

The Hunt Institute's

BLUEPRINT

for Education Leadership

James B. Hunt, Jr.
Foundation Board Chairman

Judith A. Rizzo, Ed.D.
Executive Director and CEO

April D. White, Editor
Director of Communications
awhite@hunt-institute.org

Created by the University of North Carolina Board of Governors in 2001, the James B. Hunt, Jr. Institute for Educational Leadership and Policy works with leaders to secure America's future through quality education. Working at the intersection of policy and politics, the Hunt Institute connects leaders with best strategies for developing and implementing policies and programs to improve public education.



JAMES B. HUNT, JR. INSTITUTE
for EDUCATIONAL LEADERSHIP and POLICY
1000 Park Forty Plaza
Suite 280
Durham, NC 27713
p: 919.425.4160
f: 919.425.4175
www.hunt-institute.org

For additional copies of *Blueprint*,
or to receive a copy via e-mail,
please call 919.425.4167.

PREPARING TO IMPLEMENT NEXT GENERATION ASSESSMENTS

By Stephanie Dean, Director of Research and Evaluation

Why Assessment?

Assessment of student progress is a critical ingredient of teaching and learning. There are many types and forms of tests and assessment, and it is essential that each is used for the specific purpose for which it was created. State assessments are generally given at the end of the school year to generate data for accountability. They are not designed to provide the student-specific data that teachers need to determine appropriate instruction or that parents and students need to understand academic progress. A **comprehensive assessment system** would provide both kinds of information in an integrated way.

Why Now?

Two recent events have presented states with an unprecedented opportunity to design comprehensive state assessment systems:

1. Thirty-seven* states and the District of Columbia have adopted Common Core State Standards (CCSS) that clearly define what students need to know and be able to do in math and reading in order to graduate from high school career- and college-ready.
2. Two newly formed state consortia – the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC) – have been awarded Race to the Top grants to develop assessment systems that are aligned with the CCSS. Forty-four states and the District of Columbia are participating in one or both of these consortia and have committed to implementing the new assessments in the 2014-15 school year.

Earlier this year, the Hunt Institute commissioned the workshop series “**Best Practices for Student Assessment**,” convened by the National Research Council’s Board on Testing and Assessment to explore key issues for states to consider as they develop assessments aligned with the CCSS. The meetings brought together researchers, practitioners, and state and federal policy leaders who reviewed research findings and considered lessons learned from state attempts to establish innovative assessment tools. This issue of *Blueprint* reports on those findings and discussions, suggests what we can expect from the assessment consortia, and offers several ideas that states can utilize to prepare for the changes that will result from this exciting work.¹

*Minnesota has adopted the English Language Arts standards only.

CURRENT CONTEXT AND OPPORTUNITIES

Characteristics of Current Assessments

The summative assessments states currently use – end-of-year or end-of-course assessments that measure student achievement for accountability purposes – have evolved during the last few decades. This evolution has been shaped by many forces, including the requirements of the *No Child Left Behind Act* (NCLB), limitations of the assessment industry, and budget cuts and constraints. During the Hunt Institute-sponsored NRC workshop series, which included an array of individuals with experience developing and administering state assessment systems, participants reflected on key characteristics of current assessments. In order to appreciate the scale of changes that the state assessment consortia aim to generate in a very short timeframe, it is helpful to first consider several aspects of the current landscape.

- **Accountability testing occurs more frequently than in the past.** In response to the requirements of NCLB, states have increased the frequency of summative tests that are used for accountability. In accordance with federal requirements, end-of-year testing now occurs at least in grades 3-8 and one year of high school. Each student receives a score, and scores are disaggregated to monitor the performance of traditionally underperforming student groups. Such data, typically unavailable prior to NCLB, now provide what is perceived to be vital information for accountability purposes.
- **Test results lack clear meaning.** NCLB requires each state to determine its own cut scores – the test score that students must achieve to be designated “proficient.” According to studies by the National Center for Education Statistics and the Fordham Institute, a far greater percentage of students achieve proficiency on their state assessment than on the National Assessment of Educational Progress (NAEP)². This discrepancy is attributed to lower cut scores in the states and sends unclear messages about student achievement to state leaders, educators, students, and parents.
- **Multiple choice tests narrow instruction, but higher quality tests are much more expensive to design and score.** More frequent accountability testing has required tradeoffs in the way that assessments are designed and administered. Assessments are costly – particularly performance assessments that require students to demonstrate or apply complex skills. Budgetary constraints are a primary reason that states limit the use of performance assessment. As states have increasingly come to rely on less expensive, multiple choice summative tests, test preparation within the classroom has narrowed to focus on the low-level skills that these assessments tend to measure.
- **Many teachers struggle to interpret meaning from assessment data.** Teachers need timely, detailed data about each student’s progress, but also need to be prepared to translate that data into effective instructional strategies. NRC workshop participant Margaret Heritage of the National Center for Research on Evaluation, Standards and Student Testing (CRESST) noted that a number of studies have documented that teachers have difficulty using assessment information to plan instruction. Heritage noted that pre-service teacher education generally does not provide teachers with the professional knowledge that they need to accomplish this complicated work.
- **The capacity of state departments of education has been hit hard by budget cuts.** Reports by the Government Accountability Office (GAO) and Education Sector have documented a loss of relative capacity in state assessment leadership.³ Most states have seen at least a three-fold increase in the number of tests developed and administered, yet have seen nowhere near that increase in personnel. Except in larger states, the same personnel oversee and manage both assessments and the state’s accountability system. This situation will only worsen as state budgets diminish.
- **Pockets of skepticism may linger as a result of past efforts.** As state leaders set out to improve their assessment systems they might encounter skepticism based on previous failed attempts. Educators are accustomed to state assessments designed to generate accountability data only. As a result, some educators fear that plans to develop “comprehensive assessment systems” will ultimately generate more accountability testing and will not actually be designed to inform instructional efforts.

Comprehensive Assessment System Consortia

The Partnership for Assessment of Readiness for College and Careers (PARCC) includes 26 states and will develop end-of-year assessments and “through course” assessments administered via computer at designated intervals during the school year.

The Smarter Balanced Assessment Consortium (SBAC) includes 31 states and will develop end-of-year assessments that adapt to the test taker’s ability and performance tasks to be administered during the school year.

Both PARCC and SBAC will develop optional formative assessment tools and professional development resources that are aligned with the Common Core State Standards.

Opportunities for State Consortia to Change the Assessment Landscape

As noted by NRC workshop participants, assessment is an integral aspect of education. Whether it is done well or poorly, it greatly influences the actions of educators and students. There are several ways that the assessment consortia could dramatically change the assessment landscape in states.

State accountability testing occurs at the end of the school year, but assessment occurs throughout the year. Teachers find and develop their own assessment tools to analyze student progress and plan accordingly. Some states or districts also mandate periodic practice tests to predict later scores on the year-end state assessment. In a comprehensive assessment system, assessment tools will be carefully designed to generate valuable information while avoiding duplicative or meaningless testing. State assessment consortia are particularly seeking to **balance the need for accountability data with the need for data that inform instruction**. Each consortium has proposed a system that will give teachers and school administrators periodic data about student progress and teacher effectiveness, and at the end of the school year will provide policymakers the data they need for accountability purposes.

To create a system that strikes a balance between accountability and instruction, consortia will need to tap the knowledge of test developers *and* learning experts (cognitive psychologists). These two groups of experts historically have worked in separate spheres, developing assessment tools and curricular/instructional materials apart from one another. Experts from these fields bring their respective theories and professional criteria to the task of developing a comprehensive assessment system, and their competing ideas will also need to be reconciled with the expectations of policymakers and boots-on-the-ground educators. This will be complicated and intensive work – very challenging for any single state to take on amidst the current strain on budgets and human resource capacity. Working together, members of state consortia will create the space that is needed for leaders and experts to come to agreement about the goals for the new assessment system and can ensure that the necessary expertise and perspectives are included in the planning and design phases.

State assessment consortia also have the potential to **generate answers to lingering questions about the development of innovative assessments**. The ambitious plans of each consortium enter uncharted territory for the testing industry. Professional test developers are currently well-equipped to produce large-scale multiple choice items and some test items that require students to construct responses. However, NRC workshop participants noted several other types of assessment items for which there is not a strong enough knowledge base to develop effective templates. For example, the industry has not come to agreement

about specifications for effective interim assessments, which are a critical tool for providing teachers with immediate feedback on instruction and learning. Additionally, Stephen Lazer of Educational Testing Service noted that research on simulation-based test items is still in its infancy. These knowledge gaps have persisted despite growing interest in such assessments. By banding together in consortia, states will generate a collective demand and the resources for much-needed research and development.

Likewise in the area of technology-based assessment, research, development, and implementation have been prohibitively costly for states to undertake on a large scale. The state consortia aim to **use technology to improve assessment quality and effectiveness**, including greater efficiency in test production and delivery, adaptability to generate more precise measures of an individual student's learning, simulation activities to present “real life” test items, and artificial intelligence for scoring that gives teachers and students immediate feedback. State consortia will benefit from federal funds to develop new assessment technologies, and individual consortia members will also benefit from interaction with a network of states that will simultaneously tackle the challenges of implementing new technologies. Consortia member states such as Delaware, Hawaii, and Oregon have significant experience with implementation challenges in this area, and can share their deep understanding with consortia partners.

Innovative assessment will be more expensive than multiple choice-based tests, but members of state consortia also stand to **realize significant cost savings** in this regard. Ed Roeber and his colleagues at Michigan State University recently set out to measure the financial burden of moving to innovative assessments. The typical, multiple choice summative assessment costs around \$20 per student. Roeber found that a high-quality summative assessment that includes a combination of multiple choice items, constructed response items, and performance tasks would cost approximately \$55 per student if it were developed and implemented by a single state. However, if states join together in a consortium, economies of scale would reduce this cost by an average of \$15 per student. Roeber also estimated that additional cost savings of \$3-4 per student can be achieved with strategic use of technology, and an additional \$10-20 per student can be saved if teachers score the assessment. The actual per-student cost of consortia assessments is not yet known; SBAC and PARCC have estimated that ongoing costs could range anywhere from \$17 to \$50 per student. Roeber's study suggests that the ongoing costs of consortia assessments could be similar to current costs, but states would acquire a much improved product.

State assessment consortia will also **establish a consistent measure of student performance that is aligned with college and career readiness** – a cut score to be shared across all member states. To ensure that this achievement measure is truly aligned with the requirements of postsecondary demands, the consortia have secured commitments from institutions of higher education (IHEs) to assist in the development of high school summative assessments. Higher education has not traditionally been involved in this work, though member states of Achieve’s American Diploma Project have made progress in opening conversations about what it truly means to be ready for postsecondary success.⁴ The state assessment consortia will build on this work to ensure that K-12 and postsecondary expectations are aligned. Indicative of this commitment, IHEs that have signed on to participate in the state assessment consortia have promised that their respective schools or systems will use the resulting achievement standard as a mechanism to place incoming freshman into credit-bearing coursework.

In many states, the move to align K-12 assessments with college and career readiness will raise the bar for student performance. New York is one of a few states that have recently grappled with this situation. This year, the state set tougher cut scores for its accountability assessments, which produced much lower passing rates. This has prompted a public reaction that includes surprise and frustration, and some doubt that the assessment instrument is reliable.⁵

The collective nature of the consortia is a powerful ally to states in this situation. As has been noted by leaders of the New England Common Assessment Program – a partnership that formed when four small states pooled their resources and expertise to develop improved assessments – member states found that they were able to set more challenging expectations than they would have developed on their own. The partnership also helped state leaders remain resolute in resisting pressure to lower cut scores when poor test results generated political backlash. Likewise within the assessment consortia, cross-state agreement that a cut score is truly aligned with college and career readiness will establish a network of support for policymakers and education leaders as they struggle to sustain that bar.

As the assessment consortia undertake the research and development efforts described above, policy and education leaders will need to work within their own individual states to ensure that the ground is prepared to successfully implement the new system. This will require clear communications with stakeholders about what is to come and sustained awareness of the consortia developments – particularly opportunities for piloting assessment items, infrastructure requirements for administering the assessments, and the training resources that will be developed for educators. Policy and education leaders will need to carefully consider staff capacity, system capacity, and the climate in which the new comprehensive assessment system will be introduced. As states look ahead to these implementation challenges, they have the opportunity to benefit from lessons learned in the past.

STEPPING STONES TO INNOVATION: LESSONS FROM PAST EFFORTS

As noted by Brian Gong, anyone considering innovative assessments for large-scale use is likely to have a much clearer idea of how to meet technical and operational challenges than did pioneering developers in the 1990s. Past efforts to innovate have generated examples and lessons to consider as states prepare to implement the comprehensive assessment systems being developed by SBAC and PARCC. During the NRC workshop series “Best Practices for State Assessment Systems,” Brian Stecher of RAND and state leaders who played a role in three programs – the Maryland School Performance Assessment System, the Kentucky Instructional Results Information System, and the Minnesota Comprehensive Assessment in Science – reflected on lessons that were learned in test development and implementation.

Key lessons learned from past efforts:

- **The purpose and intended use of the assessment was not always clear.** Before setting out to develop an assessment, it is critical to clearly define what is to be measured and how test results will be used. During the workshop series, several assessment leaders speculated that their state efforts would have achieved greater long-term success if sufficient time had been dedicated to educate and gather input from policymakers, teachers, and the public about goals for the assessment. Each of these groups has unique expectations about what a test should accomplish, and early communication might have helped to reconcile differing expectations before the assessment tools were designed.

- **Teacher training was sometimes overlooked.** Assessment expert Brian Gong of Kentucky explained that in the past, his state ran into implementation obstacles because teachers were unsure how to adapt to expectations of the new assessment. Research and development is an intense and costly phase when establishing new assessment tools, which can mean few resources are left to help teachers adjust instruction accordingly and make the most of a new assessment tool.
- **Without a clear plan for implementation, a tight timeline can be detrimental.** The assessment consortia are operating on an ambitious timeline, with research, design, pilot testing, and full implementation to be completed by 2014. In the past, some states have been able to make valuable changes to their assessment systems within a short timeframe. However, states have also learned that if educators and parents aren't given the chance to hear about and understand the plan in the earliest stages, they are likely to perceive that the process is moving too fast and could view it as reckless. In such cases, support for the change will lag and resistance might emerge.
- **No assessment can compensate for poor learning targets.** Learning targets must be set for each grade level and should map out a steady progression of learning that culminates in college and workforce readiness at the end of grade 12. If milestones are identified in isolation within each grade level, expectations of progress are likely to be inconsistent from year to year. In 2007, the Fordham Institute found that many states set the bar significantly lower in elementary school than in middle school.⁶ This causes confusion when students appear to be on-track academically in the early years then find that they are not prepared for middle school expectations.
- **Technological tools take time and infrastructure to implement.** Technology can be used to provide more individualized assessment. However, as noted by Randy Bennett of Educational Testing Service, not every state has an adequate technological infrastructure to support wide-scale testing of large groups of students. States need time and expertise to develop the required infrastructure to meet the bandwidth and security requirements of a computer-based assessment system. For example, Tony Alpert explained that Oregon's Assessment of Knowledge and Skills is now delivered almost exclusively online, but it took the state about five years to achieve full online delivery.
- **Implementation will not succeed without stakeholder support.** It is also important to establish support for new assessments across a broad base, as leadership turnover can leave a narrowly supported effort without champions. Stakeholder involvement is essential to ensure that the individuals who will be most directly affected by the new assessment have a voice in the development process. If states are able to cultivate informed support from educators and other stakeholders, these groups will be more inclined to allow some flexibility for states to make adjustments when technical obstacles arise.

STATE LEADERSHIP IS CRITICAL

Each member state will need to implement the assessment system of its consortium by the 2014-15 school year. However, there are plenty of activities that states need to tackle in the interim. In the short term, policy and education leaders need to clarify goals for the assessments that are being developed and establish open lines of communication with stakeholders. Over the long term, state leaders will need to sustain clear communication about the developing assessment system; ensure that instructional materials and professional development are aligned to the CCSS; and foster an implementation environment that allows for adjustment when technical issues arise.

- **Communicate the vision and importance of the new assessments.** As assessment design gets underway, policymakers and education leaders need to articulate the vision of the comprehensive assessment system to key stakeholder groups. State leaders also need to develop an understanding of public attitudes toward the change. If these perspectives are not fully understood or taken into account, unexpected opposition could be encountered as the assessment program is rolled out. For instance, both of the state consortia will be creating systems that include assessment during the school year, and parents might perceive this to be "over-testing" if the vision and benefits of the new assessments are not clear. It is also important for state leaders to establish the expectation that flexibility will be needed during the implementation phase, particularly as technology becomes the new platform for assessment administration.

- **Design for gradual implementation and continual improvement.** Policymakers and education leaders need to establish a clear implementation plan that phases out current assessments and gradually phases in the new assessment system. Gradual implementation allows the state education agency the opportunity to identify problems and to seek the resources within the assessment consortium to navigate those issues. States can compare their current assessment frameworks with the CCSS to identify items that can be omitted and items that should be given more weight. For example, if a state currently assesses statistics and probability in the second grade, those concepts can be moved to assessments in later grades.
- **Ensure that measures are being taken to support teachers in this transition.** Teachers will need ample opportunity to learn about the curricular implications of innovative assessments. Teachers will also need ongoing support as they reshape instruction in light of that knowledge. To encourage such activity in states, the assessment consortia are planning to establish a shared bank of professional development resources that are aligned with the new assessment systems. State leaders can maximize the potential of these anticipated resources by ensuring that their respective states have a system in place to effectively identify and address professional development needs.
The state assessment consortia are also required to produce data that can be used in evaluations of teacher and principal effectiveness. These data will need to provide timely and constructive feedback and establish a basis for targeted professional development. Individual states that applied for Race to the Top grants have formulated plans to ensure that evaluation and professional development occur, and consortia member states can share their strategies for accomplishing these plans.⁷
- **Ensure that supports for struggling students are in place.** Once the state has implemented a new assessment system that is aligned with college and career readiness, many students could find that they are falling short of the achievement goals for their current grade level. This situation will generate anxiety among students, parents, and educators. State leaders need to begin discussions now about what it will take to design supports for students who fall short of the new academic goals. State leaders will also need to cultivate the belief that all students in their state can achieve college and career readiness. Leaders could highlight examples of schools and districts that are making this vision a reality. Each year, the Education Trust highlights schools across the nation that are accelerating learning and achievement of formerly low-performing students⁸, and a recent report by the Council of Great City Schools identifies several urban school districts that have achieved impressive gains.⁹

State leaders also need to anticipate that they will feel pressure to lower cut scores and make concessions. It is critical to plan for this moment – lowering cut scores may help ease short term political difficulties but will significantly undercut efforts to ensure that all young people are prepared for success in college or careers.

Who Should be Engaged in Conversations about Assessment?

Assessment is a cornerstone of CCSS implementation and has significant implications for the work that occurs in each community's schools and classrooms. It is critical to listen to and engage a variety of stakeholders during the design of a comprehensive assessment system.

- Educators: teachers, principals, superintendents, school board members
- Parents
- Civil rights community
- Community-based organizations
- Higher education
- Business leaders

DEFINITIONS

Below are definitions of several assessment phrases that have been used in this issue of *Blueprint*.

Comprehensive assessment systems must be aligned with college and career readiness, measure complex knowledge and skills, provide an accurate measure of student growth and achievement during the academic year, and generate data that can be used for accountability and instruction. Such systems are the focus of the U.S. Department of Education's Race to the Top Assessment Program, and a detailed set of requirements has been articulated by the Department.

Formative assessments are typically embedded in instructional activities to provide teachers with immediate feedback about student learning.

Interim assessments are administered periodically throughout the school year and may be used to predict student performance on state assessments, to provide data on an instructional program or strategy, or to provide diagnostic information on a particular student.

Summative assessments measure what students have learned at the end of a set of learning activities. State and national achievement tests administered at the end of a school year are summative assessments.

Through-course assessments are defined within the U.S. Department of Education's Race to the Top Assessment Program. These assessments are administered in components periodically during the academic year. A student's results from through-course assessments are combined to produce the student's total summative assessment score for that academic year.

Innovative assessments can be defined as anything that is not a multiple-choice, paper-and-pencil test. NRC workshop participants noted that a test might be innovative because it:

- Incorporates more complex prompts than are typical in a printed test, such as hands-on materials, video, or multiple types of materials;
- Offers different kinds of response options, such as written responses, collections of materials (portfolios), or interactions with a computer; and
- Is delivered in an innovative way, usually through technology.

Performance assessment items require students to demonstrate their knowledge or skill by completing a task, generally in a manner that closely resembles a real-life situation.

Constructed Response items require that the student provide the correct answer rather than choosing from a menu of possible answers.

Computer Adaptive Tests (CATs) adjust to the test taker based on the level of difficulty of items answered correctly or incorrectly. CATs aim to produce more precise information on an individual student's learning than non-adaptive assessments, particularly for very high- and low-achieving students.

Psychometricians study ways to infer what students know, understand, or can do within an assessment setting. These experts are primarily concerned with the construction and validation of measurement instruments.

Cognitive Psychologists study the ways individuals perceive and process information. In recent years, cognitive researchers have begun to conduct more studies with teachers in actual classroom settings.

REFERENCES

- ¹ National Research Council. (2010). *State Assessment Systems: Exploring Best Practices and Innovations: Summary of Two Workshops*. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. The National Academies Press. Available online: http://www.nap.edu/catalog.php?record_id=13013#aboutprepub
- ² National Center for Education Statistics. (2007). *Mapping 2005 State Proficiency Standards onto the NAEP Scales*. Research and Development Report. Available online: <http://nces.ed.gov/nationsreportcard/pubs/studies/2007482.asp>
Petrilli, M. (2007). *Presentation to the National Research Council Workshop on Assessing the Role of K-12 Academic Standards in States*. Available online: <http://www7.nationalacademies.org/cfe/Petrilli%20Presentation.pdf>
- ³ Government Accountability Office. (2003). *Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies*. GAO Report-03-389. Available online: www.gao.gov/cgi-bin/getrpt?GAO-09-911
Toch, T. (2006). *Margins of Error: The Testing Industry in the No Child Left Behind Era*. Education Sector. Available online: http://www.educationsector.org/research/research_show.htm?doc_id=346734
- ⁴ Michael Lawrence Collins. (2008). *It's Not About the Cut Score: Redesigning Placement Assessment Policy to Improve Student Success*. Jobs for the Future. Available online: <http://www.policyarchive.org/handle/10207/8669>
- ⁵ Jennifer Medina. (2010). "Students' Passing Rates Plummet in New York." *The New York Times*. July 28, 2010. Available online: <http://www.nytimes.com/2010/07/29/education/29scores.html>
- ⁶ John Cronin, Michael Dahlin, Deborah Adkins, and G. Gage Kingsbury. (2007). *The Proficiency Illusion*. Thomas B. Fordham Institute and Northwest Evaluation Associates. Available online: http://www.edexcellence.net/doc/The_Proficiency_Illusion.pdf
- ⁷ U.S. Department of Education. (2010). "Race to the Top Fund: States' Applications for Phase 2." Available online: <http://www2.ed.gov/programs/racetothetop/phase2-applications/index.html>. See section D(2) of the application.
- ⁸ Information about winners of The Education Trust's "Dispelling the Myth" award is available online: <http://www.edtrust.org/dc/resources/success-stories>
- ⁹ Council of Great City Schools. (2010). *Beating the Odds: An Analysis of Student Performance on State Assessments and NAEP*. Available online: <http://www.cgcs.org/publications/achievement.aspx>

The James B. Hunt, Jr. Institute for Educational Leadership and Policy, an affiliate of the University of North Carolina at Chapel Hill, is a non-partisan, non-profit service entity that does not lobby for, or take positions on, state and federal legislation.