commissioners and deputy commissioners of education acknowledged that the tests then current in each of their states would not independently meet NCLB’s requirements for testing in grades 3-8 and still maintain the desired high level of quality. Joining forces produced efficiencies and an opportunity to create better quality tests for less money. The current administration is prioritizing the efficiencies of joint state efforts in assessment by explicitly targeting $350 million in its Race to the Top Fund for the development of new common assessments.

Additionally, given the regional similarities and shared investment in graduates who often shuffle among New England states, this process also provided a venue to establish collective “must haves” for future members of the 21st century workforce. Measured Progress, Inc. was contracted as the consortium’s assessment developer and the National Center for the Improvement of Educational Assessment, Inc. as the facilitator. As a first step and to develop capacity for cross-state comparisons for joint assessments, the states had to reconcile their standards. Common Grade-Level Expectations (GLEs) were determined for mathematics, reading, and writing first for grades 3-8. Later, targets were set for science assessments, and Grade-Span Expectations (GSEs) were created for high school students in mathematics, reading, and writing. The states integrated GLEs and GSEs into curriculum frameworks that were subsequently approved by their state boards.

While NECAP is often regarded as the best example of an assessment/accountability consortium to date, this work was neither clean nor easy. Based on his experiences as NECAP’s assessment developer, Measured Progress CEO Stuart Kahl urges states to consider the following issues before building or joining a consortium around testing:

- **Timing:** The work is not a one-time commitment. States that are not scheduled to revise standards may have to make significant changes to their workscope calendars to begin such work. Once engaged in a collaborative, before assessment development can start, states have to agree on testing windows and shared standards. For example, Maine’s commitment to NECAP was delayed until the state passed legislation necessary to adopt the fall testing schedule followed by the rest of the collaborative.

- **Shared Values and Proximity:** The NECAP states share similar budgetary, resource, and capacity obstacles that helped draw them into initial collaboration. Moreover, they hold similar priorities for assessment: they all value constructed responses, challenging tests, and high standards enforced by rigorous cut scores. Regional proximity also allowed NECAP members to readily conduct face-to-face meetings. On the development end, the common geography also allowed for the use of “regionally flavored” content (e.g., reading passages on cold weather as opposed to hurricanes), which can boost student engagement and motivation during testing.

- **Strong Leadership:** The nature of compromise in consortia is both tricky and critical. In order for the people who will be implementing an assessment system to buy into a new strategy, strong commitment to a unilateral approach must be solidified at the top. Some degree of coercion can keep members from expending too much energy on small differences that could lead to sacrificing quality to a least common denominator of agreement. Due in part to its strong leadership, this effort has sustained long-term institutional value across the four states, which speaks volumes to the issues of capacity and continuity that often pose substantial barriers.

- **External Management:** Contracting with a program manager outside of the states and assessment developer also helps build a successful consortium. An independent
facilitator offers logistical support in terms of planning and chairing meetings, as well as serving as a mediator and ensuring that one vision is conveyed to the test contractor.

- **Identical Standards and Tests**: Efficiencies are not assured—if states do not accept identical standards and the same tests, the payoffs in production costs and scoring comparability are lost.

- **Cost Sharing Formulas**: Establishing how states of differing sizes will pay for the various expenditures associated with building assessments is an important step in the planning process. A hybrid formula allows states to equally share fixed costs such as those for development while dividing variable costs such as shipping and scoring according to number of students in each state.

- **Standards and Assessments Will Not Do It All**: While the benefits can be significant, there are two cautions states should acknowledge. First, common standards do not allow for comparability across states unless the same test is used. Second, the disparity between student performance in the United States and Singapore, which uses U.S.-created standards, is a testament to the reality that standards do not guarantee improvements. Performance growth happens at the classroom level. 

  
  
  Competence and conceptual development within a particular subject area.51

  Furthermore, the RAND analyses of standards-based assessment and accountability systems showed that, overall, tests rather than standards tend to drive instruction: schools and teachers spend more time and resources on tested subjects and on content included in the test. 52 Because many standards are overly broad or poorly written, while at the same time tests often focus on easy-to-measure items with lower cognitive demands, there is poor alignment between standards and assessments. As a result, the focus on the content and the format of the tested sample can lead to score inflation—where gains on the high-stakes test overestimate the knowledge and skills learned in a subject area. 53

  One example of this human tendency to “game the system” was brought to light this summer when 49 schools in 12 districts were accused by the Florida education department of teaching “template writing” for the Florida Comprehensive Assessment Test. While technically not cheating, the teachers in the schools were accused of instructing students how to showcase colorful, creative writing by including certain phrases, such as “a kaleidoscope of colors encircled me.” As a result, many exams came in with the same or similar phrases on them. 54

  Given the propensity to focus on tests, it is not surprising that the RAND researchers concluded that the type of testing that dominates our accountability systems does not appear to promote fundamental improvements to pedagogy. They found that instructional activities tended to mirror the format of test items, which provided few opportunities for students to engage in deep processing or meaningful application of knowledge by debating, evaluating, or synthesizing new ideas and concepts. In addition, the test results had a marginal impact on instructional decision-making. Time lags between administration and receipt of results and the lack of specific information on individual needs limited the utility of mid-course or individualized instructional changes.

  In addition, the increased use of tests and the application of consequences based on test scores have raised considerable controversy regarding equity and the disproportionate impact on certain groups. The National Research Council (NRC) states that numerous problems have arisen in using assessments that have been designed to evaluate the effectiveness of programs and schools to make judgments about individual students. 56

  Indeed, in its series of reports on high school exams, the Center on Education Policy (CEP) cited a number of serious issues related to the adoption of policies that require students to pass exit exams in order to graduate. CEP reports that 26 states require or will require students to pass exams in order to graduate from high school by 2012, which by then will include an estimated 75 percent of all public high
In 2009, the state-led effort to create a more meaningful set of standards for the nation’s youth gained substantial momentum. As of September 2, 51 states and territories—representing more than 43.5 million students, 87 percent of the nation’s K-12 student population—formally committed to the development of common state standards in English language arts and mathematics. Convened by the Council of Chief State School Officers and the National Governors Association in partnership with a Standards Development Working Group composed of representatives from Achieve, ACT, and the College Board, the initiative’s objective is to create “research and evidence-based, internationally benchmarked” standards that are “vertically aligned with college and work expectations and include rigorous content and skills.”

At the federal level, the Obama administration has set a priority to adopt rigorous standards and high quality assessments and supports the state-led initiative to develop common standards. U.S. Secretary of Education Arne Duncan plans to allocate part of the $5 billion in “Race to the Top” incentive funds toward a common standards effort including $350 million for the development of new common assessments. According to the Working Group, developing common standards is designed to help students by:

- Creating consistent expectations for all students—indepedent of their zip code—that ensures they are prepared for the global marketplace;
- Fostering more self-directed learning by providing students with clearer standards of what they are expected to know and do; and
- Minimizing disruptions in schooling as students transition between states.

The initiative is designed to help states by allowing them to align curricula to high-achieving countries, create professional development for educators based on best practices, provide the foundation for the development of a common assessment, and permit comparison and evaluation of policies that will impact student achievement across states.

The 51 states and territories committed to the initiative have agreed to two provisions: 1) To make a “good-hearted effort” to adopt the developed common standards, and 2) To ensure the core will represent at least 85 percent of a state’s standards.

The timeline for the Common State Standards Initiative is ambitious. Phase I was dedicated to the development of college- and career-ready standards including a review by the Feedback Groups composed mostly of college and university professors and a public comment period beginning in September. The development of fewer, clearer, and higher standards was guided by the following principles, which hold that the standards should be:

1) Able to articulate to parents, teachers, and the general public expectations for what students will be know and be able to do grade by grade and when they graduate from high school;
2) Internationally benchmarked;
3) Based on research and evidence from lead-
ing national organizations and high-performing states and countries; and

4) Ready for states to adopt.

The next step of this fast-tracked process involves an independent validation committee composed of state-nominated national and international experts on standards. By the end of the year, Phase II will be completed whereby the group will release grade-by-grade “learning progression standards” which are following a similar development, feedback, and vetting process from national and international experts.

Ultimately, the authority to adopt the K-12 and end-of-high-school standards circles back to the states. An advisory group, which includes NASBE, the State Higher Education Executive Officers, and Student Achievement Partners, will work with NGA and CCSSO to facilitate the sharing, coordination, and messaging of the common core standards. The focus of this effort is to engage state boards of education, other governing bodies, and the larger public in supporting the adoption and implementation of a rigorous and usable set of academic expectations on which a robust and coherent system of standards-based assessments and learning can be based.

Tradeoffs and Considerations

In a presentation to the NASBE Assessment Study Group in June, Wayne Camara, Vice President for Research and Development, College Board, discussed implications for state adoption, noting that it will be important for states to adopt the core set of standards without changes or deletions.

As in any consortium, the potential group benefits such as cost savings, capacity efficiencies, and comparability are lost if states add, delete, or modify the content standards intended to be shared. Additionally, while participation in this effort is voluntary, not all states are in the same place with their existing standards – some have more ground to make up, while others worry that a common core could mean lowering the expectations currently in place. States also differ in compounding factors such as predetermined standard adoption calendars and budgets.

Standards Are a Small Step

States acknowledge standards revision as necessary but insufficient to improve the postsecondary opportunities for America’s K-12 population. While the impact of the common standards movement will resonate differently in every state, it will be incumbent upon state boards of education to ensure the support systems of teacher induction and development, curriculum content and instructional materials, as well assessments and accountability are in place to make those policies effective. Standards may elevate the content but can only go so far. As state boards approve and adopt new standards, they must also consider contending factors such as student mobility and disparate resources. States will have to build the infrastructure required to ensure teachers can teach the standards and students have the resources to meet them to bring student learning improvements to scale.50

More information is available at www.corestandards.org/.
schools students and 84 percent of students of color.\textsuperscript{57}

The CEP reports cite numerous challenges to the fairness and effectiveness of exit exams given the lower pass rates for students of color, English language learners, low-income students, and students with disabilities. In the case of California, CEP cites a number of reports that found increases in the dropout rates in 12th grade that are correlated with the implementation of the state’s high school exit exam policy.\textsuperscript{58} The California Dropout Research Project (CDRP) provides further support for the detrimental impact of graduation exams in a Statistical Brief released in May 2008.\textsuperscript{59} CDRP reported that the dropout rate in grade 12 increased by about 40 percent in the past decade, with the biggest increase taking place in the 2005-06 school year, the first year of the California High School Exit Exam (HSEE) requirement. At the same time, University of California researchers released a report that looked at the impact of state HSEE on long-term trends of academic achievement of mathematics and reading. It found little evidence of effects in either area and concluded that such effects “are hardly worth the substantial economic and personal costs of state HSEEs to students, parents, teachers, and the general public.”\textsuperscript{60}

California and Arizona continue to face lawsuits challenging their exit exam policies. The suits are based on the claim that the state has failed to provide the quality of education and services students need to reach academic standards, and as a result unfairly penalizes students who have not received adequate learning resources.\textsuperscript{61} The key issue of fairness rests with ensuring that students have the opportunity to learn the material tested, and as such requires that states both hold high expectations for students, as well as provide the interventions, supports, and resources students need to reach standards.\textsuperscript{*} According to the NRC, “If the assessments are not aligned with what students are being taught, it is not fair to base promotion or rewards on the results, especially if less advantaged students are harmed disproportionately by the outcome. If current assessments do not effectively measure the impact of instruction or fail to capture important skills and knowledge, how can educators interpret and address gaps in student achievement?”\textsuperscript{63}

Clearly, if we are truly trying to significantly raise achievement levels for all students—as many countries around the globe have already done—policymakers must rethink the intent of reform efforts and the consequences of using test scores as the sole basis for action with respect to students, teachers, or administrators. How we retool our system of standards, assessments, and accountability to meet the challenges laid out in this chapter is taken up in the next sections.

\textsuperscript{*} The requirement for states to demonstrate whether a test used for graduation covers materials actually taught was established three decades ago in the landmark federal appellate court case of Debra P. v. Turlington following a challenge to the Florida high school graduation test.\textsuperscript{62}
II. Linking Assessments with Teaching and Learning

“It has often been said—and it must be understood—that even very well-designed assessments cannot by themselves improve learning. Indeed, for many years those who were skeptical of our assessment systems frequently used the old agricultural adage that “you can’t fatten a hog just by weighing it” to make their case that much of the testing going on in classrooms seemed to have little to do with learning. And looking at assessment practices in this country over the last decade makes it all too easy to say that the critics had a point. Yet it is the contention of the Study Group that a number of new developments—including research on how children learn, advances in the field of assessment, innovations in accountability systems, and dedicated work in teacher training—afford us an opportunity to retool these systems to create a new paradigm for measuring the skills and knowledge graduates need to succeed—while at the same time improving instruction so students actually acquire the necessary knowledge and skills for life in the 21st century.

In looking at the evidence, the Study Group found considerable agreement that developing this paradigm is not only possible, but necessary. To begin, presenters Eva Baker and John Tanner both emphasized the need to move beyond traditional testing, which focuses

“Improvements in learning will depend upon how well assessment, curriculum, and instruction are aligned and reinforce a common set of learning goals and on whether instruction shifts in response to the information gained from assessments.”

— National Research Council, Knowing What Students Know

Study Group Takeaway

The information gained from testing should ultimately help improve learning.
on sampling the content of a subject area such as mathematics to infer a student’s achievement level. While traditional multiple-choice tests are in fact highly predictive of performance in an educational environment, most traditional multiple-choice tests have considerable drawbacks. The sampling of discrete facts as a barometer of student achievement counters what we know about human learning and cognition and how competent performance and deep conceptual understanding develops out of learning experiences.

The Study Group concurred that information gained from testing should ultimately help improve learning. But in order to more closely connect assessment and instruction, the role of assessment must be expanded to measure complex skills in ways that reflect what we have learned from research regarding how people learn and how to assess an individual’s progress in developing competency in a subject area. For teachers to adopt practices directed at developing deep conceptual understanding and higher order thinking and communication skills, assessments for purposes of accountability must also measure more than recall of factual knowledge or performance of isolated skills.

Teachers are at a great disadvantage to work with students to build 21st century skills and competencies when it results in students performing poorly on standardized tests that focus only on surface knowledge. Only when curricula and instruction are aligned to assessment will teachers begin to use frequent assessment as part of classroom instruction to modify and refine students’ conceptual development and skills. Unless new assessment tools are aligned with new approaches to teaching, the latter are unlikely to receive attention and support at the district, school, and classroom level. The National Research Council (NRC) contends that “effectiveness and utility of assessment is predicated on the extent to which assessment educates and improves student performance, not merely to audit it.”

A. Focusing on How Students Learn

Assessment for purposes of accountability must test the high level skills that require a deep understanding of subject matter and the ability to apply knowledge to new contexts. Findings about the nature of learning offer new directions for the design of assessments to better capture information about students’ competencies and thereby enhance learning. Over the past several decades, the knowledge base regarding how children develop understanding, how people reason and build structures of knowledge, and what shapes competent performance has expanded. Research shows that knowledge-centered classrooms focus on what is taught (information, subject matter), why it is taught (understanding), and what competence or mastery looks like.

For students to deeply process subject matter that characterizes competent performance, they must have opportunities to acquire organized sets of facts and skills that are connected to meaningful problem-solving activities. To develop competence in a subject area, students need help to understand why, when, and how those facts and skills are relevant. As factual knowledge and the conceptual framework builds, students are able to plan a task, notice patterns, generate reasonable arguments and explanations, and draw parallels to other problems. Knowledge of a large set of disconnected facts will not suffice to organize knowledge in ways that facilitate deep understanding of subject matter and allow the learner to retrieve and apply information in novel ways. Most widely used assessments even for classroom instructional purposes only measure discrete bits of knowledge and are based on learning theories that are not fully in line with current knowledge about human cognition and learning. As such, they provide little information to help teachers design instructional techniques toward explicit learning goals.
Eva Baker, director of UCLA’s National Center for Research on Evaluation, Standards, and Student Testing (CRESST), shared with the Study Group her research and development work on designing assessments for 21st century learning. Researchers such as Baker have probed deeply into the nature of expertise, revealing how people learn the content and procedures necessary for competency in a specific domain. Baker described ontologies, which use a structured, empirically based approach using “experts” who arrive at consensus about domain(s) for standards, teaching, and learning. Drawing from cognitive science, the webbing or concept map describes performances of a particular skill or domain expressed in terms of principles, knowledge, key situations, and the relationships among them. While the particulars of the research and application to the design of measurement processes are beyond the scope of this report, she reiterated John Tanner’s comments about how tests historically have served to identify differences among students rather than illuminate individual performance.

Studies of expert performance have revealed how people acquire the mental structures necessary to retrieve information and apply knowledge and skills in adaptive ways to solve problems. Baker’s cognitive analyses of subject domains have helped identify how students progress toward competency and what’s needed to help learners organize knowledge, represent problems, and monitor their own learning. The results have provided important information about ways to connect how people learn with the processes of assessment and instruction. The results of these efforts have provided clearer benchmarks of student thinking so that teachers can understand what has preceded and what is likely to follow the student performance they observe in their classrooms.

In order to focus on what learners need, teachers must pay close attention to the progress of each student along this path, or “developmental corridor,” leading to a standard of expected performance. They need to tailor instruction to take the learner to the next level that is challenging but attainable. This requires assessments that make clear to both teachers and students where students are along the developmental corridor.

Here a key principle of learning is paramount: teachers must start with what students currently understand and know about a topic and build from there. Assessments can be designed to identify current student thinking, identify conceptual errors, and then move the student toward more sophisticated understandings. As Tanner told the Study Group, “The next level is a known entity, every student has somewhere to go no matter where they happen to be, and teaching is about getting students to that next level.” To accomplish this, teachers need to recognize the importance of prior knowledge, the power of practice and feedback, and the principles for facilitating the transfer of knowledge and the generalization of knowledge to new situations.

Focusing on how people learn will help teachers move beyond such false dichotomies as whether curricula should emphasize the basics or teach thinking and problem solving. Both are necessary. According to Bransford, students’ abilities to acquire organized sets of facts and skills are actually enhanced by helping students understand why, when, and how those facts and procedures are relevant to solving meaningful problems. Attempts to teach thinking skills without a foundation of factual knowledge will fail to foster problem solving ability or promote transfer to new situations. Using what we know about how people learn will bring greater coherence to curriculum, instruction, and assessment. “This cohesive function is a crucial one because educational assessment does not exist in isolation, but must be aligned with cur-

“Focusing on how people learn will help teachers move beyond such false dichotomies as whether curricula should emphasize the basics or teach thinking and problem solving. Both are necessary.”
B. Performance-Based Assessments

Baker also advocated measures that incorporate performance elements as the best way to support teachers in making the right instructional decisions for each student. During the 1980s and ‘90s, a number of states sought alternate approaches to traditional test modes that would use more open-ended tasks that call upon students to apply their knowledge and skills to solve a problem or create a product. Constructed response questions, which require students to write short-essay responses or solve problems and show their work, are currently the most popular performance items used on state tests, though some states have also worked with more extended performance tasks and portfolios.

These early efforts to bring theories of learning and cognition together with issues of curriculum and assessment gave rise to the first designs of large-scale performance assessments. These forms of assessments were intended to measure the type of complex skills that policymakers want students to develop. Students were required to perform “more authentic” tasks such as the ability to write a persuasive paragraph or to draw inferences from a science experiment. These measures were also assumed to be more in line with classroom instruction and would, therefore, better leverage improvements in teachers’ pedagogical skills, as well as respond to the individual’s progress in content and skill acquisition.

There were problematic outcomes during this initial period when performance-based tests were developed and embedded in high-stakes, large-scale assessments. First, the task and method variances were found to be high, resulting in low reliability and limited validity. “Simply because a task is ‘authentic’ does not mean it is a valid observation of a particular construct,” Baker wrote in 1997.

Second, a number of analyses revealed that there was a mismatch between what the designers intended to measure and what was actually assessed. As it turns out, many of the tests focused on a set of procedures to be followed in completing a task (task-centered) rather than engaging students in complex thinking processes tied to key learning objectives (construct-centered). Third, considerable more time and money were required to design, administer, and score performance assessments than machine-scored multiple-choice exams.

In part, these issues surfaced because performance elements were included in large-scale assessments, where a number of constraints operate in order to serve accountability and auditing purposes. The assessments must provide reliable and comparable scores for individuals as well as groups of students; sample a broad set of content standards within a limited testing period; and offer cost-efficiency in terms of development, scoring, and administration. In general, where common constructed-response questions and writing prompts have been included in state tests with
Examples of States that Use Performance-Based Assessments

**Vermont** developed on-demand performance tasks and writing and mathematics portfolios developed at the state level with extensive engagement of teachers. These ongoing assessments were used as the state’s primary assessment for accountability for more than a decade and are now implemented on a voluntary basis in addition to the annual tests required under NCLB. The assessment system enabled teachers to provide continuous real-time feedback and shifted the state’s professional development paradigm. Through collaborative networks and convenings, teachers’ discussions around what constitutes quality student work have promoted grading reliability and improved instructional practices.83

**Ohio** is considering assessments such as portfolios, senior projects, and small group collaborations as an alternative to the five-part Ohio Graduation Test currently required for a diploma. Field tests are underway with educators across the state to determine if performance tasks are aligned to the skills and processes students will need to be prepared for advanced training and work.84

**Nebraska’s** performance assessments are developed and scored by local educators who are trained to ensure inter-rater reliably. The state monitors validity through the administration of a statewide writing examination and a norm-referenced test as part of a peer review by measurement and assessment experts.

**Connecticut** adopted a statewide testing program integrating performance tasks in the mid-1990s that is still in place today. The Connecticut Academic Performance Test (CAPT) component administered in the 10th grade includes classroom-embedded tasks and involves teachers at various levels of the system from development to standard setting. The CAPT measures individual student achievement relative to state standards in four areas—mathematics, reading across the disciplines, writing across the disciplines, and science—through a range of tasks including “on-demand” experiments that run through the scientific method from design to formulating hypotheses, collecting data, and conducting analysis, critiques, and evaluations. In 2000, the State Board of Education passed a measure that prohibits the use of CAPT as an isolated determinant of graduation.85

**Maine, Vermont, New Hampshire,** and **Rhode Island** developed performance assessment components as part of their accountability systems in conjunction with more participation on the part of the state in helping local districts implement their assessments. The New England states worked together to jointly construct the New England Common Assessment Program (NECAP) that includes locally developed assessments that provide evidence of student work from performance tasks and portfolios (see text box on page 14).

**Rhode Island’s** 2008 high school graduates were the nation’s first students required to take a performance-based assessment to earn their diplomas.86 The state-mandated assessment expects graduates to show evidence of success across three elements of the performance-based graduation system: a standardized reference exam; course performance, and state-approved performance assessments such as portfolios, senior projects, and/or end-of-course exams. The results from first year implementation report greater student engagement and increased graduation rates (up to 74 percent from 70 percent).87

**Delaware’s** education department is in the process of replacing the Delaware Student Testing Program with a performance-based system, the Delaware Comprehensive Assessment System. In addition to online summative assessments, end-of-course exams, writing assessments, and computer-adaptive growth assessments, a Classroom Assessment Item Bank for teachers will be created to provide formative assessment content and tools aligned to the state’s grade level expectations.

responses to them centrally scored, reliability has not been an issue. However, the more the determination of tasks and scoring of resulting work have been relegated to local control—without auditing—the less consistent the task difficulty and scoring have been, causing problems with comparability of results across sites. States opted to adopt multiple-choice items that are generally much more efficient under these conditions—and consequently, when it comes to large-scale assessments states and their testing contractors have generally produced these “on-demand assessments” that are administered to all students under standardized conditions.

But as discussed in this report, multiple-choice tests do not capture rich information about the nature of student understanding. They provide minimal feedback to the examinee, instead providing information to others (district and state administrators, the public) who are removed from the process of teaching and learning. As a consequence, most multiple-choice state tests are criticized for low validity because they do not tap a broad range of essential characteristics of what it means to be proficient in a given subject, nor do they provide useful information about such important skills as knowledge organization, problem representation, and use of metacognitive strategies. As Pellegrino points out, “In 21st century learning environments, decontextualized, drop-in-from-the-sky assessments consisting of isolated tasks and performances will have zero validity as indices of educational attainments. In essence, assessment will need to transform itself to remain relevant and useful.”

Fortunately, a wider variety of measurement models and tools are currently available to support the kinds of inferences about student competence that cognitive science suggests are important aspects of knowing and learning. Moreover, we now have measurement models that can characterize students in terms of multiple aspects of proficiency; tap 21st century skills such as collaboration, problem solving, self-regulation, and communication; chart students progress over time; and address alternative methods of judging performance at the group, class, school, and state levels.

In a recent report from the Alliance for Excellent Education, Darling-Hammond details how performance measures are used in states (27 in the United States use various approaches to high school graduation decisions that combine state requirements with local performance assessments such as portfolios of work and senior projects), as well as other countries and jurisdictions such as Finland, Sweden, United Kingdom, and Hong Kong.

Darling-Hammond and Pecheone note that local in-school performance assessments serve as the dominant mode of testing in most of the high-achieving countries around the world (e.g. Hong Kong, Singapore, Finland, and Sweden). These measures incorporate rich assessment tasks at the classroom level such as research papers, presentations, and lab experiments. At the high school level, these countries often use a combination of centralized, national exams (with primarily open-ended and essay items) and locally developed tests. “The centralized assessments are often developed jointly by high school and college faculty and scored using common criteria by teachers. The classroom-based assessments are often developed jointly by high school and college faculty and scored using common criteria by teachers.”

The classroom-based assessments—which include research papers, applied science experiments, presentations of various kinds, and projects and products that students construct—are mapped to the syllabus and the standards for the subject,
and are selected because they repre-
sent critical skills, topics, and con-
cepts.” Countries and jurisdictions
such as Finland and Hong Kong
create banks of tasks that teachers
can draw from that include rich as-
essment tasks for classroom use for
formative or benchmark purposes.

The authors focus on the centrality of teachers to the process of
developing, administering, and
scoring school-based classroom assess-
ments. In this way, the deploy-
ment of in-class performance
measures serves as robust teacher
development that fosters teacher-
buy-in and readiness to adopt new
instructional practices. Teachers are
trained to administer and evalu-
ate student work using collabora-
tively determined criteria specified
through standardized rubrics and
scoring guides. This not only serves
to ensure quality and consistency, it
strengthens the connection between
assessment and instruction and
helps teachers “learn to calibrate
their understanding of the stan-
dards to common benchmarks.”

Another advantage of collective
scoring is that rather than grading
in isolation, teachers must collabor-
ate to set the criteria for judging
student efforts. This helps teach-
ers gain multiple perspectives on
learning and leads to improving
instructional practices based on a
shared definition of what consti-
tutes mastery or competency. Embed-
ning performance measures at
the classroom level permits a finer
grain analysis that allows teachers
to assess student learning along a
number of dimensions such as the
ability to frame a problem, gener-
ate hypotheses, organize informa-
tion, persist in problem solving,
and frame a coherent oral and/or
written response. Local scoring pro-
vides immediate feedback to teach-
ers and students affording the op-
portunity to diagnose how students
are progressing and why they may
be struggling.

The examples of classroom-based
performance assessments reinforce
what educators and policymakers
have recognized: that improve-
ments in teaching and the quality
of student work will require a dif-
f erent set of measures embedded
in day-to-day instruction. A more
detailed discussion about assess-
ment for learning is included later
in the report, but of note here is that
studies have shown strong effects
for using a diverse array of inter-
ventions, all of which have featured
some formative use of assessment
data or processes. To date, despite
strong evidence of its substantial
benefits, classroom assessment
has been seriously underutilized
in our schools. Pellegrino writes
that “efforts to use assessment to
drive academic outcomes provide
relatively little evidence that as-
essment external to an ongoing
process of learning and teaching
can in fact produce the desired edu-
cational outcomes.” At the same
time, measurement experts warn
that classroom assessments will im-
prove educational outcomes only if
they become an essential part of the
design of contemporary learning
environments and an integral part
of teacher preservice training and
professional development.

C. Defining the Purposes of Testing

The Study Group participants
agreed on the foundational premise
that an assessment system must be
designed to improve student learn-
ing, but recognized that no single
test would suffice. Inherent in the
design of all tests are trade-offs and
constraints that derive from their
specific purposes. The sharp con-
trast between classroom and large
assessment, for example, arises be-
because they serve distinct purposes
and provide different information
to different audiences.

In the case of large-scale assess-
ment, policy leaders look for trends
across schools and districts and
need broad patterns of performance
to plan funding and make policy
decisions. They need to know
whether larger aggregates of stu-
dents have had the “opportunity

“Measurement experts warn that classroom
assessments will improve educational
outcomes only if they become an essential part
of the design of contemporary learning
environments and an integral part of teacher
preservice training and professional
development.”
to learn” essential content and skills identified in the state standards. Standardized multiple-choice tests offer efficiency in terms of testing time and cost and provide reliable and comparable metrics for individuals and groups of students across time and place. They sample a broad set of content standards within a limited testing period that yields evaluative and trend data to individuals external to the classroom. Test designers respond to what states need in terms of testing within a limited testing period the material covered during an entire year of instruction. Designers create “on-demand assessments” that are administered to all students under standardized conditions and include many tasks that can be responded to quickly. These types of assessments serve the purposes of an accountability and auditing model—signaling the aspects of academic competence that are valued by a society and worthy of recognition and reward.92

Yet, because the information serves the interests of those who are removed from the processes of teaching and learning such as district administrators, program evaluators, and state policymakers, their utility at the classroom level is severely limited. They do not capture rich information about the nature of student understanding in relation to the personal learning experiences of examinees.93 As a consequence, it is essential to use other forms of assessments closer to the classroom that enable a finer-grain analysis of students learning and performance.

Classroom assessment helps to maximize learning by linking instruction and assessment more explicitly. Students learn more by understanding the goals of instruction and how their work will be evaluated. Teachers gain insight into the particular qualities of the students’ work and about what they can do to improve student’s performance on the pathway to competence.

A better balance between large-scale and classroom assessment is needed, particularly given the strong evidence of the potential impact of the latter on student learning. The enormous resources invested in the design of large-scale assessments far outweigh the research and funding available to improve the quality and use of classroom assessments. Instead, the Study Group recommends that states use an assessment system that includes multiple assessments or components aimed at a set of explicit learning goals and results. What’s needed is a new paradigm to craft a coordinated assessment that operates in tandem, along with curriculum, and instruction to promote learning. Pasquale DeVito, who served as the director of the Board of Testing and Assessment for the 2001 National Research Council report, presented information to the Study Group on the design of comprehensive assessment systems and cautioned participants on the use of tests:

- The important thing about a test is not its validity in general, but its validity when used for a specific purpose;
- Tests are not perfect. No single test score can be considered a definitive measure of a student’s knowledge; and
- Neither a test score or any other kind of information can justify a bad decision.

Finally, while the purpose of the test drives its design and use, it is critical that the component assessments that form a comprehensive system are not at cross purposes. For example, since teachers model their own tests after the high-stakes accountability tests, it is important that the latter model good classroom testing. All assessments should be based on a model of student learning that is grounded in the research on cognition and learning. How assessments are designed and used will depend on sets of questions such as: What do we want to learn from this assessment and why do I want to learn it? Who will use the information and what action steps will be taken as a result? What professional development or support structures are needed to ensure successful follow up? How will student learning improve by using this assessment? Will it improve more than if the assessment system was not used?94

“The enormous resources invested in the design of large-scale assessments far outweigh the research and funding available to improve the quality and use of classroom assessments.”
Basic Types of Assessments

**Summative assessment** describes the extent of student learning (content and/or skills) over time, relative to content standards and may include proximal measures of classroom learning such as end-of-course or semester tests that tend to contribute to students grades. State assessments are summative measures, including interim or benchmark assessments, that are designed to provide broad domain or sub-domain coverage of content covered in state assessments. The results, expressed most often in a number or score, are usually used for accountability purposes such as assigning student grades, evaluating a teacher’s instruction, or evaluating a school’s success. Classroom summative assessments such as unit tests or other graded performances are administered after material is taught and as such do not provide a diagnostic lens on student performance. These may be constructed locally or provided with curriculum materials.

**Formative assessments** are used to guide teaching and learning processes. It refers to a number of assessment activities that teachers and students engage in on a day-to-day basis to gauge students’ understanding of content along with the performance skills used to demonstrate understanding. It may include any number of instructionally embedded activities such as observations, homework, student writing, quizzes, or project-based work. It should be developed locally in accord by teachers to diagnose gaps in student learning and adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes.

**Interim or benchmark assessments** that fall between formative and summative may be used for different purposes (instructional, evaluative, predictive). They generally evaluate students’ knowledge and skills to specific goals, within a limited time frame after instruction has occurred, and are designed to inform decisions both at the classroom, school, and district level. In contrast to formative assessments, they are not embedded within instruction to guide immediate actions to further learning. Results from interim assessments can be meaningfully aggregated and reported on a broader scale and so they tend to be administered by districts and schools rather than teachers.
III. Creating a Balanced Assessment System for the 21st Century Learner

A. The Need for Multiple Assessments

The previous chapters have emphasized several points that are worth repeating here:

- First, policymakers must recognize the need for multiple purposes for assessments, each of which provide the different types of information required at different levels of the education system.

- Second, the two primary purposes of assessments are accountability and improving teaching and learning.

- Third, of these purposes, assessments for improving teaching and learning are most notable by their absence.

- Finally, what’s most critical is that state education leaders integrate these multiple assessments with their varied purposes into a coherent, comprehensive system that works together toward explicit learning goals.

In order to achieve this aim, states must develop a vision and overarching objective for the design of a balanced assessment system that is based on the latest research about how students learn. For multiple assessments to support learning effectively, the purposes served by each assessment, the type of information the measure yields, and how the information will be used must be made explicit to users. It will require states to develop a theory of action about how assessments provide information important to improving student learning and how the information will be used to drive educational decisions at each level of the system.

Because the effectiveness of a test in advancing learning ultimately depends on how connected it is to curriculum and instruction, assessment design should be guided by current models of student learning that describe the progression of

Study Group Takeaway

States need to generate new assessment and accountability systems focused on:

- improving student learning at levels needed to meet postsecondary expectations and
- reducing disparities in performance at all levels of the system.
how students attain competency within a given subject area. These models serve to describe how students learn in ways that are clear for users and that are maximally useful for guiding instruction and assessment. The learning models, then, effectively serve to align the system.

Problems arise in test usage and interpretation when there are conflicting assessment goals and the resulting information causes confusion for end-users. It will be critical for states to move away from one-shot testing to crafting an approach to collect multiple sets of observations that enhance the value and utility of the information at each level of the system. Tests should be nested within a system, each providing finer- or coarser-grain information appropriate for use at the state, district, school, classroom, and student level.

In their separate presentations to the Study Group, Stuart Kahl, CEO of Measured Progress, and Doris Redfield, CEO of Edvantia, identified critical state actions that can promote this process, including:

- Shifting investments in research and development from a single-point focus on large-scale assessment toward classroom assessment where teaching occurs;
- Incorporating multiple assessments into a system of curriculum, instruction, and educator development that focuses on effective instructional practice;
- Defining a set of learning goals that functions as part of a comprehensive, balanced assessment system that is vertically aligned; and
- Ensuring that the resulting information from the testing system has maximum utility for guiding instruction in relation to the learning goals (as identified by curriculum standards and informed by cognitive research and theory).

**B. The Need for Changes to Large-scale Assessment Models**

In order to create a balanced assessment system, states will need to consider alternate models that relax some of the constraints that currently drive large-scale assessment practices. This is needed because, as we’ve seen earlier, the big standardized tests typically in use today do not provide the kind of rich information about more complex student knowledge and skills that should, ideally, be part of our 21st century learning goals. As is commonly said about testing, if you don’t ask for it, you won’t get it. Or to put it another way, it’s also commonly said in education circles that it’s okay to teach to the test—if the test is a good one.

Fortunately, there are a number of ways to craft blended approaches that combine externally controlled assessments with work samples drawn from coursework in order to produce richer information about student learning. For example, a summative measure can include evidence extracted from student performance on classroom work (curriculum-based assessment). Test designers can consider broader approaches to combining different types of tasks to ensure their assessments capture what is being asked for in new learning goals. Student work or scores on classroom assessments can be used to supplement the information collected from an on-demand assessment—recognizing that on-demand tests are not directly related to the examinee’s classroom learning experiences.

A balanced system could include assessments administered at a fixed time or give students the option to select items from a bank of tasks that match the students’ particular selection of courses or designated career path. For example, the International Baccalaureate Program, designed for 11th and 12th graders, assesses student learning using both school-based performance assessments and external exams at the end of each course. Both of these assessments capture students’ knowledge and skill levels relative to the curriculum objectives detailed in “subject outlines” for each course. The school-based measures (e.g., oral exercises in language subjects, student portfolios, laboratory and mathematical investigations, artistic performances) contribute 30 to 50 percent of the final examination score. The exam includes mostly essays, constructed responses, data response questions, case study questions, and text response questions, with a limited use of multiple-choice items.
Many high-performing nations combine a national standardized exam with school-based curriculum-embedded assessments that include projects, presentations, and essays. In the United Kingdom, for example, secondary school students who have concentrated on physics as one of three chosen subjects must take a university entrance exam consisting of a variety of elements: multiple-choice questions, short answer questions, responses to questions on a reading passage; tasks requiring the examinee to draw conclusions from a set of presented data and to explain phenomena described in short paragraphs; and short problems with equipment set up in a laboratory. In addition, students’ performance on school-based tasks (an investigation and a research paper about a topic chosen by the student) is included. The combination of tested material strengthens the linkage between assessment, curriculum, and actual instruction.101

Another alternative for reducing the constraints of large-scale tests is the use of matrix sampling rather than census sampling, particularly if the assessment’s purpose is program evaluation. Instead of all students taking the same test, different students take portions of the larger assessment and the results are combined to obtain an aggregate picture of student achievement. NAEP, for example, uses a complex form of matrix sampling to provide its national and state achievement reports.

Finally, technology will continue to offer innovative approaches to learning and assessment. Sophisticated information technologies can support the creation of more powerful learning environments; provide multiple elements for enhancing learning and instruction; offer opportunities to measure complex reasoning and problem solving; and help manage data warehousing and interpretation. A range of technologies are already available to create embedded assessment strategies such as simulations and virtual reality environments that can enhance students’ knowledge and skills while providing feedback in real time to improve performance.

Regardless of the particular mix of assessments, a comprehensive assessment system should subscribe to a unifying, research-based model of student learning and, of course, be technically sound and timely for their given purposes. Users must be able to trust the accuracy of the information, to draw inferences about individual performance, and to make educational decisions for specified purposes. While much of the recent focus has been on the technical quality of external large-scale assessments, classroom assessments must also meet standards for reliability, validity, and fairness if they are used as part of a composite measure of student achievement.
C. Rethinking Accountability

From the classroom to the boardroom, education systems face considerable challenges in retooling assessments to ensure that graduates can meet expectations for advanced training, a global economy, and civic participation. For state policymakers, one significant challenge will be revamping accountability to work as part of a comprehensive assessment system that fits multiple purposes and uses and that shifts the focus toward improving classroom instruction and student learning. This is important because it is accountability systems that so often drive the nature and use of tests.

Assessments for accountability do have a place and serve to provide valuable data to inform school and program improvement efforts. However, because they now include such high stakes, accountability systems are problematic, and indeed, the validity of state accountability systems as framed under NCLB have come under considerable scrutiny. The focus on a single test score as the measure of school quality and student achievement is suspect when the increase in state test scores does not translate into deeper learning and subsequent academic success. Along with the lack of major gains on NAEP exams as states have shown improvements in their own test results, other indicators suggest that the current focus on test-based accountability is flawed.112

A 2009 report by the Northwest Evaluation Association examined the impact of decisions states have made in implementing NCLB.113 The authors concluded that what drives the discrepancies as to whether a school is judged in need of improvement rests not with how “good school” and “bad school” are defined, but the arbitrary nature of selecting cut scores, minimum subgroup sizes, and the confidence intervals around those scores. States have latitude on a number of aspects of those elements that are required as part of NCLB including setting their own academic standards, developing their own tests, and defining proficiency. The proficiency standards or the cut scores on the state tests vary widely. Given this variance, the authors examined whether test-based accountability produced consistent judgments about schools across states.

Researchers administered an independent test to students in 18 elementary and 18 middle schools and looked at those performances in relation to each states’ proficiency cut scores and 2008 annual targets. Within the elementary sample, the number of schools that made AYP varied widely by states. Almost all of the sampled schools failed to make AYP in some states, and nearly all of these same schools made AYP in others. If the schools were in Massachusetts, which has high proficiency cut scores, only 1 of 18 schools would have made AYP. But if they were in Wisconsin, 17 schools would make AYP. In other words, the same student performance measured on the independent test generated far different results and interpretations across different states. The authors wrote that “one of the adages of the NCLB era is that a child’s zip code shouldn’t determine her life chances. Indeed. But neither should a school’s zip code determine whether or not it makes AYP. Yet regrettably it does. And so the success or failure of a given school under NCLB is driven as much by the way the law is implemented by its home state as it is by the performance of its students and the amount of progress they’ve made over the course of a year.”114

Given such idiosyncratic outcomes, disenchantment with our current accountability system is widespread. While recognizing the important contribution NCLB has made in forcing disaggregation of student subgroup data to focus attention on achievement gaps, education leaders across the country have urged federal policymakers to abandon the law’s disproportionate attention to basic skills narrowly defined:

- First, the problems in accountability systems—whether in education, health, or other fields—that focus on quantitative data alone are well established.118

Study Group Takeaway

States need to expand the types of data used in accountability systems rather than relying so heavily on test scores.
Integrating Technology in Assessment Systems for the 21st Century Learner

States are aware of the growing saturation of technology in the global marketplace and higher education, and the need to align the next generation of K-12 assessments with technological advances to ensure relevancy and readiness to these expectations. The convergence of the power of technology—to redefine parameters of space, time, communication, and personalization—with the implications of assessment on educational practice provides a significant opportunity for state-level investments to get closer to the classroom level and focus on actual gains in student performance. Yet as highlighted in this report, the potential of technology rests with how well its design is grounded in the sciences of cognition and measurement and the degree to which it integrates curriculum and instruction to improve student learning.

Clearly, extremely powerful technologies will become more ubiquitous throughout K-12 schooling. Already during the last school year, 26 states offered some form of computer-based assessment, including low- and high-stakes testing for both formative and summative purposes.\(^ {102}\)

**Benefits of Technology**

In the short term, online assessments offer the following immediate potential benefits:

- **Speed:** improves efficiency in administering tests, expedites input of student demographic data, and reduces time needed to score and report results.

- **Transparency:** widens test monitoring capabilities and accessibility of results to the public.\(^ {103}\)

Over the longer term, researchers note other ways technology-based assessments could more fundamentally change the learning landscape.\(^ {104}\)

- **Structure:** allows for assessment of more complex and authentic skills and understandings, as well as the use of a broader choice of content (audio, video, Internet) and formats (e-portfolio, blog, hand-held student response system) to deliver, demonstrate, and measure learning.

- **Relationships:** encourages collaboration, co-creation, and continuous evolution—including teachers networking to share best practices, students soliciting feedback without logistical barriers, and fluidity in the designation of students and teachers in expert and participant roles.

- **Human Capital:** offers potential to train educators in innovative instructional practices, develop teachers and principals in data analysis to focus on classroom and school decision-making, and provide students with experience using the tools and thinking processes needed to productively contribute to the knowledge-based economy.

- **Individualization:** provides rich insight into student approaches, challenges, and performance, as well as the capacity to customize resources for educators to address these identified learning gaps.

**Examples at the National, State, and District Level**

During the last 20 years—with heightened intensity in the last decade—states have made systematic and substantial investments in technology for education. Integrating technology into assessment systems poses steep costs upfront for broadband access, equipment, and training. Over time and as these systems are brought to scale, however, proponents believe payoffs in student performance and teacher quality are potentially transformative. Below are several examples of how technology is addressing assessment and accountability:\(^ {105}\)

- **Simulating multi-step, real-world problems and the processes of human cognition in a structure not easily presented or measured using a traditional static test format.**

- **The United States Department of Education is developing a new technological literacy component of National Assessment of Educational Progress (NAEP) which will be administered to 4th, 8th, and 12th graders in 2012. A pilot of the Problem-Solving in a Technology-Rich Environment given to 8th graders required students to search a simulated version of the Internet, locate and synthesize information, and complete repeated trials and experiments. In a compelling and interactive...**
structure that combines higher-order thinking and domain-specific knowledge, the test yields a score for the overall task as well as marks for particular skills embedded in the assessment and information about the approaches students use to obtain their answers.106

- Similarly, a performance-based assessment run by the Council of Aid to Education tests student ability to read complex texts. The computer-based College and Work Readiness Assessment gauges students’ dexterity in scanning and organizing information from various texts, evaluating the merits of their arguments, and drawing conclusions.107

- The Diagnoser, developed by Facet Innovations, shows how computerized testing can also address some of the criticisms of multiple-choice testing. The assessment module better aligns with human cognitive processes by breaking down facets, or pieces, of student thinking and adapts based on student answers. Honing in on a small area with layers of “right/wrong” responses specifies “what students know and are able to do on the way to competence” and allows more educated inferences to be drawn from tests about where and why student understanding waivers.108

- Resolving the disconnect between assessment data and instructional improvement by facilitating more periodic testing and building platforms to better inform district- and school-level decision making.

- Indianapolis Public Schools adopted software that comingles its core curriculum and formative assessment data. By pinpointing gaps in student data and providing intentional, concrete solutions for educators, this model has accelerated student achievement. In 2007, nearly half of the students who were labeled as “high risk” for reading difficulty in fall 2006 went on to meet or exceed the benchmark level, with significant numbers of students also progressing out of warning categories.109

- Pennsylvania’s Value Added Assessment System (PVAAS) is a web-based tool that provides all districts in the state with student progress data to consider alongside student achievement data. Growth and projection information from PVAAS offer a more precise way to evaluate how schools add to students’ educational experience and to troubleshoot gaps identified among subgroups and individual students.110

- Empowering the exchange of responsive, adaptive tools that drive deeper integration in the day-to-day instruction that’s proven to impact individual student achievement.

- Arizona has an online formative assessment bank, including more than 5,500 sample items aligned to its state standards and sample pre- and post assessments for classroom use. This provides structure with flexibility for educators to experiment and make adjustments relevant to their students’ specific needs. Alaska’s Grade Level Expectations Item Sampler has a similar purpose and audience.

- Virginia implemented a web-based test delivery system and paired it with an online Educational Information Management System beginning in 2001. One test in the state, the Algebra Readiness Initiative, uses adaptive diagnostic tests to determine and allow teachers to appropriately modify instruction for students who are in need of intervention to pass Algebra I, a course where mastery is highly predictive of postsecondary success. The state is working toward administering all of its state achievement tests online and administered more than 1.4 online tests in the spring of 2008 alone.111

Considerations and Conclusion

The innovative ability of technology does not come without potential disadvantages and difficulties. States must consider issues including ongoing maintenance, security, educator preservice curriculum, measurement, equity (e.g., students’ varied exposure to technology), how to use data most productively, and the potential setbacks imposed by accountability system pressures.

No longer isolated as an add-on consideration, integrating technology in assessment redesign is emerging as a chance for states see high returns on their investment in broad educational reforms, all rooted in the goal of improving learner outcomes.
• Second, it is essential to determine the purposes of assessment and to validate and use tests strictly in terms of those purposes, but few of the tests used for NCLB have been validated for the way they have been used (e.g., for evaluation of schools and districts; for measuring the impact of instruction and programs).\textsuperscript{119}

• Third, approaches are needed to attend to the development of the whole individual including academic competency, physical health, character, and civic and social development throughout schooling.

Accountability systems should promote goal-driven action to ensure students meet 21st century expectations and to ensure that educators use appropriate strategies to meet these goals. The focus should be on how the system performs overall in a number of key areas and should make use of multiple indicators—of which summative assessment is only one. Examples of possible indicators to drive school improvement and the quality of instruction and student learning include:

- school self-evaluation;
- external peer evaluation;
- existence of enriched academic content;
- existence of high-quality early childhood programs;
- existence of comprehensive literacy program and supports to struggling readers;
- extent of advanced placement classes;
- structured opportunities for extended learning time;
- implementation of response to intervention programs;
- extent of quality professional development for staff; and
- existence of full service school-connected health centers for all disadvantaged children.\textsuperscript{120}

States should use a broader range of data for different purposes and move away from test-based accountability by using, for example, a composite quality index. A number of education leaders recommend that states should institute the type of inspectorate system used in England, Scotland, Netherlands, Belgium, France, New Zealand and other nations to determine if schools are following practices that will lead to the attainment of broad long-standing educational goals.

Darling-Hammond also points to the example of high-performing countries that have a stronger focus on teacher development and overall school improvement rather than using tests largely for accountability purposes. She recommends enhancing school and district performance through the use of accreditation and site-inspection processes to identify the specific improvements needed and guide the intervention efforts.
of state departments and districts. These may include school self-appraisals or inspections along with external peer evaluations using a set of performance indicators. Such measures should assess how well the school 1) uses multiple forms of classroom-embedded assessments to effectively gauge students’ performance with respect to knowledge, skills, and attitudes and 2) uses curriculum evaluation data to inform planning, improve teaching to standards, and to share knowledge and reflect on practice. In summary, schools and districts must be held accountable for student achievement, but using an approach that goes beyond test scores alone and includes appropriate qualitative and quantitative methods.121

In a similar vein, Rothstein, Jacobson, and Wilder recommend that states conduct mandatory inspections in each school and in related community institutions about once every three years to determine whether students are achieving adequate outcomes in eight goals that have remained constant throughout U.S. history—basic academic knowledge and skills; critical thinking and problem solving; appreciation of the arts and literature; preparation of skilled employment; social skills and work ethic; physical and emotional health; and citizenship and community responsibility. Inspection teams would use comparisons with high-performing schools with similar demographics to assess what areas of school quality require improvement.122

The authors conclude that the next generation of accountability systems should create more productive incentives that drive improvements as intended; measure growth, not just status in student achievement; report the full score scale as well as performance levels, not just a small number of cut points; reduce incentive to teach to the test by varying the test content and format; and ensure that the supports and incentives to improve practice operate directly at classroom level. “Perhaps the most important reason why NCLB, and similar testing systems in the states, got accountability so wrong is that we’ve wanted to do accountability on the cheap. Standardized tests that assess only low-level skills and that can be scored electronically cost very little to administer—although their hidden costs are enormous in the lost opportunities to develop young people’s broader knowledge, traits, and skills.”123

To summarize, states need to experiment with alternate approaches to the design, frequency, and universality of tests deployed for purposes of accountability. At the same time, multiple assessments should be used for guiding instruction, identifying an individual student’s strengths and weaknesses, and certifying students’ preparedness for graduation. These assessments should be of high quality and focus not only on academic skills and cognitive growth, but on the knowledge and skills needed for students to become effective participants in a global environment. In his presentation to the Study Group, Stuart Kahl noted, “No matter how high or low the cut scores are set for state tests, our students are not performing particularly well with respect to the content standards states currently use. Our basic need is not necessarily common standards, but rather, high expectations for our students with respect to well-defined content in every state. And of course, common content standards would be no substitute for effective teaching practice.”124

It’s a complex challenge to balance accountability with the design of assessments intended to maximize student learning. We need a more flexible, supportive accountability system that takes into account the needs of students and teachers and attends to what the broader system must provide to advance important outcomes that will ultimately benefit both individuals and society at large.

Study Group Takeaway
Many experts recommend that we use qualitative measures such as school inspections in addition to quantitative data (test scores, attendance rates) in order to get a fuller picture of school performance.
A. The Advantages of Formative Assessments

This report closes with a focus on what may be the most important use of assessment: formative assessment or assessment for learning. In contrast to summative assessment that takes place after learning is supposed to occur, formative assessment refers to those measures used to promote, not merely judge or grade, student progress. It has become increasingly evident to experts in measurement science as well as policymakers that assessments conducted solely external to the ongoing teaching and learning processes are unlikely to produce desired outcomes. In their presentations to the Study Group, measurement experts Stuart Kahl, Doris Redfield, Laura Hamilton, Pasquale DeVito, and Eva Baker all emphasized the importance of embedding richer and more valid assessments into the context of day-to-day classroom instruction if we are to realize the goal of producing high-level outcomes for all students.

The advances that connect cognitive theory with assessments can enhance teachers’ abilities to chart individual students’ progress and tailor instruction to help them achieve the next level of understanding and competency. This requires assessments that make visible to both teachers and students where students are along a developmental trajectory for acquiring essential knowledge and competencies within a particular subject area. Teachers can use the assessments to identify current student thinking, identify conceptual errors, and then move the student toward more sophisticated understandings. Yet most of the investment in terms of money, time, and research has been devoted to large-scale testing as the centerpiece of state accountability systems. Their limitations and distal relation to classroom instruction as discussed earlier in this report have been well-documented. “The challenge in schools today is not just to get more students to pass tests, but to create new knowledge about how to improve instruction.”

There is considerable evidence that assessment for learning, or formative assessment, can have a powerful impact on student learning. A major review of more than 250 books and articles on the effects of classroom assessment was conducted a decade ago by Paul Black and Dylan William. The researchers concluded that using formative assessment that includes feedback to students can produce significant and substantial learning gains across grades and subject areas.

In fact, the analyses and reported gains were definitive—proving to be among the largest found for any
educational intervention—about \(0.4\) to \(0.7\) standard deviation units. Regular use of formative assessment yields gains of about 15 to 25 percentile points, or two to four grade equivalents on standardized achievement test scores scales. Stiggins reported even greater achievement gains of one-half to two standard deviations on high-stakes tests following regular formative assessment practices that provided students with the opportunity to take control of their own learning and achievement, with the largest gains made by low achievers. Other researchers found that if such gains by low achievers were applied to how well students perform on international assessments, the United States’ rank would improve from the mid-range of the 42 nations tested to the top five.

Given these positive effects, education leaders and measurement experts have expressed frustration with the current focus on large-scale assessments and emphasize the imperative of paying attention to what works. Research shows that well-designed formative assessment achieves its impact by combining the central elements of effective learning environments. These include:

- Powerful content organized around big ideas;
- The use of a learner-centered approach;
- Assessment-centered elements that help make thinking visible to teachers; and
- The creation of learning communities among teachers and students within classrooms and schools.

These elements are essential to support ongoing processes that deepen students’ conceptual development and ensure their steady progress toward competency levels within and across subject areas. Through curriculum-embedded assessment, students engage in integrated cycles of teaching, learning, assessment, and feedback. “When assessment is combined with other strategies,” Pellegrino asserts, “it can have a substantial impact. But for this to happen, it must be more valid and informative than is currently the case….In essence, it must become an essential part of the design and enactment of contemporary learning environments.”

In order to diagnose students’ needs, support their work, and help them monitor their own progress, teachers need more than a once-a-year summative standardized test to guide day-to-day instructional decisions. Teachers need ongoing, embedded curriculum assessment that provides teachers and students with the kinds of information they need to make decisions that promote continued learning. Formative assessment occurs while the learning is still happening, sometimes so early on in the learning process that students’ scores will not be high. This should not be couched as failure—it simply represents a point where students are along the pathway to ultimate success.

One caution for policymakers is that in order to generate the kind of impact documented in research studies, teachers must be able to use assessments to their best advantage. Yet researchers stress the consistent finding that teachers lack understanding of the features of high-quality formative assessment and receive little training in how to use it to improve student learning. Classroom instruction requires substantial transformation in how students are encouraged to express their ideas and how teachers provide feedback to them that can enhance their skills in managing and guiding their own learning. Fundamentally changing instructional practice will require students and teachers to jointly set goals for what to learn next and understand the need to scaffold content and performance requirements towards greater attainment.

To this end, research shows that using formative assessment augments the role of both teachers and learners in structuring learning processes and providing feedback to students on how to improve by:

- Providing a clear view of the learning goals;
• Providing information to the teacher about the state of the learner; and

• Requiring action to provide feedback to the learner and adjust instruction towards continuous growth and improvement.133

B. What about Interim or Benchmark Assessments?

Using their large banks of multiple-choice items, publishers of standardized tests are repackaging summative assessments for interim or “formative” purposes. While these benchmark tests can be used to identify students not making progress toward standards as measured on the summative test, this is not the type of powerful assessment practice described by Black and Wiliam and other researchers. These interim assessments often mirror the format of summative measures and as such do not provide the detailed, fine-grain picture that can help the teacher and student adjust instruction in relation to clear criteria for quality student work. Feedback that is limited to right and wrong responses fails to inform the teacher or the learner in how to make improvements. Furthermore, the consumers of interim assessments are teachers, not the students themselves, nor do interim tests provide focused guidance specific to the learning objective.

In contrast, formative assessments show students how to improve the quality of their work one feature at a time, as well as helping students evaluate their own work in order to see where they are now and where they need to be.

C. Feedback and the Role of the Learner

The question posed by Lori Shepard a decade ago remains apt as states consider the critical elements of their assessment systems: “How should what we do in classrooms be changed so that students and teachers look to assessment as a source of insight and help instead of its being the occasion for meting out reward and punishments?”134

Part of the answer to that question is, “provide useful feedback.” A central attribute of formative assessment is that it offers students timely feedback on their performance, which in turn increases their active engagement in their own learning and expands their understanding and motivation to improve. A great deal of research substantiates the importance of specific feedback that provides students with information that can improve the quality of and sense of ownership in their work. It’s also important to understand that how the feedback is conveyed to students significantly affects how they respond. Not all feedback is helpful—just giving grades, for example, can be counterproductive.

But the evidence is clear that feedback improves learning when it is designed to provide students with immediate information about particular attributes of their work and specific guidance on how to make improvements. “The evidence is clear that feedback improves learning when it is designed to provide students with immediate information about particular attributes of their work and specific guidance on how to make improvements.”

Moreover, as Stiggins points out, a school-centered versus a learner-centered culture that focuses on what’s taught rather than what’s learned has had profound implications for students’ decisions about their own learning trajectory. Their perceptions of their abilities have been shaped by assessment practices that largely use final exams and grades to ascribe their capability as a learner. Students have made decisions about whether success is within or beyond reach, whether the learning is worth the required effort, and whether dropping out is the better option on the basis of how their performance is judged.135
The use of assessment for learning highlights the role of the learner and his or her responsibility to use assessment information to make progress toward learning goals. In comparison to assessments of learning that focus on informing others about students, assessment for learning focuses on informing students about themselves, producing results unique to individual students, providing continuous feedback, enhancing teachers’ role in promoting success, and empowering students to monitor and improve their own learning.

The implications suggest that states must envision how to create powerful, equitable learning environments in which students participate as part of a community for purposes of inquiry. Students should be taught to ask questions about their own work and revise their learning as a result of feedback and reflection. As a result, they are empowered to improve their learning and the learning of their peers based on shared goals and understanding of the criteria for attaining them.

**D. The Role of the Teacher**

As schools are asked to raise achievement levels for all students, retraining teachers and establishing school and district cultures that support new forms of teaching, student engagement, and assessment for learning will be essential. Teachers and school leaders must believe that all students can achieve a level of academic success and competency and bring students to believe in this as well. In order to do this, teachers must receive highly-quality professional development and support to accommodate the fact that students learn at different rates. Teachers also need to develop skills in adjusting learning and providing students with ongoing evidence that they are making progress.

Through professional learning communities, teachers will need to examine their assumptions about instruction and its impact on student learning and become more intentional about engaging students based on their knowledge about how students learn.

Teachers need to be empowered to alter their pedagogical methods in ways that actively engage students in their own learning. For example, they need to alter the wait time in eliciting responses from students, engage students using work with peers in pairs or groups to solicit discussion, ask students to explain their reasoning in order to understand possible misconceptions, use information technologies to enhance students’ problem-solving skills, and frame high-quality questions to evoke important aspects of students’ thinking and learning. They empower students by co-constructing criteria by which students can assess themselves and their peers in order to identify their strengths and areas where improvement is needed. Teachers require a range of competencies in using student self-evaluation, student peer reviews and collaboration, involving students in their own parent/teacher meetings, using technology to tap students’ cognitive competencies, and developing reporting systems that are designed to provide feedback to individual students concerning their own progress and achievement.

In summary, investing in teacher development is critical so that teachers can:

- Turn broad content standards into classroom level targets, set clear learning goals, and define the criteria for success;
- Frame appropriate learning tasks that chart the learner’s progression toward competency within a given domain;
- Use assessments and other strategies to induce student engagement and provide students with feedback;
- Interpret, summarize, and record assessment information and use results correctly depending on the design and purposes of assessments;

**Study Group Takeaway**

States must have clear expectations that teachers and administrators can effectively use assessment for learning and ensure that educators gain essential assessment literacy skills through licensure, preparation programs, evaluation, and ongoing professional development.
“The premise that educators know what to do and all they need are the correct incentives to do it is essentially wrong. There is simply no way to solve the problem of large-scale improvement in educational performance without connecting policy and practice more directly and powerfully. It is this connection that we have been avoiding in a variety of ways, through a variety of pretexts, throughout the twentieth century, but especially since A Nation at Risk. Schools simply cannot do what they are being asked to do without more explicit and powerful guidance and support for instructional practice and without major changes in investments in knowledge and skill for educational practitioners.”

—Richard Elmore, School Reform from the Inside Out: Policy, Practice, and Performance

- Communicate results correctly; and
- Involve students in setting targets and assessing the goals.²³⁷

(See text box on middle school science vignette, page 43.)

Despite the importance of ensuring that educators have the skills needed to meet current demands for preparing all students for college and careers, there is a well-documented absence of preservice and ongoing training in assessment literacy skills. Moreover, traditional structures for professional learning fall well short of the training needed to build teacher capacity. Research shows that few teachers have access to high-quality, intensive professional development: more than half (57 percent) of U.S. teachers responding to the Schools and Staffing Survey (SASS) said they had received no more than 16 hours of professional development in the previous 12 months in their content area. More than two-thirds of teachers nationally reported that they had not even had one day of training in supporting the learning of special education or limited English proficiency students during the previous three years.²³⁸

E. The Role of the District and School Leaders

District and school administrative support is essential to create a school culture and the conditions conducive to integrating effective assessment and teaching practices in classrooms. District and school leaders will need to understand the role and effective use of different types of assessment, then collaborate with teachers to create a professional learning environment conducive to strengthening classroom assessment and teaching. They must attend to building assessment literacy of all staff and provide organizational structures and time for professional development and collaboration. Teachers need to collect student data and work and have opportunities to share the information regularly with their colleagues to determine whether or not their teaching strategies are working and whether students are making sufficient progress. Principals play a central role in ensuring there are regular, focused discussions about formative assessment and its use to improve instruction and student learning.

The ultimate impact of formative assessments on student learning will depend on a confluence of multiple factors acting in tandem within schools and districts that include:

- Monitoring the learning of each student on a timely basis and providing feedback on their progress;
- Facilitating department-wide and building-wide collaboration;
- Reshifting school culture to focus not on what’s taught, but on what’s learned by every student;
- Creating a professional learning culture to co-construct effective classroom assessment and instructional practice;
• Contributing to the development of supportive school policies;

• Providing information essential to an effective system of academic intervention;

• Informing the practice of individual teachers and teams;

• Engaging and motivating students as central actors in learning knowledge and new skills; and

• Fueling a continuous improvement process as part of school culture.\textsuperscript{140}

Yet, few district and school administrators are well-prepared for the current demands to lead organizational change around a powerful culture of teaching and learning. They must be able to differentiate between sound and unsound practices. Consistent with the current status of many teacher education programs, leader development systems are not providing superintendents or principals with the competencies to drive school-wide improvements in assessment, curriculum, and instruction. Studies show that the training principals receive across the nation leaves the majority ill-equipped for the job of promoting powerful teaching and learning, particularly with those students who need it the most. A national study of 31 preparation programs by Hess and Kelly found a critical lack of emphasis on results-oriented management and accountability, hiring quality teachers, making personnel decisions on the basis of performance, or using data or technology to manage school improvement.\textsuperscript{141}

**F. Role of the State**

As the bar has risen for all students to achieve at higher levels, states will need to consider how to scale effective assessment and teaching practices that result in school- and district-wide improvement. For their part, policymakers will need to help cultivate a culture of shared responsibility for student learning among individuals throughout the system, promote the use of frequent assessment to personalize instruction, and nurture a professional, collaborative teaching culture. Not that this is easy: scaling up what works and leading systemic change are notoriously difficult because they require changing the behavior of all individuals at every level of the system. The implications are that states will need to pilot new comprehensive approaches to assessment, accountability, and educator development that include changes in the way states work with universities, technical assistance partners, districts, and schools.

Formative assessment must be a major component of this change process. Ultimately, improving school performance requires significant improvements at the classroom level in the quality of ongoing assessment, instructional practice, and the level of student engagement.

To that end, education and policy leaders will need to create balanced assessment systems that focus on improving teaching and learning at the classroom level. States must have hard conversations about what’s at stake, how to leverage improvements, how build sufficient capacity, and then create models that help define what successful integration of assessment, curricula, and teaching will look like. In the long run, building the capacity for teachers and students to engage around equitable powerful learning in response to the level of student understanding rests with the caliber of people in the system. In order for an assessment system to support the goal of preparing students for college and careers, states must work explicitly toward ensuring that teachers have the preparation, training, and tools to assess accurately and to use the results to guide student understanding and learning. Despite decades of standards-based reform, improving educators’ assessment literacy has been long neglected, particularly with respect to how to deeply integrate assessments with good instruction.

“State standards for teacher and leader competencies should include requirements on the proper integration of learning and assessment that reflects the knowledge base on how students learn a content area and how assessment tools and practices can capture information about their competency.”
Programs for preparing educators, for example, continue to be driven by what providers want to offer—not by what schools or staff need; and licensure remains poorly connected to how well educators impact student learning. State standards for teacher and leader competencies should include requirements on the proper integration of learning and assessment that reflects the knowledge base on how students learn a content area and how assessment tools and practices can capture information about their competency. Improving teacher education will require states to revisit licensure, program approval requirements for teacher and leader preparation programs, and guidelines for professional development for working educators.

Moreover, as states act to leverage improvements in teaching, evaluations of teachers and principals should include information about how assessment for learning is effectively integrated in classroom instruction. Such performance appraisal systems would serve not as an isolated tool administered at a single point in time, but as a way to assess and respond to information about how teachers are gauging their students learning and how well they are responding to the individual learner’s performance.

Training that solely revolves around interpreting the results of large-scale assessments is not adequate nor targeted at the competencies needed to effectively craft, use, and interpret classroom formative assessment. States must be clear about expectations for using such assessment for learning and require those responsible for educator preparation and training to provide coursework and clinical training in school settings as part of program approval requirements. States must also ensure that the national faculty responsible for training teachers and leaders has the requisite training in the fundamentals of effective classroom assessment. States such as West Virginia, Delaware, Ohio, Illinois, South Carolina, and Kentucky are beginning to focus much more intently on providing ongoing in-service programs to help teachers improve the design and use of assessments for learning.¹⁴²

States must also partner with districts to develop models of learning that closer connect curriculum, instruction, formative assessment, and educator development by:

- Co-constructing guidance to support professional learning of classroom assessment practices and strategies;
- Developing tools and platforms that allow teachers to manage and use “just-in-time” results to improve individual student instruction;
- Deconstructing state standards to provide a curriculum map (learning progression) that is articulated within and across grade levels;
- Providing information management systems to meet the needs of instructional decision-makers, including students and parents;
- Using student friendly learner systems that allow students to manage their own learning; and
- Developing longitudinal data-management systems to measure student progress over time and so educators and parents can project whether a student is on the path to proficiency and is meeting important benchmarks for college- and career-readiness¹⁴³

- Increasing investments in research and development of formative assessments that can serve as prototypes for gathering and analyzing student work, providing effective feedback, and adjusting instruction;
Formative Assessment: An Example of Classroom Practice

The power and utility of formative assessment to improve instruction and learning outcomes rest with the quality and nature of the feedback it provides. A number of attributes can contribute to its overall effectiveness:

- The degree to which sub-goals are clearly articulated as part of a learning progression;
- Clear communication of specified goals and criteria for success;
- Feedback that is evidence-based and linked to instructional outcomes;
- Self and peer assessment that are incorporated to help students monitor their progress; and
- Conducive classroom culture that serves as a community of learners.144

Developed by Wylie from ETS, the following vignette is an example of formative assessment practice developed and implemented as part of a middle school science class:

After looking at her students’ lab reports, an 8th grade science teacher concluded that her class was struggling with the content of weekly science inquiry experiments. She structured the lab experiences around the “big ideas” presented throughout the science unit. In order to strengthen the connections that students made in framing their weekly lab reports and the key science concepts, she decided to revise the “criteria-for-evaluation form” she used to outline her expectations for the students’ write-ups.

The teacher decided to work directly with her students in constructing the revised version in order to increase their understanding of how to apply the criteria to their own reports. Using four student lab reports from the year before, she reviewed the criteria-for-evaluation form, asked students to work in small groups to rank the four reports according to the criteria and then justify their rankings. A member of each group was randomly chosen to explain his or her evaluation of one of the reviewed reports.

Following additional discussion and questions from students, the teacher reviewed how she used the criteria to grade the reports, emphasizing the purposes of the experiments and lab reports and focusing on how students should connect the experimental results and the big ideas of the course. The teacher asked students working in small groups to draft their own checklists, first by reviewing the two higher scoring reports against the criteria-for-evaluation form and second by articulating central features of the reports in their own words. The lists of important features were compiled in a subsequent class period resulting in the final student-developed criteria list. Throughout the co-construction process, the teacher emphasized the links between the central ideas and the evaluation criteria.

On subsequent lab reports, students were instructed to use the final criteria list to develop and evaluate their lab reports. Each week for about 15 minutes, students exchanged lab reports with another student and provided feedback using a 2 + 2 strategy (two positives and two suggestions for improvements). Students had the opportunity to make revisions based on the peer feedback before submitting their reports.

The results showed that the quality of the lab reports improved significantly and reflected students’ expanded understanding of key concepts within the science unit. The teacher could focus on students’ understanding of the “big ideas” and use the information to adjust instruction. This vignette highlights a number of key attributes of formative assessment, including structuring inquiry experiments as part of a learning progression that accentuates big ideas, collaborative learning, clear communication of descriptive feedback, and peer and self assessment to help students monitor the quality of their work.

The current environment affords states a critical window of opportunity to produce a new generation of assessments that addresses the longstanding issues regarding disparities in education and the demand for a well-educated workforce for the 21st century. Through their research, discussions, and experiences in their home states, the Study Group members concluded that it will be absolutely essential for states in short order to achieve two goals:

1) improving the rigor and relevance of standards to meet what is needed in today’s knowledge-based economy and

2) creating high-quality assessments that afford substantial improvements in curriculum, instruction, and the ability of teachers to address individual learning needs.

States will need to engage in honest and open discussions about the level of commitment needed to retool current assessment and accountability systems. This work will require creating greater integration and coherence in curriculum, instruction, and assessment; fostering a culture of learner-centered policy development and decision-making; expanding expertise and capacity through strategic use of networks, regional centers, and technology; crafting prototypes of effective approaches to integrating assessment and learning; using continuous evaluation to scale up what works; and building a human capital system to ensure high quality assessment and instruction.

Finally, in summing up the imperative for a balanced assessment and accountability system, we could not put it better than assessment expert Samuel Meisels and colleagues:

“When accountability is seen as a system that incorporates both instructional assessment and on-demand tests, both teaching and learning can be affected positively. Moreover, this methodology provides policymakers with clear documentation not only of summative accomplishments, but also of the process of teaching and learning. The approach described…places emphasis where it belongs: on teaching and learning, rather than on testing. And it does so without sacrificing either the student or the teacher on the altar of accountability.”145
VI. Study Group Recommendations

Assessment Systems

• States need to define 21st century learning in terms of fewer, clearer, and higher standards that identify what students need to know and do when they graduate from high school. Standards should specify grade-level expectations based on valid, reliable research for predicting valued performances that are internationally benchmarked and should be grounded on the latest advances in the sciences of thinking and learning.

• Assessment systems must be designed to improve student learning. Recognizing that no single test serves all purposes, states need to create a comprehensive, balanced assessment system that includes both assessment of learning (reporting on what’s been learned) as well as assessments for learning (providing ongoing feedback to teachers and students as learning progresses). The assessments—summative, formative, interim—should function as a coherent system that uses a variety of approaches to integrate assessment as part of the fabric of classroom teaching.

• All assessments, whether for large-scale accountability or day-to-day instruction, should be based on current knowledge about the nature of learning and validated for their specific purposes. States should conduct frequent evaluations to ensure validity, reliability, and fairness of assessments and to determine their impact on teaching and student learning.

• Policymakers need to shift more attention to classroom-based assessments that permit a finer-grain analysis of student understanding through the use of a variety of performance-based tasks (e.g., open-ended responses, portfolios, technology-based items).

• States should ensure that teachers have the tools and training they need to strengthen the connection between assessment and instruction based on our knowledge of how students learn and how such learning can be measured.
• Assessment results should provide user-friendly, transparent information that clearly describes differences in learning in a subject area in order to communicate effectively about student performance. Results should be communicated to a range of users, including teachers, students, and parents, in ways that position teachers and students as central actors in using results to guide teaching and individual instruction and to engage students in their own learning.

• States must develop appropriate assessments and accommodations for special education students and English language learners through extensive research and testing to ensure they are of high technical quality (e.g., valid, reliable, and aligned to standards). They should provide for a range of options (e.g., emphasis on universal design, the development of high-quality accommodation policies, and provision of alternate assessments) that adhere to professional testing standards and support high achievement levels.

• States should participate in national and state-level international assessments such as NAEP, PISA, and TIMSS in order to examine student attainment in an international context and thereby ensure that students are receiving an education that prepares them for the 21st century global economy.

• States must take advantage of the enormous possibilities offered through technology and its applications to integrate assessment and classroom teaching toward specific learning goals. Technology can contribute to powerful learning environments by embedding well-designed formative assessment strategies using highly engaging and innovative approaches consistent with how students learn.

Accountability

• Policymakers at all levels should support assessment models that measure the growth or progress of students in key subject areas and that provide statistically rigorous, multi-level analyses of the education system and its impact on students’ progress over time.

• State accountability should 1) focus on how the system (including school, district, and state levels) performs in a number of key areas and 2) make use of multiple indicators, of which summative assessment is only one. States should collect qualitative and quantitative measures, including student growth over time across the entire achievement continuum, as well as other indicators of school progress. The accountability index or composite should include long-term data that measure whether or not students have been effectively prepared for college or the workplace, including graduation data, college or workplace entry, and college completion.

• States should develop longitudinal data-management systems so student progress can be measured over time, and so educators and parents can project whether a student is on a path to proficiency, college readiness, and other important benchmarks along a PK-20 continuum. States should include elements to:
  • match teacher and student data;
  • hold preparation programs accountable for the performance of educators they prepare and license;
• provide data on students’ successful transition from secondary school to postsecondary education;
• identify factors that correlate to students’ ability to successfully engage in and complete postsecondary-level general education coursework; and
• inform education policy and practice in order to better align state academic content standards and curricula with the demands of postsecondary education and the 21st century workforce.

Preparation and Professional Development

• To ensure that assessment systems achieve their purposes, states must establish standards for teacher and leader competencies regarding their knowledge and skills of how students learn, how learning can be assessed, and how these two must be closely integrated to guide classroom assessment and instruction.

• States must establish consistent teacher development standards that position assessment literacy as a major component for teacher licensure, accreditation for preparation programs, and teacher evaluations. States must also ensure that the national faculty responsible for training teachers and leaders throughout the United States has the requisite training in the fundamentals of effective classroom assessment.

• States must ensure that at all levels of the system—classroom, school, and district—educators are provided with ongoing, high-quality professional development, along with the guidance, tools, infrastructure, and technology, to improve educators’ assessment literacy and their use of multiple assessments to measure students’ progress and respond to individual learning needs.
Appendix A. Glossary of Assessment Terms

**Accommodations**: Alterations made in the administration, not the content, of an assessment in order to “level the playing field” for a student with a disability. Accommodations can include providing a student with special equipment (such as an amplifier), producing the assessment in a different format (such as Braille), administering the assessment in a different setting (such as a hospital), altering the test schedule (such as giving extended time), or providing alternate ways to respond (such as the use of sign language).

**Alignment**: Process used to ensure that the knowledge and skills asked for on assessments is the same knowledge and skills specified in the content standards.

**Alternate assessment**: Tool used to seek different information by measuring different skills than a regular assessment. Alternate assessments are commonly administered to students with disabilities who are unable to access a standard curriculum, such as students with severe cognitive disabilities.

**American Diploma Project**: Network convened by Achieve, Inc. committed to raising state standards, graduation requirements, assessments and accountability systems and ensuring academic content and expectations are sufficiently rigorous, relevant, and aligned for high school graduates to be prepared for the demands of college, career, and life. Thirty-four states—responsible for educating 41 million students, nearly 85 percent of all public school students in the United States—are currently part of ADP.

**Benchmark**: A subcomponent of a standard that more specifically defines what students should know and be able to do at a particular point in their educational career.

**College- and career-readiness**: Student access to and mastery of the course work and teaching needed to perform without remediation in entry-level, credit-bearing courses at two- and four-year programs. Research shows that the targets of preparedness for postsecondary success must be at the same level of sophistication whether students are in pursuit of an associate’s degree, bachelor’s degree, technical certification, or immediate entry into a career. The convergence of aptitudes and experiences needed for post K-12 success be it in college or career is attributed in part to the shifting nature of the economy and the depreciating availability of low-skilled, blue-collar jobs.

**Common Core Standards Initiative**: A state-led process currently including 50 states and territories committed to developing research and evidence-based, internationally benchmarked K-12 standards for English language arts and mathematics that are aligned with college and work expectations and skills. The effort is convened by the National Governors Association Center for Best Practices and the Council of Chief State School Officers in partnership with Achieve, Inc., ACT, and the College Board.

**Comprehensive exams**: Assessments aligned with state standards and generally targeted at the 9th or 10th grade level.

**Content standard**: What students should know or be able to do in a specific subject area. *Benchmarks* define the point in a student’s career (usually a grade level) when he or she should have acquired such knowledge or skill.

**Criterion-referenced test**: Test that compares student performance to clearly defined curricular objectives, standards. Assessment results are usually reported
as a pre-defined level of performance or a numerical score.

**End-of-course exam:** Assessment used to gauge whether students have mastered the content of specific high school courses; these exams are usually standards-based, and students take each test after completing a specific course. In some cases, results have implications for graduation.

**Elementary and Secondary Education Act (ESEA):** U.S. federal statute enacted in 1965 and most recently reauthorized as the No Child Left Behind Act of 2001. ESEA is responsible for primary and secondary education funding, including professional development, instructional materials and resources to support educational programs, parental involvement promotion, and allocations for low-income students (defined in Title I as 40 percent or more of school population). The next anticipated reauthorization of ESEA is in 2010.

**Exit exam:** Test a student must pass to graduate. Twenty-six states have or are in the process of implementing exit exams by 2012. Those in favor of exit exams cite the importance of ensuring that a conferred diploma signals attainment of basic academic proficiency and job readiness. Those opposing exit exams bring equity issues to the fore and highlight the disproportionate impact such tests have on minority students and increased drop-out rates.

**Formative assessment:** Activities and feedback process between teacher and student used to guide day-to-day teaching and learning processes. It may include observations, homework, student writing, quizzes, or project-based work. Formative assessment differs from evaluative assessment in that its process from development to implementation by teachers is embedded in instruction. The detection of gaps or misunderstandings in student learning creates opportunity for adjustments to ongoing teaching and learning to improve students’ achievement of intended outcomes.

**Interim assessment:** Test typically administered with less frequency than formative assessment and more frequency than summative assessment to evaluate students’ knowledge and skills within a particular timeframe of instruction. Interim assessment results may be used for instructional, evaluative, or predictive purposes and tend to be administered by districts and schools rather than teachers.

**Minimum competency exams:** Assessments generally focused on measuring basic skills below the high school level.

**Monitoring:** A non-testing tool used for accountability. In England, for example, in addition to traditional student achievement assessments, school inspections are used as part of the country’s accountability system to provide a broader picture of program implementation, the quality of academic instruction, and school culture, when auditing schools’ strengths and weaknesses. As a result, accountability has both evaluative and supportive functions. Peer evaluations and self evaluations are also mechanisms for monitoring.

**Multiple-choice assessment:** Assessment that requires students to select their responses from among a set of specific, pre-determined choices. When carefully designed, multiple-choice assessments can provide reliable information about what students know. Usually, multiple-choice assessments are scored by computers, which provide quick, cost-efficient, accurate, and impartial results. Although multiple-choice assessments yield very reliable scores, their validity for more complex skill sets may be low.

**National Assessment of Educational Progress (NAEP):** (also called “The Nation’s Report Card”) Test created 20 years ago by Congress and administered by the National Center of Education Statistics, U.S. Department of Education. This test is the only ongoing nationally representative measure of what students in the United States know and can do in 10 subjects and three grade levels (4th, 8th and most recently 12th grade). NAEP is governed by the 26-member bipartisan National Assessment Governing Board (NAGB), with a chair appointed by the Secretary of Education.

**Norm-referenced test:** Test that compares individual performance against the performance of a representative national sample. Assessment results are usually reported according to “percentile” rank (a student’s percentile rank on a test is the percentage of students he or she outscores).
Performance standard: Definition of “how good is good enough” in terms of achieving a content standard. Frequently, this division comes by way of categorical levels such as “advanced,” “proficient,” “basic,” and “below basic.”

Performance assessment: Assessment that requires students to formulate an original response to a question and to communicate that response through the performance of some act. For example, a performance assessment may ask a student to produce a written essay, model, diagram, or persuasive speech. Performance assessments are usually criterion referenced. Educational reforms and standards that emphasize application of knowledge and expression of ideas are most effectively evaluated with performance assessments.

Portfolios: A type of performance assessment. Portfolios are collections of student work that demonstrate progress over time and level of performance.

Professional Learning Community (PLC): An organizational structure—often integrated as a component of a development or improvement strategy—to promote continuous learning and collaboration among school staff. PLCs may be school-, district-, or virtually-based and involve investing in the peer-to-peer exchanges among educators to promote school culture and student achievement. Activities including data analysis, mentorship, and shared age- or subject-group planning may require reprioritizing and restructuring the school day.

Program for International Student Assessment (PISA): An internationally comparative measure of student competencies in reading, mathematics, and science administered as students approach graduation. In contrast to other international tests that focus on school-based curricular achievement, PISA includes performance-based tasks and focuses on domain literacies and students’ mastery of processes, understanding of concepts, and application of knowledge in real-world contexts. PISA is sponsored by the Organisation for Economic Cooperation and Development.

Progress in International Reading Literacy Study (PIRLS): An internationally comparative measure of student reading achievement, behaviors, and attitudes administered to fourth grade students. The test content is collaboratively developed by international reading experts and taps into student performance as well as reading curricula and classroom practices across nations. It was first administered in 2001, then again in 2006, and will be administered in 2011.

Reliability: The extent to which a student would receive a similar result if a test were administered again. Reliability is discussed in terms of “consistency” (e.g., will a student’s score on a test taken one day be close to his or her score on the test if it is taken the next day) and “generalizability” (e.g., does a student’s score on a 60-minute math test accurately reflect his or her knowledge of the entire subject).

Rubric: Guide used to score performance items that describe the characteristics or components of responses to questions earning various point values in a standardized fashion. For example, a rubric would describe what an answer should contain to earn the student one point, two points, three points, and so forth.

Sampling: Giving a test to a randomly chosen subset of students in order to provide a statistically equivalent picture of how all students in that group would perform. On some assessments, each randomly chosen student takes only a portion of the test. This technique, called “matrix sampling,” is an efficient way to provide reliable group (e.g., district, state, nation) results.

Summative assessment: Measure of student competency attainment after completion of a particular segment of study. Summative assessments include end-of-unit tests, final course grades, and state- or district-wide large-scale tests.

Trends in International Mathematics and Science Study (TIMSS): An internationally comparative measure of student achievement in mathematics and science administered to U.S. students in 4th and 8th grade. TIMSS content focuses on curricular coverage.
for each age range and also includes demographic data collected through student, teacher, and school questionnaires. It was first administered in 1995, then again in four-year intervals to more than 40 nations worldwide.

**Validity:** The degree to which a test actually measures what it purports to measure (e.g., the extent to which an I.Q. test actually measures intelligence); also, the degree to which particular uses and interpretations of assessment results are justified (e.g., the extent to which a certain score on an SAT or ACT test predicts a student’s chances of succeeding in college).
Appendix B. State Actions in Assessment and Accountability

Alabama
The state board-endorsed proposal would include paying for all students to take the ACT college entrance exam and also recommends doing away with five-section Alabama High School Graduation Exam and instead requiring end-of-course tests in the same subjects as well as four different ACT college- and work-ready assessments administered during 8th to 12th grade.146

Arkansas, Colorado, Illinois, Kentucky, and Maine, Michigan, Wyoming
These states administer national college entrance examinations to all students—not just those categorized as college-bound—as part of their state assessment system.147 The states do not require a specific score to graduate.148

California
California adopted a “multiple pathways” approach to high school that combines a college-prep academic instruction with a technical curriculum and work-based learning experiences. The college- and career- curriculum is intended to have two-fold benefits for the state economy and individual well being: reduced incidents of dropout and improved match between graduates’ skills and local job needs.149

Colorado
In August 2009 Colorado launched the Colorado Growth Model, an academic data system that links students’ scores from the Colorado Student Assessment Program across multiple years and reports results at the school, district, and state level. The data is available for public access at www.schoolview.org. Colorado also adopted new 12th grade standards in July 2009 as part of the state’s three-year initiative—Colorado Achievement Plan for Kids (CAP4-K).

CAP4-K charged the state board of education and higher education commission with developing new standards aligned with the business world, vocational-tech schools, and universities. The next component of the plan rollout is course-by-course, grade-by-grade standards, which are expected to be ready for release by the end of the year. Also being considered in the state: creating new assessments aligned to the college- and workforce-ready content standards, changing diplomas to make students eligible for an “honors diploma” that signifies completion of rigorous college-prep courses, and potentially dropping the Colorado Student Assessment Program in favor of end-of-course exams.150

Florida
Florida’s assessment system allows students who fail the high-stakes 10th-grade Florida Comprehensive Assessment Test three times to still earn a high school diploma if they score high enough on the ACT.151

Kentucky
In February 2009, state education leaders released a position paper from the Kentucky Department of Education and Kentucky Board of Education that outlines key principles in the design of the state’s next generation of assessments and accountability. The annual state assessment should measure both the knowledge and higher-level thinking required by standards; provide diagnostic, longitudinal growth data and overall
proficiency at the individual student level; and include interim and classroom assessments that support the annual assessment.152

**Massachusetts**

Building on the longitudinal data model developed in Colorado, Massachusetts is creating a state data system that allows for student-level growth analysis. Tracking the performance of individual students annually based on the statewide assessment in English and math will allow the state to predict the likelihood that students will meet expectations or show improvement. This system is potentially powerful because it provides greater accountability at the school level where, in the past, results from small classes easily skewed by a handful of students have caused many instructional leaders to de-emphasize or discount grade-to-grade score comparisons.153

**Montana**

Montana is building awareness and investment in the state around formative assessment and educator training. The Office of Public Instruction (OPI) acknowledged that a new assessment system will require supporting educators on the ground. The Montana OPI asserted its commitment to provide professional development, learning communities, and expert support and consultation, and underscored the need for all of these investments to be ongoing and available to teachers in remote regions. Webinars are one strategy the state plans to use to engage with educators in rural areas and to facilitate professional conversations around best practices among educators in different areas of the state.154

**New Jersey**

New Jersey students will now complete individual learning plans with personal, academic, and career goals beginning in middle school and continuing until graduation as part a new plan approved by the state board of education in June 2009. The Secondary Education Redesign plan, also increases the number of credits needed to receive a diploma, adds more rigorous math and science requirements including a half-year financial literacy course, and requires students to take an additional seven exams in different subjects throughout their high school career. As part of the individual learning plan, New Jersey is also piloting a formalized mentorship program that pairs students with teachers, parents, and community members. New Jersey also plans to make changes to the high school graduation test and move toward a series of end-of-course tests.155

**North Carolina and West Virginia**

North Carolina is collaborating with West Virginia as part of the Partnership for 21st Century Skills and has revised its standards and creating assessment systems aligned to 21st century skills along with web-based tools to assist districts and schools in implementation.156 North Carolina created a Blue Ribbon Commission on Testing and Accountability report on design of comprehensive system that includes essential standards built on 21st century skills/content and formative, benchmark, and summative assessments. The North Carolina Graduation Project, a performance-based and rubric-scored element of the state’s new assessment system design slated to begin with the Class of 2010, has since been put on hold by state general assembly; at the earliest, it would go into effect for the Class of 2015.157

**Ohio**

Ohio officials are considering alternatives to standardized testing that may help students stay focused in their high school studies and better prepared for postsecondary opportunities. A statewide student group, Ohio Youth Voices, petitioned Governor Ted Strickland to consider alternatives to the five-part Ohio Graduation Test, which seniors must pass to graduate. State educational officials secured a $1.3 million grant to explore alternate assessments such as portfolios, senior projects, journals, and small group collaborations. A group of educators from across the state will select and field test a range of alternate assessments beginning this next school year. Earlier this year, Gov. Strickland proposed dropping the Ohio Graduation Test and replacing it with a four-part graduation requirement—ACT test, end-of-course exams, senior thesis, and community service project—aligned to college and career expectations.158

**Oklahoma**

The state is using college entrance exams as an alterna-
tive to passing other elements in its assessment system. Students who do not pass end-of-instruction tests required for graduation can take the corresponding subject section of the ACT to demonstrate mastery.\textsuperscript{159}

**Pennsylvania**

The Pennsylvania State Board of Education approved a series of high school graduation exams in core subjects by a 14-2 vote. The Keystone Exams would require students to take exams in core subject areas and would count for at least a third of their final course grades. The first exams would be given in 2010-11, with the class of 2015 being the first class to be required to pass the exams or an alternative assessment. Districts would be allowed to use their own exams if they are independently validated. Although the board approved the tests, the new rule must first pass through the House and Senate Education Committees and the Independent Regulatory Review Commission, which ensures the testing mandate would not conflict with other regulations or statutes. The state legislature could also block implementation with majority votes in both chambers.\textsuperscript{160}

**Rhode Island**

Rhode’s Island’s class of 2009 was the first in the nation to take performance-based assessment as a state requirement to earn a diploma. The requirement took effect in accord with a 2003 policy change by the State Board of Regents and guidance from the Rhode Island Department of Education. In addition to students’ class grades and scores on the New England Common Assessment Program (NECAP), seniors must choose and pass two of three possible state-mandated but locally-designed performance-based assessments: a portfolio work selected from their four years of high school, a senior project, and a comprehensive course assessment.\textsuperscript{161}

**Tennessee**

Tennessee is the most recent state to add a requirement for high school students to take the ACT as part of the state’s assessment system. As part of the Tennessee Diploma Project, a state-based version of a national initiative, the state is also swapping end-of-course tests in replacement of gateway tests, requiring more rigorous course completion, and charging teachers to “learn and implement new curriculum and instruction strategies”\textsuperscript{162}

**Texas**

Texas eliminated a mandate that 3rd graders must pass state tests to be promoted to 4th grade as well as a state quality ranking system that hinged on one year of data in an effort to put less weight on snapshot assessments. The state also added a new layer of college-readiness indicators to its graduation requirements, and requires districts only to meet 85 percent rather than 100 percent of indicators to address the diverse situations facing districts.\textsuperscript{163}

**Utah**

Based on recommendations from the Governor’s Blue Ribbon Panel on Assessment, Utah is piloting assessments that reposition student learning as the top priority. The adaptive tests are administered multiple times during the year via computer and change based on students’ responses to questions of varying skill level. The Panel recommendations also include dropping the Iowa Test of Basic Skills, the Utah Basic Skills Competency, and core end-of-level testing in all grades.\textsuperscript{164}

**Washington**

Washington State is replacing its high-stakes test with considerably shorter exams that will eventually be administered online. While 93 percent of students passed the Washington Assessment of Student Learning (WASL) exam in 2009, state policymakers acknowledge the results mask a dropout problem. The new testing system will include elementary- and middle-school tests called Measurements of Student Progress, and the high-school tests will be known as the High School Proficiency Exams (HPSE). The HPSE will still serve in part as a gateway—the class of 2010 will need to pass the reading and writing portions to graduate, and starting with the class of 2013, students will also be required to pass the math portion of the test to garner a diploma. However, the elementary- and middle-school tests will have a different purpose—they’re designed to give teachers and families a measure of how a student is progressing.\textsuperscript{165}

**Wisconsin**

State Superintendent Tony Evers announced that the Department of Public Instruction (DPI) will be phas-
ing out the 17-year-old Wisconsin Knowledge and Concepts Examination (WKCE) over the next two years, replacing it with new assessments that “balance the needs of students, teachers, and parents as well as providing public accountability for student learning.” The move is based on the recommendations of the 42-member Next Generation Assessment Task Force. According to the DPI, the new assessments will likely be computer-based with multiple opportunities to benchmark student progress during the school year, allowing for immediate and detailed information about student understanding and facilitating the teachers’ ability to adjust instruction. At the high school level, the WKCE will be replaced by assessments that provide more information on college and workforce readiness.¹⁶⁶
Appendix C. Bibliography


Alliance for Excellent Education. From No Child Left Behind to Every Child a Graduate. Washington, DC: Alliance for Excellent Education, 2008. www.all4ed.org/publication_material/reports/ECAG.


Endnotes


5. Ibid.

6. Alliance for Excellent Education, From No Child Left Behind to Every Child a Graduate.

7. ETS, America’s Perfect Storm.

8. Alliance for Excellent Education, From No Child Left Behind to Every Child a Graduate.

9. ETS, America’s Perfect Storm.


14. NGA, CCSSO, and Achieve, Benchmarking for Success.

15. See, for example, Daniel Koretz: “Clearly, caution must prevail in making comparisons with other countries based on spurious inferences about the root causes for how students perform solely on the basis of test scores. “Test scores describe what students know and can (or can’t do). And in most cases, that is all test scores do. Except in unusual circumstances, such as a planned experiment with random assignment of students, scores by themselves do not explain why students know what they do. For that, we need additional data, and those additional data are rarely available to the commentators who are so quick to tell us why students can or cannot perform as we wish.” Daniel Koretz, Measuring Up: What Educational Testing Really Tells Us (Cambridge, MA: Harvard University Press, May 2008), 114.


19. NGA, CCSSO, and Achieve, Benchmarking for Success, 7.

20. ETS, America’s Perfect Storm.

21. Alliance for Excellent Education, From No Child Left Behind to Every Child a Graduate.


26. NGA, CCSSO, and Achieve, Benchmarking for Success.


33. National Research Council, Knowing What Students Know.


35. National Research Council, Knowing What Students Know.


40. ACT, 2007 National Profile Report (Iowa City, IA: ACT, 2007), 20. www.act.org/news/data/07/data.html. The report provides data from EXPLORE and PLAN exams that are used to benchmark and predict college preparedness. ACT developed its College Readiness Benchmarks (sec. 3) that are minimum scores on the ACT English, Mathematics, Reading, and Science tests that reflect at least a 50 percent chance of achieving a B or higher grade, or at least a 75 percent chance of a C or higher grade, in entry-level, credit-bearing college English composition, college algebra, social science, and biology courses, respectively.

41. ACT, ACT College Readiness System: Meeting the Challenge of a Changing World (Iowa City, IA: ACT, 2008). ACT defines a Critical Core Curriculum as: four years of English, three years of mathematics, including rigorous courses in Algebra I, Geometry and Algebra II; three years of sciences including rigorous courses in biology, chemistry, and physics; and three years of social studies.


43. ACT and Education Trust, On Course for Success: A Close Look at Selected High School Courses That Prepare All Students for College (Iowa City, IA: ACT and Education Trust, 2005).


45. Alliance for Excellent Education, From No Child Left Behind to Every Child a Graduate, 15.

46. SchoolDataDirect.org; 2007, quoted in Chris Minnich and others, “Common Standards Panel” (presentation to the NASBE Assessment Systems for the 21st Century Learner Study Group, Arlington, VA, June 12, 2009); 49 states and territories signed a Memorandum of Understanding in June 2009, while Missouri and South Carolina joined later that summer.

47. CCSSO and NGA, “Forty-nine States and Territories Join Common Core State Standards Initiative” Press


50. Richard Elmore, Improving the Instructional Core (Boston, MA: Harvard University: June 2008).


52. Hamilton, Stecher, and Yuan, Standards-Based Reform in the United States.

53. Ibid.


56. National Research Council, Knowing What Students Know.


61. CEP, State High School Exit Exams.


64. National Research Council, Knowing What Students Know, 25.


71. National Research Council, Knowing What Students Know, 54.


78. Pellegrino, The Evolution of Educational Assessment.


80. National Research Council, Knowing What Students Know.


88. Darling-Hammond and Pechone, “Reframing Accountability.”

89. Ibid, 44.


92. National Research Council, Knowing What Students Know.


97. Wurtzel and others, “The Role of Interim Assessments in a Comprehensive Assessment System.”

98. National Research Council, Knowing What Students Know.


100. Darling-Hammond and Pechone, “Reframing Accountability.”

101. Ibid; National Research Council, Knowing What Students Know.


105. Technology-Based Assessments Improve Teaching and Learning (Glen
Public Policies: The Dartmouth/OECD Conference (Hanover, NH: Public Affairs Center, Dartmouth College, 1975), 35.


120. Broader Bolder Approach to Education Campaign, School Accountability: A Broader, Bolder Approach.

121. Ibid.


123. Ibid.


130. Chappuis, Chappuis, and Stiggins, “Formative Assessment and Assessment for Learning.”


133. National Research Council, Knowing What Students Know.


and Student Progress” (Upper Saddle River, NJ: Pearson Issue Papers, no. 2, 2009); Alliance for Excellent Education, Meaningful Measurement.  


139. Richard Elmore, and The School Redesign Network at States and Abroad Learning Profession: A Status Report others,  

142. Chappuis, Chappuis, and Stiggins, “Formative Assessment and Assessment for Learning.”  

144. Pearson, “Using Assessments to Improve Learning and Student Progress.”  

145. Meisels and others, “Creating a System of Accountability.”  


147. ACT, Inc., “Statewide Administration of the ACT: A Key Component in Improving Student Access to College and Work” (Iowa City, IA: ACT, 2009), www.act.org/research/policymakers/pdf/collegenews.pdf. After the test became a requirement in Colorado and Illinois in 2001, the number of high-school graduates who enrolled in college increased by eight percent in both states from 2002-2004 with markedly higher percent increases for minority groups.  


154. Sharon Carroll, Montana state board member, e-mail message to author, March 20, 2009.  


159. Dawn Marks, “Tests Won’t Be the Only Route to High School Graduation in Oklahoma,” The Oklahoman, April 1, 2009.  


